Quantitative Fluorescence Microscopy Studies of Prostate Cancer at the Single-Cell Level Using Automated Image Analysis

Joshua Miguel Gonzalez

MSc by Research

University of York

Physics, Engineering and Technology

March 2026

Abstract

Prostate cancer is associated with altered nuclear and nucleolar organisation, but fluorescence microscopy studies in prostate cell lines commonly use permeabilised immunofluorescence workflows and rarely include a non-permeabilised baseline. It is often unclear whether permeabilisation is a neutral preparation step or whether it changes the measurements used for interpretation. This thesis established a fluorescence microscopy workflow for preparation-aware, single-nucleus analysis in two prostate cell lines: LNCaP (malignant) and PNT2-C2 (non-malignant). Fixed cells were analysed under three matched conditions: non-permeabilised DAPI staining, permeabilised DAPI-only staining, and permeabilised DAPI plus anti-fibrillarin (DAPI+FIB) immunofluorescence. Nuclear morphology, DAPI-derived intensity measures, and fibrillarin-defined nucleolar measures were quantified per nucleus, with distribution-level analysis alongside mean values.

Convergence testing showed that descriptor estimates stabilised with increasing nucleus count, and that LNCaP required larger sample sizes than PNT2-C2, consistent with greater heterogeneity. In the non-permeabilised baseline, LNCaP nuclei were $\sim 45\%$ larger, $\sim 20\%$ more elongated, and showed greater dispersion in intranuclear DAPI staining/texture, with mean DAPI intensity coefficient of variation $\sim 40\%$ higher and the kernel density estimate (KDE) mode $\sim 60\%$ higher. However, permeabilisation materially altered several measurements. Most clearly, it reversed the cell-line ordering for nuclear area, and strongly impacted absolute DAPI intensity and density, showing that permeabilisation was not neutral for all descriptors. By contrast, nucleolar analysis under the DAPI+FIB workflow showed little between-line difference in nucleolar area, but clearer elevation of nucleolar intensity in LNCaP, with mean nucleolar intensity $\sim 25\%$ higher and the nucleoli:DAPI ratio $\sim 35\%$ higher.

Overall, quantitative single-nucleus fluorescence microscopy identified robust differences between malignant and non-malignant prostate cell models, but only under a preparation-aware interpretation. The non-permeabilised condition was essential as both a rarely evaluated baseline, and the reference needed to show that permeabilisation can itself reshape nuclear readouts. Key limitations are preparation-sensitive fluorescence measurements and segmentation/thresholding errors.

241 pages

Acknowledgements

Firstly, I must once again express my most sincere and heartfelt thanks to my partner, Devi. I cannot begin to describe how much their support has meant to me. Their endless patience, compassion, and kindness throughout this process has gone above and beyond what could be asked of anybody, and anything I could possibly hope to convey here would be hopelessly inadequate compared to the true extent of my feelings. From the bottom of my heart, thank you.

I am also deeply grateful to Al, Hobie, Matt, and Zoë, my knights in shining armour, it takes a village to raise a child; and this thesis has been no exception. Their willingness to put up with my (many) neuroses, and to take care of things that should rightly have been my responsibility, allowed me to fully dedicate myself to my work. Most importantly, thank you all for keeping me company while I've been in Hell. Truly, your love, friendship, and support mean the world to me.

Finally, I would like to thank my supervisor, Yvette Hancock, for the opportunity to complete this project, and for her support in helping me to get it over the line.

Everything has a beginning and an end. Life is just a cycle of starts and stops. There are ends we don't desire, but they're inevitable, we have to face them. It's what being human is all about.

Author's declaration

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for a degree or other qualification at this University or elsewhere. All sources are acknowledged as references.

Contents

1 Introduction **7**

1.1 Overview, aims, and research questions 7

1.1.1 Core study design and Results mapping 7

1.1.2 Research questions and candidate hypotheses 8

1.2 Prostate cancer clinical context 9

1.2.1 Prostate anatomy and function 9

1.2.2 Overview of prostate cancer 10

1.2.3 Histological grading systems and staging in prostate cancer 13

1.3 Current diagnostic pathway and limitations 15

1.3.1 Clinical tests for prostate cancer 15

1.4 Why analysis at the cell/organelle scale is complementary 19

1.4.1 Clinical motivation: improving discrimination of consequential disease 19

1.4.2 Scientific motivation: mechanistic characterisation of malignant state in controlled models 19

1.5 Biological models in prostate cancer 20

1.6 Cell lines and single-nucleus heterogeneity 23

1.6.1 Cell lines used in this thesis: LNCaP and PNT2-C2 23

1.6.2 Single-nucleus heterogeneity as a primary analysis target 24

1.7 Cell biology fundamentals relevant to measured endpoints 25

1.7.1 The nucleus as a cancer-relevant cellular feature 25

1.7.2 DAPI-based nuclear descriptors 27

1.7.3 Nucleolar state and fibrillarin as a nucleolar marker 28

1.8 Optical methods background and fluorescence-imaging literature 30

1.8.1 Context: other single cell measurement modalities 30

1.8.2 Fluorescence microscopy for quantitative nuclear/nucleolar assessment 30

1.8.3 Fluorescence imaging synthesis table: what is commonly quantified/reported . 32

1.9 Study positioning and novelty 36

1.9.1 Positioning relative to fluorescence-imaging practice 36

1.9.2 Novelty pillars 36

2 Methods **38**

2.1 Fluorescence microscopy theory 38

2.1.1 Photophysical basis of fluorescence contrast . . . 38
2.1.2 Rate model, quantum yield, and measured intensity . . . 40
2.1.3 Epifluorescence microscopy instrumentation . . . 41
2.1.4 Fluorophores and labelling strategy . . . 41
2.2 Cell culture, sample preparation, and fluorescence imaging . . . 43
2.2.1 Cell culture and maintenance . . . 44
2.2.2 Coverslip preparation, seeding, and serum starvation . . . 44
2.2.3 Fixation, permeabilisation, and immunostaining . . . 45
2.2.4 Mounting, curing, and storage . . . 46
2.2.5 Fluorescence image acquisition . . . 47
2.3 Automated computational analysis pipeline . . . 47
2.3.1 Analytical rationale and data flow . . . 47
2.3.2 Implementation overview . . . 50
2.3.3 Stage 1: DAPI-driven nuclear segmentation and per-nucleus feature extraction 51
2.3.4 Stage 2: pooled descriptive analysis and repeated random subsampling convergence-series generation . . . 60
2.3.5 Stage 3: KDE estimation and parametric distribution fitting . . . 64
2.3.6 Stage 4: convergence modelling and convergence-point estimation . . . 66

3 Results: Analysis of non-permeabilised cells 71
3.1 Convergence testing . . . 74
3.1.1 Convergence behaviour and robustness checks . . . 74
3.1.2 Metric selection . . . 79
3.1.3 Summary and implications . . . 82
3.2 Morphological and intensity analysis . . . 89
3.2.1 Morphology . . . 89
3.2.2 Intensity Analysis . . . 101
3.3 Summary . . . 113

4 Results: Analysis of permeabilised cells 116
4.1 DAPI-only analysis . . . 118
4.1.1 Convergence testing . . . 118
4.1.2 Morphological and intensity analysis . . . 124
4.1.2.1 Morphology . . . 124
4.1.2.2 Intensity Analysis . . . 134
4.1.3 Interim conclusions . . . 146
4.2 DAPI and anti-fibrillarin analysis . . . 147
4.2.1 Convergence testing . . . 148
4.2.1.1 Key quantitative outcomes . . . 148
4.2.2 Morphological and intensity analysis . . . 152
4.2.2.1 Morphology . . . 152
4.2.2.2 Intensity analysis . . . 166
4.3 Summary . . . 187

5 Conclusion 191

5.1 Overview and final answer to the overarching thesis question . . . 191
5.2 Principal findings . . . 192
5.2.1 Sample-size sufficiency and nucleus-level heterogeneity . . . 192
5.2.2 Baseline nuclear differences in the non-permeabilised condition . . . 192
5.2.3 Effect of permeabilisation on DAPI-derived descriptors . . . 193
5.2.4 Direct nucleolar findings from fibrillarin staining . . . 193
5.3 Scientific meaning and contribution . . . 194
5.4 What the data do and do not show . . . 195
5.5 Limitations of the study . . . 196
5.6 Recommendations for future work . . . 197
5.7 Final statements . . . 198

Bibliography **199**

A Methods **234**
A.1 Exact implementation details for Stage 1 DAPI segmentation and FIB nucleolar assignment 234

B Convergence **236**
B.1 Non-permeabilised data . . . 236
B.1.1 Summary table . . . 236
B.2 DAPI-only permeabilised data . . . 238
B.2.1 Summary table . . . 238
B.3 DAPI+FIB permeabilised data . . . 238
B.3.1 Summary table . . . 238

Chapter 1

Introduction

1.1 Overview, aims, and research questions

This thesis is positioned as a comparative, cell line study that interrogates nucleus and nucleolus resolved fluorescence descriptors in a malignant prostate cancer model (LNCaP), and an immortalised non-malignant comparator (PNT2-C2), using epifluorescence microscopy [1–4]. The primary measurement unit here is the individual nucleus, with nucleoli defined as fibrillarin-positive subnuclear regions within each nuclear mask (formalised in Section 1.7.3). The overarching motivation is complementary to primary cell culture and tissue-based studies; quantitative nucleus/nucleolus scale descriptors are analysed to provide mechanistically interpretable descriptors of the malignant state, on the premise that controlled characterisation of malignant and non-malignant cells in a simple model system may reveal differences that can subsequently be examined further in more complex and heterogeneous systems.

Two explicit scope limits are imposed throughout. First, all biological claims are framed at the cell line level, under controlled preparation and acquisition; no authoritative claims are made regarding aggressiveness in primary cell cultures or patients, clinical performance, or population risk [5–7]. Second, fluorescence intensity-derived variables are treated as representative readouts, explicitly dependent on sample preparation and imaging parameters. Where intensity shifts are discussed, they are interpreted comparatively under matched protocols rather than as absolute molecular quantification. As no calibration was performed using DNA standards, absolute DNA content can only be inferred rather than quantified exactly (more detail given in Sections 1.7.2 and 1.7.3) [8, 9].

1.1.1 Core study design and Results mapping

The Results are organised around three preparation/marker conditions that enable controlled attribution of observed measurement shifts:

1. Non-permeabilised, fixed cells stained with DAPI (baseline, detergent-free preparation),
2. Permeabilised, fixed cells stained with DAPI (control condition to quantify the effect of detergent permeabilisation on nuclear descriptors);
3. Permeabilised, fixed cells stained with DAPI and anti-fibrillarin (DAPI+FIB, nucleolus-resolved

condition enabling fibrillarin-defined nucleolar characterisation).

Epifluorescence is used as the primary imaging modality because it is aligned with high throughput nucleus-resolved sampling in 2D monolayers, while recognising the reduced optical sectioning relative to confocal microscopy (with a more detailed discussion carried out in Section 1.8.2) [10–12]. Within each condition, the emphasis is on distributional analysis; per-nucleus descriptor distributions are retained and examined, rather than being collapsed into a single mean or median (more details given in Section 1.6.2).

1.1.2 Research questions and candidate hypotheses

The overarching question of this thesis is whether quantitative single-nucleus fluorescence microscopy can identify reproducible and biologically interpretable differences in nuclear and nucleolar state between the malignant LNCaP model and the non-malignant comparator PNT2-C2, once sample-size sufficiency and preparation-dependent effects are accounted for explicitly.

This overarching question is addressed through four linked research questions (RQs), each mapped directly onto the Results structure above.

1. **Sampling sufficiency for single-nucleus analysis.** What sample sizes are required to stabilise key nuclear and nucleolar descriptor estimates in LNCaP and PNT2-C2 under each preparation condition, and do the differences in sampling requirements mirror expected differences in heterogeneity between cell lines?

2. **Preparation dependence (permeabilisation as an experimental factor).** To what extent does detergent permeabilisation affect nucleus-level descriptors, and are permeabilisation-induced shifts similar across LNCaP and PNT2-C2?

3. **Comparative distributional behaviour (malignant vs comparator).** Across preparations, do LNCaP and PNT2-C2 exhibit distinct distributional patterns for measured nuclear descriptors, and if so, are these differences mechanistically consistent with malignancy-associated nuclear state at the model level?

4. **Nucleolar state via fibrillarin.** In the permeabilised DAPI+FIB condition, does LNCaP exhibit increased nucleolar prevalence and/or nucleolar signal density relative to PNT2-C2, consistent with the expected altered nucleolar state in the malignant model?

The candidate testable hypotheses synthesised from the above research questions are therefore:

1. **Heterogeneity as a quantifiable metric requires sample size validation.** Statistically converged sampling enables nucleus level heterogeneity to function as a reproducible descriptor. As a working expectation, the malignant model is predicted to exhibit greater heterogeneity than the non-malignant comparator, implying a larger required number of nuclei to achieve a fixed precision target in LNCaP than in PNT2-C2 [13–15].

2. **Permeabilisation is not a neutral preparation step.** Detergent permeabilisation is hypothesised to induce measurable shifts in one or more nuclear descriptors, and the direction and magnitude of these shifts are treated as empirical outcomes rather than assumed constants [11, 16].

3. **Altered nuclear and nucleolar state in the malignant model.** Relative to PNT2-C2, LNCaP is hypothesised to exhibit both an altered nuclear and nucleolar state characterised by (at minimum) increased nuclear and nucleolar area, and/or increased nucleolar signal density [2, 17, 18].

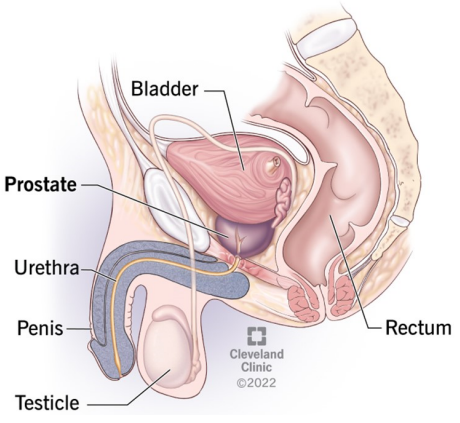
In brief, the novelty of this thesis is that it combines distribution-level, single nucleus fluorescence-based characterisation with explicit verification of sample size sufficiency by statistical convergence testing and quantification of preparation-induced shifts (such as the effects of permeabilisation), extending beyond simple nucleolar size/number to also include nucleolar signal density; the detailed novelty defence is deferred to Section 1.9 and grounded by the fluorescence-imaging synthesis in Table 1.1.

The remainder of the Introduction aims to establish the prostate cancer clinical context and key grading definitions, summarise the contemporary diagnostic pathway and its limitations, motivate analysis at the nuclear/nucleolar scale, justify the model choice and distribution level single nucleus framing, define the biological interpretation context for the measured descriptors, and provide a brief primer for the optical methods used in this study, and literature synthesis that support the stated novelty framing.

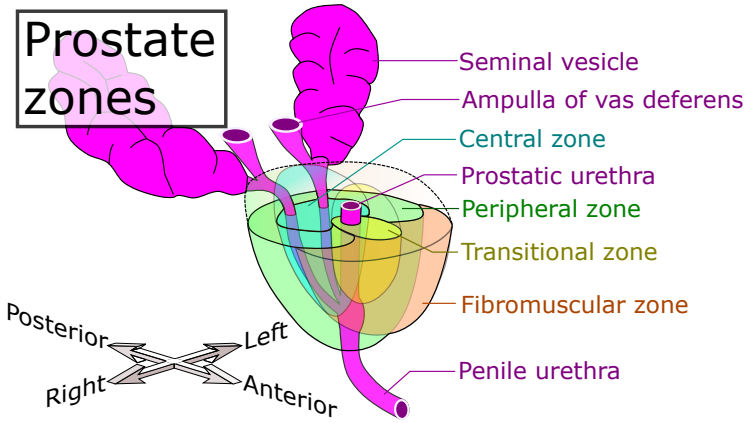
1.2 Prostate cancer clinical context

1.2.1 Prostate anatomy and function

The prostate is a small (2 cm to 3 cm in width), roughly conical shaped fibromuscular gland present in male mammals [19]. It forms part of the reproductive system and is located just below (inferior to) the bladder, surrounding the proximal urethra [20, 21]. It is separated from the rectum by a thin membrane known as Denonvilliers' fascia [22, 23]. This anatomical relationship is clinically relevant because disease in the posterior gland can be more readily detected by digital rectal examination (DRE), whereas the anterior and lateral aspects are less accessible to palpation [24]. A more detailed discussion of DRE within the diagnostic pathway is provided in Section 1.3.1. The anatomical location of the prostate relative to other pelvic organs, along with a standard zonal schematic is shown in Figure 1.1.



(a) Cross-section of the pelvis showing the location of the prostate from Cleveland Clinic (2022) [22].



(b) Detailed view of the prostate showing the various zones it comprises created by Häggström (2019) [25].

Figure 1.1: Diagrams showing (a) the location of the prostate relative to other pelvic organs and (b) the major zonal subdivisions commonly used in clinical descriptions.

The prostate is commonly divided into glandular zones, each with different functions and disease susceptibility. The transitional zone (TZ, yellow in Figure 1.1(b)) surrounds the proximal urethra and is typically small, however may enlarge with age, eventually progressing to benign prostatic hyperplasia (BPH) [20, 26]. The central zone (CZ, blue in Figure 1.1(b)) lies at the base of the prostate and surrounds the ejaculatory ducts. The peripheral zone (PZ, green in Figure 1.1(b)) occupies the posterior and lateral portions of the gland, and makes up the majority of its volume. Nearly three quarters of prostate cancers arise within the PZ [23]. Finally, the anterior fibromuscular stroma (non-glandular tissue sometimes known as the fibromuscular zone, orange in Figure 1.1(b)) lies in front of the TZ and urethra [20]. These zones are enclosed by a fibrous layer known as the prostatic capsule.

One of the main functions of the prostate is the production of prostatic fluid, which makes up about 25-30% of the volume of semen [27]. This is an alkaline, milky fluid, rich in enzymes (including prostate-specific antigen; PSA) and nutrients (citrate, zinc and proteins). Prostatic fluid helps neutralise the acidity of the vaginal tract, provides energy and support for sperm, prolongs survival and aids motility during intercourse, and helps propel semen during ejaculation [21, 23]. PSA also contributes to the liquefaction of semen, further aiding mobility. In addition to secretion, the smooth muscle fibres of the prostate contract during ejaculation to expel fluid into the urethra, alongside sperm from the testes and fluid from the seminal vesicles [20]. Overall, the prostate plays a vital supportive role in the reproductive system by producing a nutritive, buffering fluid and helping to transport sperm through the urethra.

Prostatic tissue can be affected by a variety of conditions including benign enlargement and inflammation, as well as more serious diseases such as cancer [28, 29]. BPH is a non-cancerous growth of glandular and stromal tissue, typically arising in the transitional zone as mentioned earlier [26]. As a person ages, the TZ often expands, which compresses the urethra and may cause difficulty with urination. Prostatitis (infection in or inflammation of the prostate) can also occur, causing pain and urinary irritative symptoms, but is not typically a risk factor for cancer [30]. In contrast, prostatic carcinoma, a malignant tumour arising from epithelial cells (most commonly adenocarcinoma, which originates from glandular cells) usually develops in the peripheral zone [21, 31]. Early-stage prostate cancer often produces no symptoms, partly because tumour growth in the PZ does not immediately block urine flow. When advanced, tumours may invade locally into adjacent structures such as the bladder neck or seminal vesicles, potentially causing urinary obstruction or haematuria (blood in urine). These considerations are relevant to the later tests discussed in Section 1.3.1 because the zonal anatomy shown in Figure 1.1 shapes both anatomical access and the distribution of clinically relevant disease.

1.2.2 Overview of prostate cancer

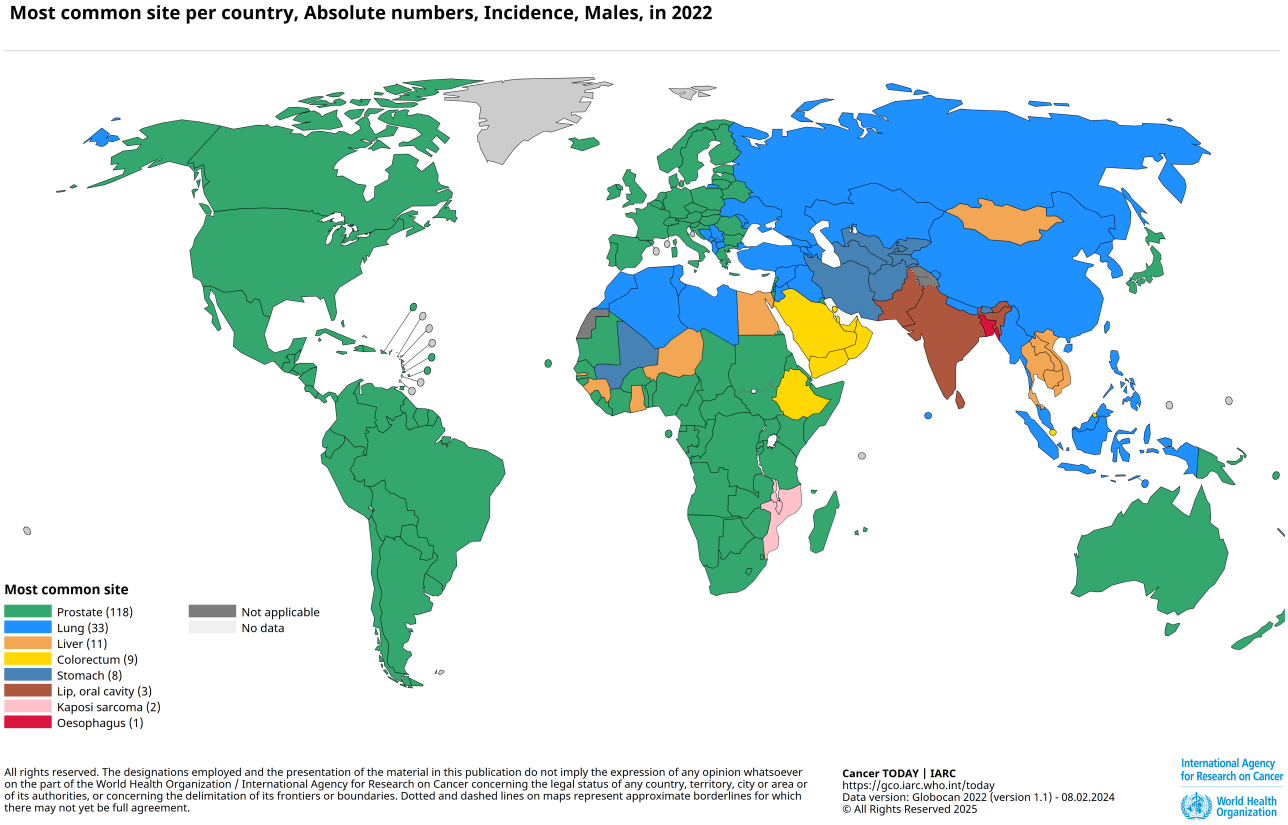
Prostate cancer (PCa) is among the most prevalent malignancies affecting men globally and remains a major cause of cancer mortality [32, 33]. Although microscopic, indolent prostate lesions are highly common in ageing populations, such that some studies have suggested that all men will likely develop some form of prostate cancer if they live long enough [34]. The clinical burden of the disease is driven disproportionately by the subset of tumours that progress into locally advanced or metastatic disease [35, 36]. The central clinical challenge is thus not merely detection, but stratification; stated plainly, distinguishing indolent disease that can be monitored safely without requiring intervention from biologically aggressive disease that necessitates timely treatment [37–39].

Global incidence and mortality vary significantly by region, and it is important to state explicitly what is meant by such disparities. Visualisations of the leading cancer site among men by incidence and by mortality are shown in Figure 1.2 using GLOBOCAN 2022 estimates (Cancer Today) [32, 40, 41]. These maps illustrate the stated disparity in incidence and mortality clearly: PCa frequently dominates male cancer incidence in all of the Americas, as well as most of Western Europe, Africa, and Oceania, however it is the leading cause of cancer related deaths only in a select few countries in the Americas and sub-Saharan Africa [32, 40]. Early detection is a key factor in reducing PCa mortality [33, 35, 36, 42], and so these geographical differences in incidence and mortality may be due in part to substantial differences in stage at diagnosis, access to early detection pathways, and access to curative or life prolonging treatments between wealthier, more developed (i.e. having a higher Human Development Index) countries and less affluent, developing countries (such as many of those in the Global South) [35, 36, 43–46].

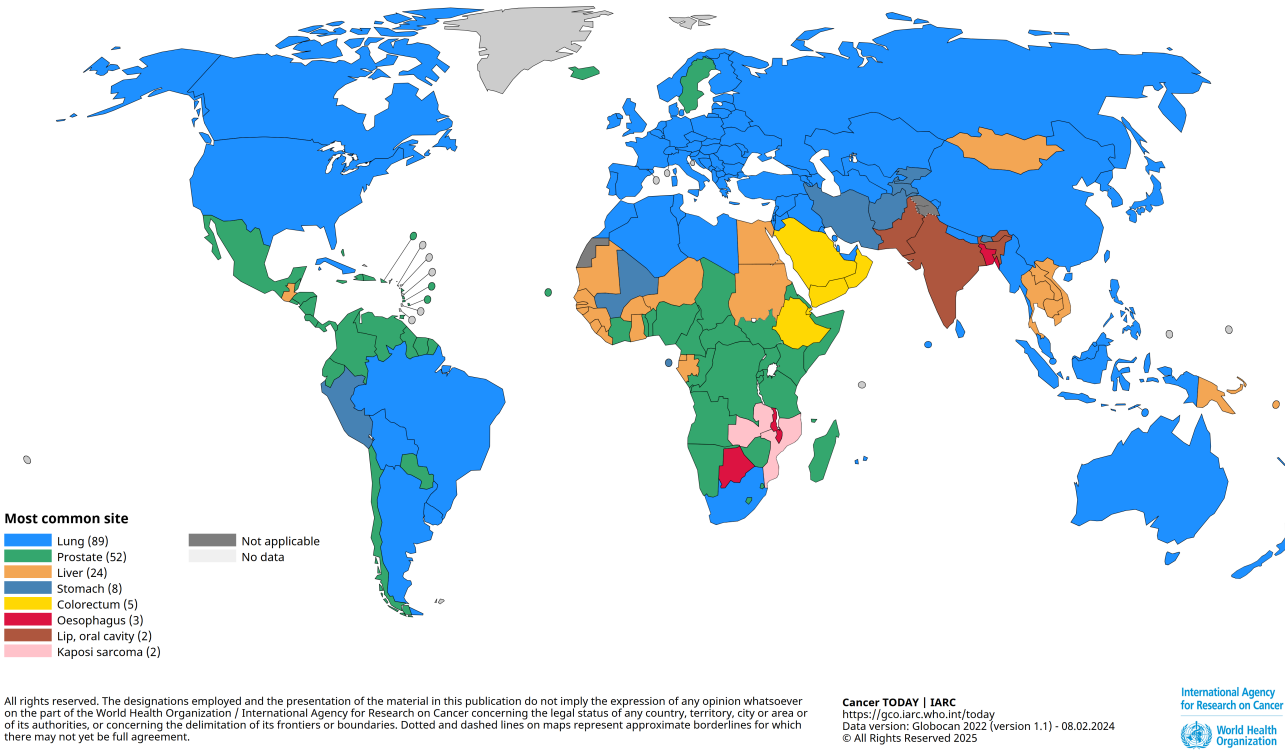
In the UK specifically, PCa is the most common cancer in men, and therefore constitutes a major population level health burden, despite the predominance of indolent disease in many diagnoses [47]. Low-grade/risk PCa makes up around 70% of new diagnoses; these forms of PCa are relatively harmless and may not cause any significant symptoms during the lifetime of the patient, and it is a clinically important subset of cases which behave aggressively and carry a substantially greater risk of progression and death [32, 37, 38, 48]. People with low-grade PCa can often survive many years and die from causes unrelated to the disease, whereas those with high-grade cancers face a much higher risk of death (28% mortality for low-grade vs. 64% mortality for high grade) [39]. It should be noted that while there is a general correlation between low grade cancer and low risk, this is not always the case. Patel et al. (2025) for example found that 1 in 6 men who had grade group 1 PCa actually had intermediate or high risk disease [49]. In the US the overall risk of death from PCa is around 2.5% [50], so accurately diagnosing and categorising the condition is important to ensure the patient receives the correct treatment (if any is required), and is not subject to unnecessary, potentially invasive procedures [29, 51, 52].

There are many established risk factors for PCa including increasing age, family history, and ancestry-associated differences in incidence and severity (black people for example are known to have an elevated risk of developing more aggressive prostate cancer) [33, 53, 54]. Here, these factors are introduced only to contextualise the disease burden, however, a representational limitation is relevant later when interpreting cell-line results; widely used prostate cell lines are derived from a relatively small number of donors, a disproportionately large number of whom are white men, and as a result high risk groups are often under-represented in the wider prostate cell line literature [55–58].

At the time of writing, there is no universal PCa screening programme in the UK due to the need to balance the benefits of a given programme against the disadvantages and avoid unnecessary tests and treatment (though a report from the National Screening Committee based on a recent review of practices is due to be published in March 2026 [59]). Instead, PSA testing is available via symptomatic assessment pathways and through informed choice testing on request (i.e. opportunistic case finding), with earlier discussion/testing commonly recommended for higher-risk groups [60, 61]. Here, "screening" is used in the strict public-health sense of an organised, invitation-based programme targeting an asymptomatic population. A more detailed discussion of the current diagnostic pathway is carried out in Section 1.3.



(a) Leading cancer site among men by estimated incidence [41].



(b) Leading cancer site among men by estimated mortality [41].

Figure 1.2: Maps showing the leading cancer site among men by estimated incidence (a) and estimated mortality (b), generated using data visualisation tools from the Cancer Today website [32, 40, 41]. Note that prostate cancer frequently leads incidence in many regions, whereas the leading cause of male cancer mortality is often not prostate cancer with a few exceptions, highlighting geographically variable diagnostic and treatment contexts.

1.2.3 Histological grading systems and staging in prostate cancer

Gleason patterns and Grade Groups

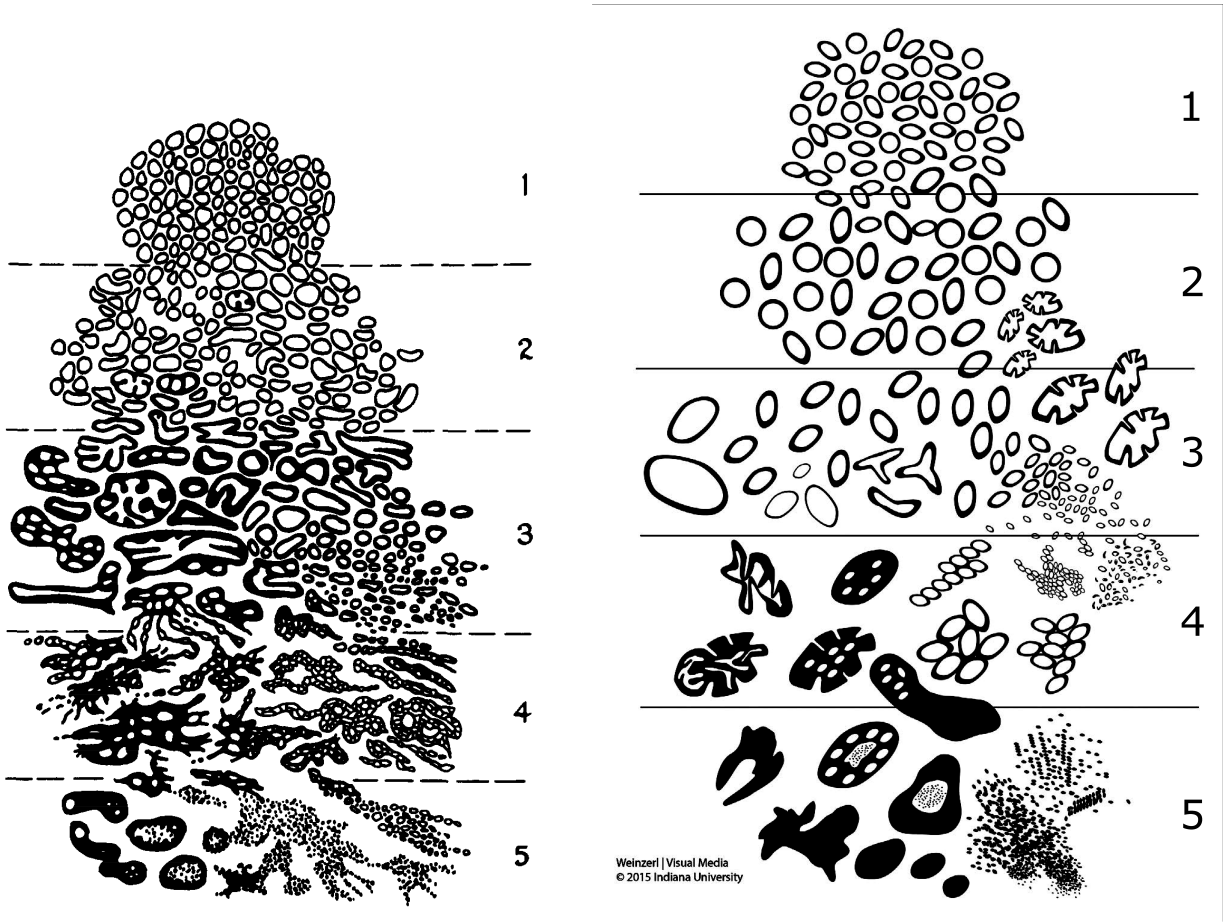
Across prostate cancer literature, many diagnostic performance endpoints and clinical risk stratification frameworks are expressed in terms of histological grade. It is therefore necessary to define the grading system typically used to define "clinically significant" disease before discussing pathway performance metrics that are defined against grade thresholds.

The Gleason system is the dominant histological grading framework for prostate adenocarcinoma, originally proposed in the 1960s and subsequently refined through multiple updates [62, 63]. Gleason grading is based on architectural pattern recognition in haematoxylin and eosin (H&E) stained tissue sections, rather than individual cell morphology. In the classical schematic formulation, pattern 1 represents well-defined nodules of uniform, closely packed glands, pattern 2 shows less uniform and more loosely arranged glands with early infiltration, pattern 3 denotes clearly infiltrative but still discrete gland units, pattern 4 encompasses poorly formed, fused, or cribriform gland architecture with loss of discrete gland separation and pattern 5 represents essentially absent gland formation (manifesting as solid sheets, cords, or single cells), sometimes with comedonecrosis [64, 65]. Modern practice for needle-biopsy based reporting effectively restricts clinically assigned patterns to 3-5 for cancer; patterns 1-2 are rarely assigned other than in very specific exceptions, and are not typically used for cancer diagnosis in contemporary reporting [66–68]. A diagram illustrating the different patterns within the Gleason grading system is shown in Figure 1.3.

In practical terms, a given specimen is assigned both a primary pattern (most prevalent) and a secondary pattern (second most prevalent). The reported Gleason score is the sum of these two pattern numbers (e.g. 3+4). Ordering matters because it informs of the dominant pattern: $3 + 4$ indicates predominantly pattern 3 architecture with a lesser component of pattern 4, whereas $4 + 3$ indicates predominantly pattern 4 architecture, and is typically interpreted as more adverse within the same summed score band [63, 66]. In contemporary reporting (and especially needle biopsies), Gleason scores most commonly range from 6 to 10, reflecting the practical restriction of defining cancer as patterns 3-5 [68]. Additionally, in the case of needle-biopsy grading, the secondary pattern is often replaced by the highest pattern present (i.e. a patient with mostly normal prostate tissue, but a small amount of highly abnormal tissue may be graded as 1+5, even if there was also a substantial proportion of pattern 2 tissue) [69–71].

The Gleason grading system has continued to evolve as additional evidence has accumulated, with the latest major modern refinements formalised in the 2014 ISUP consensus and subsequent recommendations being published in 2016 [66, 68, 71]. One of the key improvements coming from these recommendations was in collecting various Gleason scores into Grade Groups (GG), with the aim of improving interpretability, reproducibility, and communication. Grade Groups provide a five-tier ordinal categorisation running from GG1-GG5: in the standard usage of this framework, GG1 corresponds to any Gleason score ≤ 6, GG2 corresponds to Gleason 3+4, GG3 corresponds to Gleason 4+3, GG4 corresponds to Gleason score 8 (most commonly 4+4, but also 3+5 or 5+3), and GG5 corresponds to Gleason scores 9-10 (i.e. 4+5, 5+4, or 5+5) [63, 66, 67].

Although Gleason score and Grade Group are indispensable clinical frameworks, they are fundamentally tissue-architecture based systems defined on H&E sections, and therefore are not reproduced



(a) Illustration of the original Gleason grading system created by Donald Gleason, adapted from Gleason (1992) [62].

(b) The latest updated Gleason grading system, adapted from Gordetsky and Epstein (2016) [66].

Figure 1.3: Schematic overview of Gleason grading system, used to assess the architectural patterns in prostate adenocarcinoma tissue samples (not individual cell morphology). Tissue can be graded from 1-5, ordered from well-formed, discrete glands (low Gleason pattern) towards poorly formed/fused glands, and ultimately solid growth with absent gland formation (high Gleason pattern). Gleason scoring is based on the two most prevalent architectural patterns in the specimen; the dominant (primary) pattern and the second most prevalent (secondary) pattern are each assigned a number (1-5) and reported as a sum (e.g. 3+4) (note however that in the case of needle biopsies, Gleason score may be calculated as the primary pattern + the highest pattern present, or alternatively as the primary + secondary pattern, with the highest pattern present included as a tertiary grading) [69, 70]. Ordering is important, even though 3+4 and 4+3 both have a score of 7, they have different dominant architectures, and are therefore interpreted differently in terms of risk stratification. [62, 66]

directly by the present workflow, which instead interrogates fixed cell cultures at single-nucleus and nucleolus-resolved scale. This grading discussion remains relevant, however, because prostate pathological interpretation has long involved not only glandular architecture, but also accompanying cytological features, including nuclear atypia and prominent nucleoli, particularly in diagnostically challenging material [63, 72, 73]. As such, this thesis makes no attempt to infer Gleason grade from cell lines; rather, it asks whether controlled, preparation-aware distributions of nuclear and nucleolar descriptors can capture reproducible malignant-associated differences in a simplified model system that is complementary to tissue-based grading.

Prostate cancer staging: the TNM system

Prostate cancer stage is conventionally described using the tumour-node-metastasis (TNM) system, originally developed by Pierre Denoix and subsequently maintained and revised by the Union for International Cancer Control (UICC) [74, 75]. TNM is an anatomical classification: T denotes the

extent of the primary tumour, N the absence or presence of regional lymph node metastasis, and M the absence or presence (and site) of distant metastasis [75]. Summarily, TNM provides a concise, standardised shorthand for the disease extent at diagnosis, and is distinct from histopathological grading systems (such as Gleason grading as detailed in Section 1.2.3), which quantify tumour differentiation rather than anatomical spread.

Within the prostate-specific TNM framework, organ-confined disease corresponds to T1-T2: T1 denotes a clinically inapparent, non-palpable (i.e. not detectable by touch) tumour (including biopsy-detected disease), whereas T2 denotes a palpable tumour confined within the prostatic capsule [75]. More extensive local disease is captured by T3 (extraprostatic extension, including seminal vesicle invasion) and T4 (tumour fixed to, or invading, adjacent structures) [75]. Regional nodal status is defined with respect to the true pelvic lymph nodes (below the bifurcation of the common iliac arteries): N0 indicates no regional lymph node metastasis, while N1 indicates regional lymph node metastasis [20, 75]. Distant spread is encoded by M: M0 indicates no distant metastasis, while M1 indicates distant metastasis, further subclassified as M1a (non-regional lymph node(s)), M1b (bone), and M1c (other site(s)) [75]. As such, "localised" prostate cancer most commonly corresponds to T1-T2 (with N0 and M0), whereas metastatic disease is defined by any M1 category, irrespective of T and N [75].

Prostate Cancer UK defines advanced (metastatic) disease as cancer "that has spread from the prostate to other parts of the body" via the bloodstream or lymphatic system [76]. In other words, metastasis is the process by which malignant cells detach from the primary tumour, invade the surrounding tissue, intravasate into lymphatic or blood vessels, and colonise distant tissues [77]. Metastatic spread from the prostate typically occurs through one of several defined pathways. Prostate cancer cells may migrate directly into adjacent tissues, but typically spread via the pelvic lymphatic network and venous channels [78, 79]. The most common sites of metastasis are regional pelvic lymph nodes and the axial skeleton. Bones (especially vertebrae, pelvis, ribs) are also a favoured site, as tumour cells can lodge in the marrow and exploit the rich growth factors there, and prostate cancer preferentially binds to bone marrow cells (though the exact mechanism governing this is still not understood) [80]. Importantly, even after spreading, metastatic lesions (typically) remain histologically identifiable as prostate carcinoma, rather than forming a new cancer of the bone or lymph node, though they may undergo alterations such that they do not appear identical to the tumour from which they originated [81–84].

1.3 Current diagnostic pathway and limitations

1.3.1 Clinical tests for prostate cancer

In the UK, the diagnostic pathway for suspected prostate cancer typically begins with prostate specific antigen (PSA) testing, with a digital rectal examination (DRE) offered as an additional test where clinically indicated. Patients are then referred into secondary/tertiary care for multiparametric magnetic resonance imaging (mpMRI) as the first line imaging investigation, with biopsy and histopathology used for tissue confirmation when indicated. These authoritative statements of the typical pathway are guided by National Institute for Health and Care Excellence (NICE) suspected cancer referral guidance (NG12) and NICE prostate cancer diagnosis/management guidance (NG131), alongside NHS patient-facing pathway summaries [85–87].

Diagnostic tests are commonly evaluated using two parameters: sensitivity and specificity, defined respectively as the probability of a positive test in the presence of disease, and the probability of a negative test in the absence of disease [88]. The ideal test is both highly sensitive and specific; it accurately identifies people with the disease without giving false positives when testing healthy patients, however, in most cases a highly sensitive test is not also highly specific, and so balancing these parameters is a key challenge in the field of medicine [89]. Predictive values are often more directly useful and interpretable as the probability of a patient having the disease once test results are known, but are explicitly dependent on the prevalence of the disease [89]. The positive and negative predictive values (PPV and NPV respectively) can be calculated as follows:

$$\mathrm{PPV} = \frac{\mathrm{Se} \cdot \mathrm{Prev}}{\mathrm{Se} \cdot \mathrm{Prev} + (1 - \mathrm{Sp})(1 - \mathrm{Prev})} \tag{1.1}$$

$$\mathrm{NPV} = \frac{\mathrm{Sp}(1 - \mathrm{Prev})}{\mathrm{Sp}(1 - \mathrm{Prev}) + (1 - \mathrm{Se})\mathrm{Prev}} \tag{1.2}$$

Where Se is sensitivity, Sp is specificity, and Prev is the prevalence of the target condition in the tested population, with a value of 1 indicating perfect sensitivity/specificity (i.e. all people with/without the disease are correctly identified).

In prostate cancer, the target condition (i.e. what is trying to be detected/predicted) is frequently "clinically significant prostate cancer" (csPCa), however the operational definition for this varies by study, and materially shifts reported performance. Because both imaging triage (mpMRI) and biopsy decisions are frequently evaluated against csPCa thresholds defined in terms of Gleason score/Grade Group, the endpoint definition should be treated as part of the test specification rather than a universal constant. Under the Patient-Reported Outcomes Measurement Information System (PROMIS) framework, for example, csPCa is commonly defined either stringently (University College London definition 1: Gleason grade $\geq 4 + 3$ and/or maximum cancer core length $\geq 6\,\mathrm{mm}$), or less stringently (University College London definition 2: Gleason grade $\geq 3 + 4$ and/or maximum cancer core length $\geq 4\,\mathrm{mm}$), with many contemporary guidelines also using Gleason Grade Group ≥ 2 (discussed in Section 1.2.3) as an operational threshold to define csPCa [86, 90–92].

The PSA test is widely used as an initial triage because it is cheap, accessible, non-invasive, and sensitive to prostatic pathology; however, it is not cancer-specific, and can be influenced by age, non-cancerous conditions such as benign prostatic hyperplasia (BPH) and inflammation/infection, sexual activity, medications, lifestyle, and other factors [61, 93–95]. Additionally, PSA levels may not be elevated sufficiently in certain presentations of PCa [96, 97]. Consistent with this, meta-analytic summaries emphasise high sensitivity at commonly used referral thresholds, but poor specificity, with substantial heterogeneity across studies and age strata [98, 99]. NICE evidence synthesis for symptomatic/referral contexts reports that at PSA cut-offs near typical referral thresholds (approximately $3\,\mathrm{ng\,ml^{-1}}$ to $4\,\mathrm{ng\,ml^{-1}}$), pooled sensitivity is high ($\sim 0.92 - 0.93$) whilst specificity is low (~ 0.16); increasing the threshold trades sensitivity for specificity (at a threshold of $8\,\mathrm{ng\,ml^{-1}}$, sensitivity is ~ 0.50, and specificity ~ 0.73), with evidence quality and disease heterogeneity explicitly limiting the precision of any global single-number summary [85]. In a specific set of studies carried out by Thompson et al. (2004, 2005), it was found that approximately 15 % of patients with a normal PSA level were found to have cancer, and around 75 % of patients with elevated PSA did not have cancer [100, 101]. As such, a practical implication is that PSA is best described as the "first line of defence" test, used only to

inform decisions on further testing, rather than a definitive discriminator of malignant biology.

DRE (which is now offered as an optional additional test, rather than forming part of the official diagnosis pathway) can identify some posterior palpable abnormalities, and may strengthen the indication for urgent referral, but is limited by anatomy (as only the posterior aspect can be directly assessed, as illustrated in Figure 1.1(a)), and by examiner dependence [24, 53]. Systematic reviews report strongly context-dependent performance: in primary care screening-like settings (such as forming part of a general health check), pooled sensitivity and specificity have been reported as approximately 0.51 and 0.59 respectively (very-low certainty), whereas in symptomatic case-finding contexts (i.e. where prostate cancer is already suspected), pooled sensitivity can be substantially lower (approximately 28.6 %), and specificity higher (approximately 90.7 %), implying that a normal DRE does not reliably exclude disease, but an abnormal DRE can meaningfully increase post-test predictive probability [102, 103].

Two primary implications follow this discussion: first, benign enlargement and inflammation can elevate PSA in the absence of malignancy, contributing to low specificity of the PSA test in primary-care triage [26, 93, 94]. Second, the posterior predominance of many cancers provides a mechanistic reason why DRE is able to identify some clinically relevant lesions whilst systematically missing others, and why contemporary pathways favour mpMRI for spatially resolved assessment prior to biopsy [85, 86].

According to current UK guidance, mpMRI is positioned as the first line investigation in specialist assessment of suspected localised disease, typically reported using a 5-point Likert-style suspicion score, with MRI-influenced biopsy decisions forming additional downstream analysis [35, 86, 104]. In the PROMIS multicentre UK study (with 11 NHS hospitals taking part), using template prostate mapping biopsy as the reference standard and a stringent csPCa definition (Gleason grade $\geq 4+3$ and/or maximum cancer core length ≥ 6 mm), mpMRI achieved a sensitivity of 93 % (95% confidence interval (CI) 88 % to 96 %), specificity of 41 % (95% CI 36 % to 46 %), PPV of 51 % (95% CI 46 % to 56 %), and NPV of 89 % (95% CI 83 % to 94 %) for csPCa [91]. In the context of their study, mpMRI was used to provide an estimate of the likelihood of clinically significant disease by comparing the findings to the reference standard of template prostate mapping biopsy. Definitive grading of tumours (i.e. assigning a Gleason grade) was based on histological examination of biopsy tissue and cannot be achieved through imaging techniques alone. Meta-analytic guideline panel evidence in pre-biopsy cohorts corroborates that NPV is typically high but heterogeneous, reporting a median mpMRI NPV 88.1 % (IQR 85.7 % to 92.3 %) for csPCa defined as Gleason score ≥ 7, with explicit prevalence dependence [105]. NICE guidance also emphasises residual false negative risk in MRI-negative pathways, and recommends shared decision-making when a biopsy is omitted following low-suspicion mpMRI [86]. As such, mpMRI may be described as a strong triage test upon which to base decisions of further testing via biopsy/histopathology, rather than as a definitive test suitable to diagnose csPCa in and of itself [106, 107].

Biopsy with histopathology remains the confirmatory "gold standard" within the diagnostic pathway and uses the Gleason grading system to stratify cancer aggressiveness (a diagram of the grading system can be seen in Figure 1.3), however, it is not a perfect reference standard in practice, as detection depends on sampling strategy (systematic transrectal ultrasound (TRUS) guided vs. targeted (typically via mpMRI) vs. template mapping (where a much greater number of samples are collected from each area of the prostate following ultrasound mapping)), and classification depends on grading reproducibility,

which can vary between observers/pathologists [65, 108–110]; even Donald Gleason, the creator of the Gleason grading system, was only able to replicate scores on re-examination approximately 50 % of the time (though it must be said that significant advancement has been made since then) [62]. PROMIS provides directly comparable estimates for standard TRUS guided biopsy against a mapping-biopsy reference: for csPCa, TRUS sensitivity was 48 % (95% CI 42 % to 55 %) with specificity of 96 % (95% CI 94 % to 98 %), PPV of 90 % (95% CI 83 % to 94 %), and NPV of 74 % (95% CI 69 % to 78 %) [91]. Importantly, while a biopsy is the best and most definitive method of diagnosing csPCa (purely in terms of accuracy/diagnostic power), the decision on whether to carry out a biopsy is weighted against potential harm; infectious complications including sepsis are explicitly discussed in guideline materials, and certain presentations of PCa (such as cribriform morphology and intraductal carcinoma) can be missed by biopsy, and as such it must be balanced against the risk of overdiagnosis (i.e. diagnosing cancer in cases where there is likely to be minimal impact on a person's quality of life within their lifetime), thus subjecting them to unnecessary invasive procedures, and complemented with additional methods of investigation to minimise the chance of leaving someone undiagnosed [28, 85, 86, 111].

Machine-learning approaches are increasingly explored for mpMRI interpretation and for assisting histopathology workflows, with the aim of improving efficiency and reducing variability between clinicians; however, performance remains contingent on dataset composition, external validation, and deployment context, and current evidence supports the use of AI as a tool that improves both the speed and accuracy of reporting to be used by specialists alongside other more traditional methods of interpretation, rather than as a direct replacement for expert assessment [112–118].

Active surveillance is a management strategy for people with apparently low-risk, localised prostate cancer in which treatment is deferred, and the disease is instead monitored by scheduled clinical review and repeat testing (commonly serial PSA testing, with mpMRI and repeat biopsy when indicated), with the shift to beginning definitive treatment being triggered by evidence of disease progression or reclassification to a higher Grade Group [48, 119, 120]. The clinical rationale is to mitigate overtreatment harms arising from the high prevalence of indolent, low-grade disease, and from upstream non-specific triage that increases overdiagnosis (most notably as a result of PSA testing), whilst preserving the option of timely intervention for cancers that declare more aggressive behaviour on follow-up [24, 39, 52]. An important distinction is that active surveillance is not equivalent to no treatment; however, it depends on adherence to surveillance schedules, and remains constrained by imperfect reference testing (biopsy sampling and grading uncertainty), such that a subset of patients initially classified as low grade may later be upgraded on subsequent assessment [49, 119].

Taken together, the prostate cancer diagnostic pathway in the UK increasingly leverages mpMRI to reduce unnecessary biopsies whilst retaining sensitivity to csPCa, but is unfortunately still not perfect. Upstream tests are non-specific, imaging performance is endpoint (i.e. what definition of csPCa is used) and prevalence (i.e. the proportion of people having csPCa in the population studied, which affects the predictive values) dependent, and even "gold standard" tissue sampling is itself imperfect due to the invasive nature of the procedure, as well as sampling and grading variability. This motivates undertaking further research to identify complementary biological characterisation layers that can identify key markers of the malignant/disease state more directly in standardised, controlled systems, like that which is carried out in this study.

1.4 Why analysis at the cell/organelle scale is complementary

The clinical diagnostic pathway for prostate cancer is intrinsically multi-modal, combining biochemical triage (PSA), clinical examination (DRE), imaging (mpMRI), and tissue diagnosis (biopsy with histopathology) [35, 85–87]. The diagnostic performance of each modality is conditional on cohort selection, endpoint definition, and clinical decision thresholds (see Section 1.3 for a more detailed discussion). This section motivates why cell and organelle scale characterisation should be treated as complementary to this pathway: it targets biological structure and state variables that are not directly accessible from PSA/mpMRI alone.

1.4.1 Clinical motivation: improving discrimination of consequential disease

A central challenge in diagnosing and treating prostate cancer is that detection is not synonymous with clinical consequence; some cancers are indolent and may be managed conservatively (using strategies such as active surveillance), whereas others may progress into more serious/aggressive disease, and require timely intervention [120]. In practice, the combination of imperfect upstream specificity and uncertainty in downstream risk stratification contributes to unnecessary biopsies, overdiagnosis, and overtreatment in some patients, motivating the search for complementary biomarkers that can provide additional biological information beyond PSA and mpMRI [48, 120]. Such biomarkers would be intended to support existing diagnostic investigations to form "risk-adapted" strategies, for example by aiding in discrimination between benign and malignant disease where conventional investigations are inconclusive [121].

Within this broader clinical context, quantitative molecular, cellular, and subcellular descriptors are frequently investigated as candidate biomarkers that may provide complementary biological information alongside established diagnostic modalities [122–124]. The present study examines whether reproducible nuclear and nucleolar descriptors can distinguish malignant and non-malignant prostate cell lines under standardised experimental conditions. Such descriptors may be able to provide additional discriminatory information to support clinical decision-making, however, translating any descriptor into clinical practice requires independent validation in patient cohorts, and is beyond the scope of this thesis [120, 121, 123].

1.4.2 Scientific motivation: mechanistic characterisation of malignant state in controlled models

In prostate pathology, malignant transformation is not recognised solely by glandular architecture; many other methods including examination of cytological features (i.e. the microscopic characteristics of individual cells) are also long-standing diagnostic criteria by which cancer may be diagnosed [125–129]. Lima et al. (2018), for example, used metabolomic profiling of cell culture media to show that volatile metabolites from all four prostate cancer lines (LNCaP, DU145, PC-3, 22Rv1) clustered separately from the normal PNT2 line [130]. They used principal component analysis and discriminant modelling to prove that metabolites can successfully discriminate PCa cell lines from normal prostate cell lines, and could even identify androgen sensitive vs. insensitive cell lines, and aggressive vs. more indolent cancers. Relevant to this imaging-based study, nuclear atypia/pleomorphism refers to deviations in nuclear size and shape, and alterations in chromatin patterning relative to healthy epithelium; these features are used routinely (typically qualitatively) in diagnostic assessment [63, 72, 131]. Significant

work has been done in terms of quantitative nuclear morphometry (size, shape, and chromatin texture measures) in prostate-specific literature, with evidence that nuclear descriptor variance can carry prognostic information in tissue cohorts when evaluated rigorously [132–135]. This provides a principled rationale within the context of this thesis; nuclear morphology and chromatin organisation are not arbitrary image features, but are instead clinically meaningful, and can be leveraged as measurable descriptors in vitro.

Similarly, the nucleolus is a major site of ribosome biogenesis (including rRNA transcription and processing), and increased nucleolar size/number and nucleolar prominence are widely recognised features of cancer cells, most commonly interpreted as reflecting elevated biosynthetic demand and altered growth control signalling [73, 136–139]. In prostate pathology specifically, conspicuous ("prominent") nucleoli are repeatedly emphasised as diagnostically useful cytological cues [18, 63, 72, 140]. Thus, quantifying nucleolar state in model systems can be argued as an attempt to operationalise a known cytological hallmark, while recognising that fluorescence based nucleolar measurements remain conditional on marker choice and preparation. A concrete prostate-specific mechanistic precedent links nucleolar structural changes to oncogenic signalling: Koh et al. (2011) report that the oncogene MYC drives nucleolar alterations in prostatic neoplasia, and that nucleolar biogenesis factors (including fibrillarin) lie within this axis [17]. While this study is mechanistically informative, it does not by itself justify causal inference from fluorescence intensity measurements alone; however, it does provide a biologically coherent justification for treating the nucleolar state as a malignancy-associated feature in prostate cancer models, suitable for subsequent quantitative characterisation [18, 138, 141].

The core argument for the type of analysis performed in this study is therefore complementary rather than substitutive; clinical pathways integrate PSA, imaging, and histopathology to make decisions under uncertainty, while controlled cell line experiments are able to quantify nucleus and nucleolus level changes to characterise the malignant vs. non-malignant state with high internal consistency. The present thesis is explicitly limited in scope to comparative characterisation of two 2D immortalised prostate cell lines under explicitly defined preparation conditions, and while some interpretations are made about how this may map to the general PCa disease state, it does not assert that any measured descriptor maps directly to patient level prognosis, or constitutes a clinical decision tool [63, 132].

1.5 Biological models in prostate cancer

The objective of this thesis is to quantify nuclear and nucleolar fluorescence descriptors at the single-nucleus level under controlled preparation and imaging conditions. As such, biological models are introduced here only insofar as they justify the chosen experimental scope; 2D immortalised cell lines as a reliable model for reproducible fluorescence-based analysis. A comprehensive taxonomy of prostate cancer models is beyond the scope of this study, and so their discussion here is intentionally brief.

Laboratory models used to study PCa are many and varied, ranging from standard cell cultures, to complex animal models, but can largely be classified as in vitro (meaning "in glass", referring to studies performed on isolated cells outside the body), ex vivo (meaning "out of the living", referring to studies performed on tissue/organs which have been extracted from an organism), or in vivo (meaning "within the living", referring to studies performed inside a living organism) [142–145]. More specifically to the context of prostate cancer, in vitro models typically include standard 2D cell line studies, as well as

more complex 3D culture such as spheroids, organoids, and scaffold based models. Ex vivo models focus on cell samples from patients (including both primary cell cultures and tissue studies) or explant cultures. In vivo models use xenografts (which can be derived from patient tissue, or using standard immortalised cell lines) which are implanted into mice, as well as genetically engineered mouse models (GEMMs), in which the genes of mice are artificially manipulated to cause the spontaneous development of prostate cancer [144, 145].

Simple in vitro models are valuable as an initial step in testing biological hypotheses because they provide a highly controlled and experimentally tractable system [5, 144]. If a finding is reproduced in such simplified conditions, it can then be examined further in more complex models that better approximate the in vivo environment, such as 3D cultures, organoids, tissue explants, and animal models [145–147]. This study follows that logic by using a comparatively simple 2D cell line system to test whether nuclear and nucleolar fluorescence descriptors differ between malignant and non-malignant prostate cells before any broader interpretation is made.

Some of the key types of biological models used in prostate cancer research are illustrated in Figure 1.4, which is intended as a compact visual orientation to the model ladder which follows. Each panel corresponds to a distinct experimental "rung" that is commonly used in prostate cancer research: (a) 2D adherent cell culture (cell lines), (b) multicellular spheroids, (c) organoids, (d) tissue slices, (e) tumour-on-a-chip microphysiological systems, (f) xenograft models (cell line derived), (g) patient-derived xenograft (PDX) models, and (h) transgenic/genetically engineered mouse models (GEMMs) [144, 145, 148]. The figure is used only to support the following inference relevant to this thesis: as models move from (a) toward (h), biological context (generally) increases, but so does the complexity and difficulty of culture/sample preparation, reducing reproducibility, experimental controllability, and the certainty of interpretations (as the number of factors influencing observed changes increases substantially as systems increase in complexity) [5, 145, 146].

Stated explicitly, the main "rungs" of the model ladder relevant to the present study can be placed on a spectrum from models that are easiest to standardise and compare across repeats, to models that are harder to standardise (but increasingly resemble the in-body situation) [144–146]:

1. 2D immortalised cell lines
2. Primary cell cultures
3. 3D spheroids / organoids
4. Tissue slices / explant systems
5. Animal models (including xenografts, patient-derived xenografts, and genetically engineered mouse models)

Each model has distinct advantages and disadvantages; immortalised cell lines for example are easier to grow/work with and provide a consistent system for research, though they lack a lot of the complexity and heterogeneity of primary cell cultures, and even more so tumours [5]. Similarly, primary cultures and 3D models can better preserve aspects of tissue-like behaviour and heterogeneity, but are more challenging to standardise across batches, and often introduce additional sampling complexity and challenges in accurately segmenting images (i.e. identifying and labelling an individual nucleus from a field of view containing multiple nuclei for quantitative imaging) [5, 146, 150, 151]. Animal models

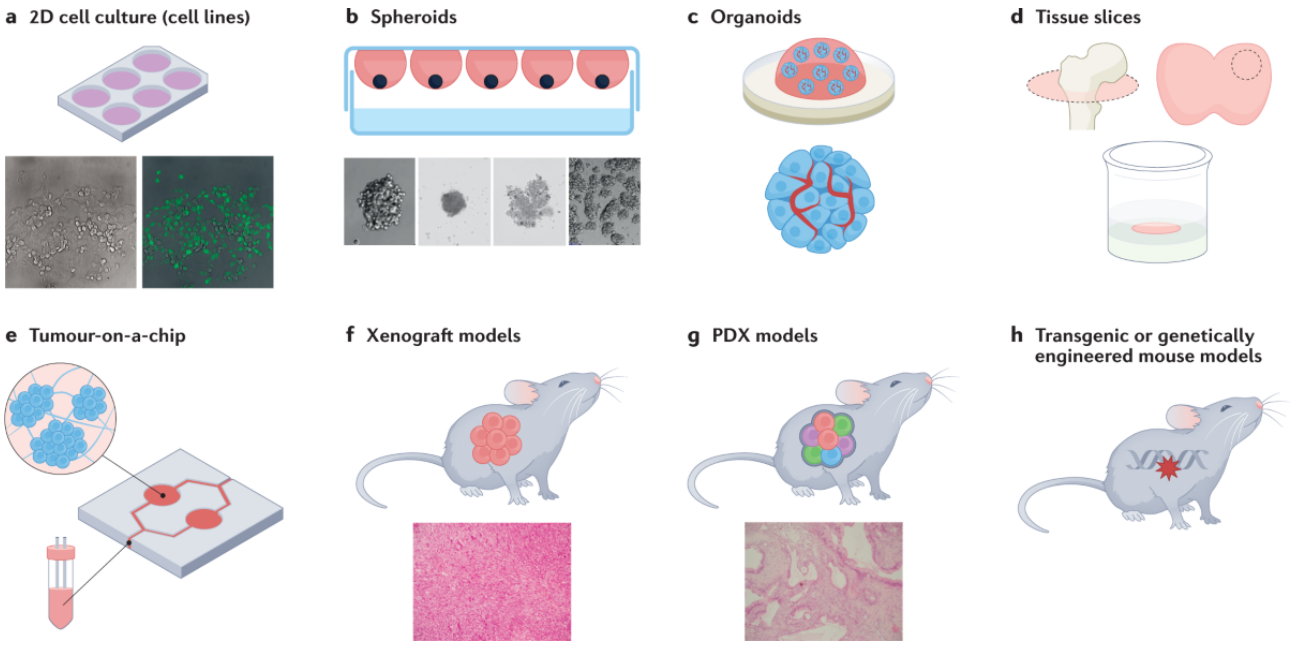


Figure 1.4: Diagram of representative biological model types used in prostate cancer research created by Sailer et al. (2023) [144]. A practical "models ladder" can be drawn from simpler, more controlled 2D cell cultures through higher-order in vitro and finally to more complex in vivo systems. (a) 2D cell culture (cell lines): immortalised prostate-derived cells grown as adherent monolayers, enabling high-throughput and tightly standardised preparation and imaging. (b) Spheroids: 3D multicellular aggregates formed from cell lines or primary cells, providing greater cell-cell contact and diffusion gradients largely absent in 2D. (c) Organoids: self-organising 3D epithelial structures derived from primary tissue or stem cells (often embedded in an extracellular-matrix-like gel), which can partially recapitulate tissue architecture and differentiation programmes. (d) Tissue slices: thin ex vivo prostate tissue sections maintained in culture for short-term study, preserving the native multicellular composition and spatial organisation of the tumour environment. (e) Tumour-on-a-chip: microengineered culture platforms that incorporate controlled flow and compartmentalisation to model tissue interfaces and transport. (f) Xenografts: implantation of human prostate cancer cell lines into immunodeficient mice to study tumour growth in a living host. (g) Patient-derived xenografts (PDX): transplantation of intact patient tumour tissue into immunodeficient mice, aiming to better preserve tumour heterogeneity and architecture compared to cell line xenografts. (h) Genetically engineered mouse models (GEMMs): mice with specific genetic alterations that drive spontaneous prostate tumour development. [144, 145, 148, 149].

further increase physiological context, but introduce additional variability between subjects, and are much more challenging to establish and leverage for high throughput applications [145, 146].

Given the trade-offs outlined above, 2D immortalised cell lines provide the most tractable platform for the central measurement goals of this thesis: high throughput fluorescence image acquisition across many fields of view and conditions, large per-condition nucleus counts to support distributional analysis and convergence testing, and segmentation and quantitative metric extraction that can be applied reproducibly and compared meaningfully across experimental batches [5, 144]. In contrast, higher-complexity models further down the models ladder more often constrain achievable nucleus counts and introduce additional sources of biological and technical variability (for example denser packing, mixed cell populations, and more structured backgrounds), which can complicate automated segmentation and reduce sensitivity to small between-condition effects [147, 152].

2D cell lines nevertheless have well-recognised limitations as biological surrogates: they do not reproduce the full tumour microenvironment, can adapt to culture conditions over time, and may exhibit passage-associated drift in both genetic composition and observable cell-state characteristics [5, 7, 146, 153]. As such, observations in this study are interpreted as model-specific quantitative descriptors obtained under controlled protocols, rather than as direct surrogates for malignant behaviour in patient tumours or for the full clinical disease ecosystem [7, 146]. Despite these limitations, the

stated advantages of 2D cell lines make them appropriate for nuclear/nucleolar state analysis in the present workflow [5, 144]. To avoid ambiguity in subsequent interpretation, it is therefore necessary to define what specifically is meant by a "cell line" in this thesis, and to motivate why intra-sample heterogeneity is treated as a biological signal and analysis target, rather than merely as measurement noise.

1.6 Cell lines and single-nucleus heterogeneity

Here, a "2D immortalised cell line" denotes a continuously proliferating human cell population, maintained as an adherent monolayer on planar substrates, with effectively indefinite replicative capacity enabled by tumourigenic origin and/or specifically engineered immortalisation [5, 144, 154]. This model choice provides several key advantages, including the ability to proliferate indefinitely in culture, allowing the replication of work and verification of findings by using a standardised set of samples, increased experimental controllability (with clearly defined media, perturbations, batch structure, and reduced heterogeneity when compared to primary cell cultures making them more suitable for mechanistic studies), ease of scalability, and measurement alignment for image-based nuclear characterisation (consistent morphology, tractable segmentation) [5, 57, 144, 153, 155]. Immortalised cell lines can also be used to model extremely specific types/presentations of disease, and can be modified to suit specific needs over long term culture (for example to study the development of drug resistance) [146, 156–158].

Two limitations, however, are particularly relevant to the present thesis. First, cell lines do not reproduce the tumour microenvironment and its coupled biochemical and mechanical signalling; therefore, observed nuclear states are interpreted as "model-specific" analytical measures, rather than direct surrogates for any patient-level aggressiveness [7, 146]. Second, long-term culture introduces evolutionary drift under selection; cells can adapt to culture conditions, and so may exhibit both phenotypical and genotypical changes when compared to cells which are at a lower passage number. As such, provenance (i.e. logging the source, passage number, handling methods and frequency etc) is a key part of the experimental specification, and the effects of this drift are mitigated by implementing strict passage number limits [153, 155].

Immortalised cell lines are essential tools that have played a significant role in prostate cancer research to date [5, 56], and a wide range of prostate cancer and normal prostate epithelial lines are available to choose from. The most commonly used malignant models include LNCaP (an androgen-sensitive lymph node metastasis), PC-3 (an androgen-independent bone metastasis), and DU145 (an androgen-independent brain metastasis), as well as their various clones and derivatives, and common benign/non-malignant comparators include RWPE-1 and PNT-derived lines, as well as BPH-1 (derived from a benign prostatic hyperplasia) [55, 56, 145, 149, 159–162]. This thesis focuses on the specific pairing of LNCaP and PNT2-C2 as a controlled malignant vs non-malignant model within a standardised 2D imaging pipeline, enabling direct, preparation-matched comparisons of nuclear and nucleolar descriptors [5, 7].

1.6.1 Cell lines used in this thesis: LNCaP and PNT2-C2

LNCaP is a malignant prostate cancer model derived from metastatic disease [1]. The LNCaP clone FGC line was established in 1977 from a lymph node metastasis from a 50-year-old white male patient

with prostate carcinoma, and is known to exhibit chromosomal instability and non-diploid karyotypic structure; this combined with its androgen receptor (AR) signalling and prostate-associated marker expression (including PSA) makes it a widely used model for androgen-regulated prostate cancer biology [56, 155, 163–165].

PNT2-C2 is an immortalised, non-malignant prostate epithelial cell line established in 1995, derived from a 33-year-old male exhibiting normal prostate function/morphology, and was immortalised by SV40 transfection; it is a PNT2-derived clone used widely in prostate in vitro work [3, 4, 162, 166–168]. Studies using PNT2-C2 in 3D Matrigel report epithelial differentiation-associated features and spheroid formation behaviour consistent with normal prostate epithelium under defined conditions, supporting its use as a reproducible baseline context for imaging-based assessment of nuclear descriptors [151].

A "non-malignant comparator" (PNT2-C2 in this case) is a cell line used as a baseline reference for nucleus and cell level observations under matched culture, preparation, and imaging conditions. Importantly, this is not necessarily synonymous with primary normal prostate epithelium; immortalisation, especially by SV40 transfection can modify DNA damage response and cell-cycle regulation by overriding p53 and pRB dependent checkpoint biology control, which constrains how strongly "normality" can be inferred from comparator measurements [3, 166, 169, 170]. Despite this caveat, PNT2-C2 is well established (along with its parent line PNT2 and related PNT1A) within the wider prostate cancer literature, and as such there is considerable precedent justifying its use as a normal equivalent in cell line studies [162, 167, 168, 171].

1.6.2 Single-nucleus heterogeneity as a primary analysis target

This thesis treats intra-sample heterogeneity as an empirically meaningful feature of prostate cancer models. "Single-nucleus heterogeneity" denotes the observed distribution of nuclear states across nuclei within a given cell line and condition, reflecting contributions from (at minimum) cell-cycle mixtures, chromatin state variability, clonal/subclonal diversity, and the natural and expected variations inherent to biological work, rather than solely from imaging noise [129, 172, 173]. Specific to LNCaP, androgen-response and proliferation-state variability have been documented at the single-cell resolution, supporting the expectation that population-averaged reporting may obscure biologically relevant substructures [155, 172].

Methodologically, this motivates a distribution-first reporting for relevant metrics. Mean-only summaries can be insufficient when the distribution may contain a mixture of sub-states, or when comparatively rare nuclear configurations occupy the tails, and in such cases, two conditions can share similar means while differing materially in higher-order structures such as skewness, or in subpopulation fractions [129, 172–174]. As such, this thesis explicitly makes use of histogram and kernel density based analyses to characterise distributional shape and identify candidate subpopulations. A key interpretive safeguard employed here is trace-back: nuclei contributing to selected regions of the histogram (e.g. low, modal, high tails, and any detected secondary modes/subpopulations) are mapped back to representative images, enabling qualitative verification of whether distributional features correspond to coherent nuclear states, segmentation artefacts, or preparation induced outliers, and provides a reproducible link between quantitative summaries and underlying image evidence.

Finally, the representational constraints of historical prostate cell line resources must be stated explicitly. Widely used prostate lines are derived from a limited set of donors, and recorded demographic labels

can be incomplete or not directly comparable across resources; though the majority of prostate cancer cell lines in use today overwhelmingly come from white donors, despite the far greater risk of developing the disease in black people [55, 56, 58]. Additionally, genetic ancestry analyses have previously demonstrated that cell line labels may not reliably encode genetic ancestry (for example, mixed ancestry structure has been reported in the commonly used prostate cancer line, 22Rv1, with explicit implications for the interpretation of preclinical findings) [57]. In terms of this work, the consequence is methodological rather than epidemiological; results are interpreted as model-specific comparisons (LNCaP vs PNT2-C2), and are not used to make population representative claims about prostate cancer biology across ancestry groups.

1.7 Cell biology fundamentals relevant to measured endpoints

This section is not intended to provide a general cell biology overview, rather it provides only the biological context necessary to understand the nuclear descriptors extracted later, namely size, shape, and DAPI intensity and texture.

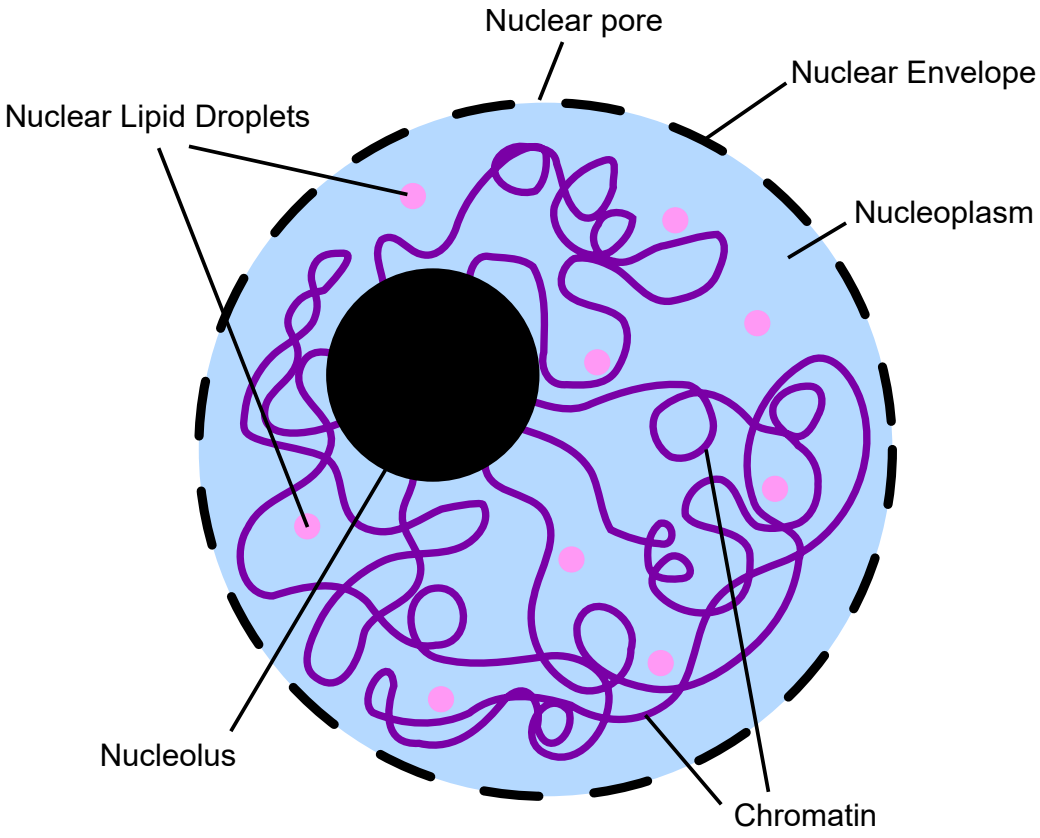


Figure 1.5: Schematic nucleus diagram created in Inkscape [175]. The figure is included with the aim of orienting the reader to the nuclear boundary (used later for the purpose of segmentation) and subnuclear organisation, which can contribute to spatial heterogeneity of DNA dye signal in 2D fluorescence images. Key features of note are the large nucleolus and nuclear lipid droplets, both of which are implicated in prostate cancer biology [137, 176–178].

1.7.1 The nucleus as a cancer-relevant cellular feature

The nucleus is the largest organelle in eukaryotic cells, and arguably the most important [179]. It is a membrane bound organelle, that contains the majority of the cell's genetic material, serving as the "control centre" of the cell, as well as the site of hosts DNA replication, RNA synthesis, and transcriptional regulation [179, 180]. The nucleus is a complex organelle containing many smaller

organelles known as nuclear bodies. The nucleolus (shown in Figure 1.5) is of specific interest to this study, as it plays a crucial role in cell growth and division; processes which are clearly disrupted in the prostate (and indeed any) cancer state [17, 137, 181–183]. In many fluorescence microscopy studies, the nucleus is often treated as a convenient segmentation target, however, in prostate cancer, nuclear structure is itself clinically and biologically meaningful. Prostate histopathology studies often use nuclear atypia/pleomorphism (including nuclear enlargement and chromatin-pattern alterations) as part of the morphological landscape that differentiates healthy from malignant cells, with prominent nucleoli frequently emphasised as a useful cytological criterion [63, 72, 131]. More generally, malignancy-associated alterations in nuclear architecture and chromatin organisation are well documented across tumour types [183, 184].

A substantial prostate literature shows that quantitative nuclear morphometry and chromatin texture features derived from 2D imagery can reliably discriminate between non-malignant and malignant states in tissue, and correlate with progression/recurrence risk, supporting the premise that nuclear size/shape and chromatin-organisation proxies carry clinically relevant information [132, 185–188]. Image analysis frameworks such as the quantitative nuclear grade (QNG) used by Veltri et al. explicitly seek to formalise this idea by treating nuclear structural alterations as an analysable biomarker class [132, 189]. The present thesis does not make use of any tissue derived measures, rather, these precedents motivate a narrower claim appropriate for cell line imaging: nuclear morphometry and chromatin-associated texture are established cancer-associated measurement domains, and therefore merit distributional characterisation at single-nucleus resolution in prostate cancer models [125, 190–193].

Nuclear morphometry in this thesis is extracted from 2D fluorescence images, and as such, quantities such as nuclear area and aspect ratio are projected descriptors of a 3D object. These descriptors aim to define enlargement and shape irregularity in a reproducible manner, but do not uniquely determine nuclear volume, surface area, or 3D shape, as they are sensitive to orientation, focal plane, and nuclear height variability across cells. Morphometric comparisons are therefore interpreted as measurement conditional differences in projected nuclear state, rather than as direct volumetric claims. However, given that cell lines were treated in an identical manner, the effects of these on comparisons between cell lines are unlikely to be significant.

Several mechanisms can contribute to nucleus-level size and shape variability in culture. Cell cycle state, for example, is known to alter DNA content and can change nuclear geometry. Nuclei in G2/M phase, for example, typically contain $\sim 2\times$ the DNA content of G1 phase nuclei, and cell cycle dependent remodelling can modify chromatin compaction and nuclear morphology [193, 194]. Malignant transformation is also frequently associated with chromosomal instability and aneuploidy (a condition in which a cell possesses an abnormal number of chromosomes), which can broaden the distribution of nuclear DNA content and indirectly influence nuclear size and chromatin organisation [195, 196]. Further, nuclear shape is constrained by the nuclear envelope and lamina, and by the cell-mechanical context; tumour-associated perturbations of nuclear envelope composition and mechanics can also therefore contribute to irregular nuclear morphology [197]. These factors also motivate a distributional analysis strategy and suggest treating observed heterogeneity as a candidate biological, signal that must be checked against technical artefacts before being relied upon for biological interpretation.

Lipid droplets, including nuclear lipid droplets (indicated in pink in Figure 1.5), have also been reported in multiple cancer contexts, including prostate cancer, and have been discussed in relation

to cancer-associated lipid metabolism as sources of energy and raw materials for the creation of cell membranes [176, 198, 199]. In this thesis, lipids were not measured, and are only discussed here to acknowledge that additional subnuclear structures which are implicated in the disease state and may be visible in some imaging modalities, without necessarily implying that they are quantified here.

1.7.2 DAPI-based nuclear descriptors

4',6-diamidino-2-phenylindole (DAPI) is a DNA-selective fluorescent dye with a preference for minor-groove binding in AT-rich regions [8, 200, 201]. In fixed-cell fluorescence microscopy, DAPI is a widely used nuclear counterstain, and under controlled conditions, provides a practical readout of nuclear DNA associated structure [9, 202]. In this study, this is used for nuclear segmentation, as well as for the extraction of morphometric and intensity based descriptors from within the segmented nuclear region.

Under ideal conditions (uniform staining, linear detector response, and identical preparation methods), measured DAPI intensity can scale with DNA content, motivating its widespread use for cell cycle profiling and ploidy estimation [9, 203, 204]. However, DAPI fluorescence also depends on dye accessibility and the local DNA environment, and as such, changes in DAPI intensity density (i.e. intensity per unit area) and texture can reflect chromatin condensation/organisational state in addition to (or instead of) changes in DNA content [205, 206]. More explicitly, a shift towards brighter and/or more spatially heterogeneous DAPI signal is consistent with increased chromatin condensation or altered chromatin organisation, but does not uniquely establish any specific molecular mechanism without orthogonal calibration/validation [207, 208].

Practical examples illustrating this non-uniqueness include both apoptosis and mitosis; DAPI-bright, morphologically compact nuclei can arise at both extremes, during both apoptotic (indicative of cell death) and mitotic (during active cell division/proliferation) chromatin condensation, meaning that increased intensity can coincide with reduced nuclear size in a subpopulation [191, 209–211]. This further motivates two interpretive safeguards adopted throughout the thesis: firstly, distributional rather than mean-only analysis, to avoid masking any significant mixture-like populations, and secondly, tracing back to individual nuclei from histogram regions in order to distinguish coherent nuclear states/measurements from segmentation artefacts.

Importantly, DAPI-derived descriptors are conditional on sample preparation and imaging settings. Key technical contributors include fixation chemistry and duration, staining concentration and incubation time, wash efficiency, illumination/exposure settings, photobleaching and potential exposure-dependent photochemical changes, detector saturation, and segmentation choices (such as thresholding and boundary refinement) [8, 9, 190, 202]. In widefield (epifluorescence) imaging, the recorded intensity also integrates signal along the optical axis; thus internuclear variability in height, focus position, and out-of-plane blur can also feasibly perturb both DAPI intensity and texture metrics, even when the projected area is similar. For these reasons, comparisons throughout this study are framed as most interpretable within matched staining and acquisition settings, with additional preparation dependent effects (most notably detergent permeabilisation) explicitly discussed.

1.7.3 Nucleolar state and fibrillarin as a nucleolar marker

As previously stated (and shown in Figure 1.5), the nucleolus is the most prominent subnuclear compartment in many mammalian cells, and functions as the major hub of ribosome biogenesis, encompassing rRNA transcription, processing, and early ribosomal subunit assembly [179, 212, 213]. It is therefore biologically reasonable to treat nucleolar state as a growth/proliferation associated feature; increased ribosomal biogenesis capacity is a common and well established requirement for sustained proliferation, and nucleolar enlargement and altered nucleolar organisation are widely reported across cancers, prostate and otherwise [136–138, 141, 214]. However, nucleolar morphology and composition are also perturbed by stress and DNA-damage signalling (the "nucleolar stress" framework), meaning that nucleolar changes are best discussed within this context, rather than solely treated as unique signatures of increased ribosome output [183, 213, 215, 216].

In prostate histopathology, conspicuous nucleoli (often described as prominent or macronucleoli) are repeatedly emphasised as diagnostically useful cytological features in invasive carcinoma and high grade prostatic intraepithelial neoplasia (PIN), albeit with non-uniform operational definitions [63, 72, 217]. Importantly, attempts to objectify nucleolar features in prostate specimens date back to the early 1980s, including studies that quantify nucleolar size/area as a prognostic variable and use nucleolar number/location as a histological biomarker in diagnostically challenging biopsy material [73, 140]. This provides significant historical precedent motivating the present fluorescence framing; nucleolar quantification as used here attempts to operationalise a well established cytological characteristic for use at the single-cell level in commonly used PCa model systems.

The nucleolus is commonly described as comprising nested subcompartments, often referred to as fibrillar centres (FC), the dense fibrillar component (DFC), and granular component (GC), that each relate to distinct stages of rRNA transcription and processing [213, 218]. Fibrillarin (FIB) is a highly conserved nucleolar protein, whose localisation to the DFC has been demonstrated using a variety of approaches, and is widely used in fluorescence microscopy workflows as a nucleolar marker [17, 218–221]. Mechanistically, fibrillarin is a core component of box C/D small nucleolar ribonucleoprotein particles (snoRNP) complexes, and is involved in rRNA processing and 2'-O-methylation associated steps in ribosome biogenesis, making FIB labelling a plausible method of monitoring the nucleolar activity state [218]. Nevertheless, as with the DAPI-based descriptors discussed previously, fluorescence intensity remains an operational proxy without explicit calibration.

A concrete mechanistic rationale for analysing the nucleolar state in prostate cancer models is provided by previous work linking oncogenic signalling to nucleolar remodelling. Koh et al. (2011) report that MYC, a well known regulator of cellular proliferation in both benign and malignant contexts, drives nucleolar structural changes in prostatic neoplasia, including increases in nucleolar number and size, and identify fibrillarin as an MYC-associated nucleolar factor within this axis [17]. In the same study, fibrillarin perturbation was reported to markedly impair the proliferative capacity and clonogenic survival of multiple prostate cancer cell lines (including LNCaP), consistent with nucleolar machinery being functionally engaged in malignant cell growth [17]. Complementary review synthesis by Brown et al. (2022)emphasises MYC as a key regulator of nucleolar activity via enhanced rDNA and ribosomal gene transcription, supporting nucleolar enlargement/activation as part of an active growth programme, rather than a purely passive morphological change [214]. Furthermore, tissue evidence from prostate cancer studies indicates that rRNA overexpression is common, consistent with

altered ribosomal biogenesis demand in prostate tumour biology [222]. These sources collectively serve to justify measuring the nucleolar state as a biologically coherent malignancy associated feature at the cell line level.

Nucleolar targeting typically requires immunofluorescence labelling, and access of antibodies to intranuclear epitopes generally requires chemical permeabilisation of fixed cells (or complex, expensive, and time-consuming genetic modification to induce expression of fluorescent proteins) [11, 12, 223]. Consequently, nucleolar descriptors in this thesis are limited to a permeabilised preparation, with an equivalent condition in which cells were prepared in exactly the same manner, but stained with only DAPI serving as a control in order to separate detergent-dependent effects on nuclear descriptors from effects which may be attributed to the introduction of an additional stain. As with DAPI intensity, fibrillarin intensity and nucleolar segmentation are conditional on fixation, permeabilisation, staining concentration/incubation, imaging settings, and threshold/segmentation choices; fibrillarin-derived intensity metrics are therefore interpreted comparatively under matched protocols, and are not stated to serve as absolute measures of protein abundance or ribosome biogenesis rate [139, 221]. Nucleoli are also known to disassemble during mitosis; low or diffuse fibrillarin signal can thus reflect cell-cycle state, as well as stress induced redistribution. As for the DAPI based analysis, inspection of nucleolar descriptors at the distributional level and tracing data back to specific nuclei are employed as safeguards against over-interpreting the data and attributing segmentation artefacts to real biological changes [194, 215, 224].

In this thesis, nucleoli are defined for the purpose of image analysis as fibrillarin-positive subnuclear regions identified by intensity thresholding within each segmented nuclear mask. Although strictly speaking, fibrillarin labels the DFC rather than the entire nucleolar volume, utilising segmented fibrillarin-positive regions provides a reproducible, biologically grounded readout of nucleolar organisation, suitable for quantitative comparison, and aligns with several previous imaging-based studies which also use fibrillarin as a nucleolar marker [18, 139, 218, 219]. The principal nucleolar descriptors used later in this study are nucleus-level summaries including total nucleolar area and nucleolar intensity per nucleus. This mirrors established nucleolar imaging practice, where nucleolar/nuclear area ratios and intensity-derived summaries are routinely extracted from Hoechst/DAPI+FIB based segmentation workflows [139, 221].

Specific to PCa, prostate cell line studies provide a strong methodological precedent for using fibrillarin as a nucleolar feature source rather than purely as an illustrative stain. Batnasan et al. (2024) used fibrillarin immunofluorescence in prostate cancer cell lines (including LNCaP), and describe feature extraction using intensity values within segmentation masks alongside artefact filtering, reporting size and intensity measures for segmented objects in a high-content analysis pipeline [224]. Saffarian et al. report nucleolar localisation of fibrillarin in both cancerous and non-cancerous prostate cells, and describe observable changes in nucleolar morphology following nucleolar protein silencing, providing additional support for fibrillarin-labelled structure as a measurable axis of variation in prostate models [18]. These precedents support the claim that nucleolar descriptors derived from anti-fibrillarin imaging constitute a biologically motivated and technically established measurement family for prostate cancer cell-line analysis, with the explicit caveat that comparisons are conditioned on identical preparation methods.

1.8 Optical methods background and fluorescence-imaging literature

This thesis uses fluorescence microscopy to characterise nuclear and nucleolar states in fixed 2D prostate cell cultures. This section aims to introduce fluorescence microscopy at a basic/conceptual level sufficient to interpret DAPI and anti-fibrillarin images, while deferring detailed optical theory to Methods, and to synthesise, making use of a focused comparison table, how nuclear/nucleolar descriptors have been quantified in fluorescence-imaging studies prior to this work, as well as how frequently they report distributional analyses and explicit sample-size sufficiency arguments.

1.8.1 Context: other single cell measurement modalities

A variety of methods exist for the interrogation of biological state at the single cell level, including genomic/transcriptomic profiling (such as DNA/RNA sequencing and microarrays), proteomics/metabolomics (such as mass spectrometry and NMR spectroscopy) alongside a range of diverse imaging and spectroscopic techniques [225–227]. In the optical domain, label free vibrational spectroscopies such as Raman and Fourier-transform infrared (FTIR) microscopy can provide chemically informative "fingerprints" without requiring the use of any stains/additives, and have been widely used in the wider PCa literature [126, 228–230]. Notably, Raman based studies across multiple cell line systems (including prostate cancer) have treated statistical representativeness as an explicit methodological requirement, and perform convergence testing as spectra accumulate (e.g. the convergence of the mean spectrum, dispersion summaries, and standard error with an increasing number of spectra included) [127, 168, 231]. The convergence testing in these studies motivates, by analogy, the question addressed later in this thesis; when heterogeneity is the object of interest, how should the sample size be justified for fluorescence-derived nuclear and nucleolar descriptors? Raman/FTIR and other non-fluorescence modalities are not discussed in any detail further here and are excluded from the synthesis table, as they are not equivalent or directly comparable to the fluorescence based analysis carried out in this study.

1.8.2 Fluorescence microscopy for quantitative nuclear/nucleolar assessment

Fluorescence microscopy is routinely used to visualise specific biomolecules, organelles, or structures. Non-linear optical techniques (e.g. two-photon excitation, second-harmonic generation) have also been applied to prostate tissue to generate high-contrast images [232]. It works by detecting emitted light from fluorophores (fluorescent chemicals which light up/re-emit light upon excitation by absorbing a specific wavelength of light) that have been excited at a shorter wavelength [233]. This enables molecularly targeted visualisation by making use of either exogenous dyes which bind to specific biomolecular classes (DAPI, for example, as a DNA-binding nuclear stain), or fluorescent labelled affinity reagents such as antibodies (a process known as immunofluorescence, or IF) [234]. In this thesis, the DAPI channel provides a nuclear counterstain used both for segmentation and for extracting nuclear morphology/intensity/texture descriptors (Section 1.7.2), while IF using anti-fibrillarin provides a nucleolar-associated channel used to define and quantify fibrillarin-positive subnuclear regions (Section 1.7.3).

Fluorescence methods, offer enhanced molecular specificity over label-free modes of analysis: common nuclear stains such as DAPI or Hoechst dyes bind to DNA and reveal nuclear morphology, lipophilic dyes

like Nile red or BODIPY can be used to stain lipids and other non-polar molecules, while fluorescent antibodies or probes can target specific proteins, RNA or organelles [224, 233, 235–237]. Furthermore, techniques like fluorescence in situ hybridisation (FISH) allow specific DNA sequences to be visualised in interphase nuclei[238]. Celep et al. (2003) emphasise that "FISH is a powerful tool for quantitative analysis of chromosomes and genes" and can be applied to cell cultures, isolated nuclei or tissue sections [239]. In prostate cancer, FISH has been used to detect gene fusions (such as TMPRSS2-ERG) and chromosomal aberrations in both patient samples and cell lines [240].

The imaging in this thesis is performed using epifluorescence (widefield) microscopy, in which the full depth of the specimen is illuminated through the objective, and emitted fluorescence is collected through the same objective using a camera [233, 234]. Relative to confocal microscopy, epifluorescence does not use a pinhole to eliminate out of focus emission, and therefore provides less strict optical sectioning. However, it offers two key advantages for the present measurement objective: first, substantially higher throughput for image acquisition across many fields of view and conditions, and second, the ability to integrate fluorescence over the full axial extent (i.e. depth) of the specimen, rather than restricting measurement to a thin optical slice [234]. Although the present cells are cultured as adherent monolayers and are therefore predominantly two-dimensional, residual variation in local cell thickness and in the axial position of subnuclear structures remains, and so capturing data through the full sample depth reduces the risk that small focus offsets or variation in the growth plane will systematically exclude relevant signal, and thereby bias quantitative comparisons.

Immunofluorescence labels specific antigens/targets using antibodies, typically using a fluorophore-conjugated primary antibody or (more commonly, as was the case in this study) an unlabelled primary antibody which binds to the target structure alongside a fluorophore-conjugated secondary antibody which binds to the primary [11, 12, 241]. For intranuclear targets such as fibrillarin, antibody access generally requires permeabilisation of fixed cells to allow antibodies to cross the various cellular membranes (and, depending on fixation, access nuclear epitopes) [11, 12, 241]. This practical requirement motivates the design separation used throughout the thesis: a non-permeabilised DAPI-only baseline, a permeabilised DAPI-only control (to quantify the shifts in nuclear descriptors resulting from detergent permeabilisation), and a permeabilised DAPI+FIB condition (to quantify nucleolar characteristics).

Fluorescence-derived morphometric and intensity-based descriptors are conditional on preparation and acquisition [241]. Fixation chemistry and timing can modify antigen accessibility and molecular retention, and detergent permeabilisation can induce redistribution or loss of material; consequently, quantitative comparisons across preparation states are not assumed to be preparation-invariant and must be treated as empirical questions [11, 12, 16]. Similarly, intensity metrics depend on staining concentration/incubation, wash efficiency, illumination/exposure, detector linearity, and photophysical behaviour of fluorophores; DAPI-derived intensity is additionally sensitive to DNA environment and can exhibit artefacts under inappropriate exposure regimes (hence the emphasis on controlled acquisition and within-protocol comparison) [8, 9]. In multichannel fluorescence, spectral bleed-through/crosstalk and filter-band overlap can bias apparent intensities when channels are not orthogonal; these risks are managed by filter choice and experimental design, with more detail being provided in Methods [233, 234].

1.8.3 Fluorescence imaging synthesis table: what is commonly quantified/reported

Because this thesis treats cell-level heterogeneity as an explicit analysis target, two terms are thus defined explicitly for the remainder of the Introduction:

"Distributional analysis" denotes analysis in which the empirical distribution of a per-cell (or per-nucleus) descriptor is treated as an object of interest (e.g. tail behaviour, multimodality/mixtures, quantiles, modal estimates), rather than being collapsed immediately to a single mean/median.

"Convergence testing" denotes an explicit procedure to evaluate whether an estimate stabilises as sample size increases, typically via sequential accumulation, re-sampling, or bootstrapping.

Table 1.1 catalogues a set of fluorescence-imaging studies that quantify nuclear morphology/chromatin state descriptors and/or nucleolar descriptors. The table is not intended as an exhaustive review of fluorescence microscopy in cancer; rather, aims to elucidate the recurrent descriptor families used in general practice, and the frequency with which per-condition sample size is explicitly reported, and distributional structure is interrogated.

Study (year)	**Model / cell system**	**Fluorescence modality**	**Markers and quantified descriptor families (examples)**	**Sample size reporting**	**Dist.**
Multiparametric morphological profiling (high-content "Cell Painting"-style assays)					
Gustafsdottir et al. (2013) [242]	Fixed and permeabilised U2OS adherent cell cultures	High-content epifluorescence microscopy	Multiplex "painted" compartments; multiparametric morphology features across compartments	N not stated	No
Bray et al. (2016) [243]	Various immortalised and primary cell cultures (but mainly A549 and U2OS)	High-content epifluorescence microscopy	$\sim$ 1500 per-cell features relating to several organelles; per-well profiling via medians	N not stated	No
clue.io Connectopedia (2018) [244]	Documentation resource	N/A (processing convention)	$\sim$ 1700 computed features	Suggests $\sim$ 1000 cells per well	No
Bray et al. (2017) [245]	Fixed and permeabilised U2OS adherent cell cultures	High-content epifluorescence microscopy	Many single cell features	N not stated	No
Nuclear-focused pipelines (morphology, DNA content proxies, texture/topology)					
Narotamo et al. (2021) [246]	Multiple fixed and permeabilised immortalised adherent cell cultures	Epifluorescence microscopy	DAPI-based nuclear features for cell cycle staging	$\geq$ 3500 nuclei per condition	No

Study (year)	Model / cell system	Fluorescence modality	Markers and quantified descriptor families (examples)	Sample size reporting	Dist.
Mandelkow et al. (2017) [191]	Multiple permeabilised immortalised adherent cell cultures including LNCaP	Epifluorescence microscopy	Nuclear morphology/intensity in apoptosis including area, aspect ratio, and total intensity	400 to 1000 nuclei per condition	No
Gertych et al. (2010) [247]	Fixed and permeabilised DU145 prostate cancer cells	Z-stacks using confocal fluorescence microscopy	Various descriptors of spatial organisation of nuclear signals	496 to 660 cells per condition	Yes
Gertych et al. (2013) [248]	Fixed and permeabilised DU145 (prostate) and Huh-7 (liver) cancer cells	Z-stacks using confocal fluorescence microscopy	Various descriptors of spatial organisation of nuclear signals including mean intensity	$\sim$ 300 nuclei per condition	Yes
Rana et al. (2021) [249]	Fixed and permeabilised PC-3 prostate cancer cells (using a publicly available dataset)	Z-stacks using confocal fluorescence microscopy	Various 3D chromatin morphology/texture DAPI-derived descriptors	N not stated explicitly, estimated $\sim$ 250 PC3 per condition	Partial
Olofsson et al. (2021) [250]	Fixed and permeabilised 3D multicellular tumour spheroids (A498 HTB-44)	Z-stacks using confocal fluorescence microscopy	Various DAPI-derived measures including total intensity, nuclear volume, major/minor axis, and surface area	N not stated explicitly, estimated 300 cells per spheroid, 10 spheroids per condition	Yes
Nucleolar-marker assays and nucleolar stress imaging (fibrillarin/nucleolin and related)					
Kodiha et al. (2011) [139]	Fixed and permeabilised HeLa cervical cancer cells	Confocal fluorescence microscopy	Fibrillarin and B24 used to assess nucleoli with DAPI counterstaining; nucleolar vs nuclear intensity and nucleolar intensity/area ratios calculated	$\geq$ 20 cells per condition	Yes
He et al. (2018) [251]	Fixed and permeabilised A549 lung cancer cells	High-content epifluorescence microscopy	Fibrillarin and nucleolin used to assess nucleoli with DAPI counterstaining; various nucleolar morphology and intensity metrics calculated	N not stated explicitly, estimate $\gtrsim$ 500 nuclei per condition	No

Study (year)	Model / cell system	Fluorescence modality	Markers and quantified descriptor families (examples)	Sample size reporting	Dist.
Ogawa et al. (2021) [252]	Fixed and permeabilised MCF 10A fibrocystic breast epithelial cells	High-content epifluorescence microscopy	Fibrillarin used to assess nucleoli with Hoechst counterstaining for nuclei; nucleolar number per nucleus, % of nuclei with ≥ 5 nucleoli, nuclear area and total Hoechst intensity measured	281 to 1097 nuclei per condition	Yes
Potapova et al. (2023) [253]	Various fixed and permeabilised immortalised adherent cell cultures	High-content epifluorescence microscopy	Various nuclear and nucleolar stains used; with nucleolar area and intensity measured	N not stated	Partial
Jamison et al. (2010) [254]	Fixed and permeabilised T24 bladder cancer cells	Epifluorescence microscopy	Various nucleolar stains used with DAPI counterstaining; only qualitative discussion	N not stated	No
Sutton et al. (2021) [255]	Fixed and permeabilised A549 lung cancer cells	Epifluorescence microscopy	Various nucleolar stains used with DAPI counterstaining; NPM1 intensity CV per nucleus and qualitative nucleolar morphology reported	≥ 80 cells per condition	Yes
Snyers et al. (2022) [256]	HeLa and Hep3B cell lines	Mixed fluorescence microscopy including confocal, widefield, and fluorescence in situ hybridisation (FISH)	Various nucleolar stains used with DAPI counterstaining; nucleolar morphology and intensity metrics measured including nucleolar area, roundness, intensity, and number per nucleus	~ 64 cells per condition	Yes
Geisler et al. (2026) [257]	Multiple fixed and permeabilised immortalised adherent cell cultures	Confocal fluorescence microscopy	Various nucleolar stains used with DAPI counterstaining; nucleolar morphology and intensity metrics measured including nucleolar area, total nucleolar intensity per nucleus, and number per nucleus	≥ 100 cells per condition	Partial
Prostate-relevant fibrillarin imaging precedents (cell-line compatible)					

Study (year)	Model / cell system	Fluorescence modality	Markers and quantified descriptor families (examples)	Sample size reporting	Dist.
Batnasan et al. (2024) [224]	Fixed and permeabilised prostate cancer cell lines LNCaP, 22Rv1 and PC-3	Confocal fluorescence microscopy with additional z-stacks	Fibrillarin and NPM used to assess nucleoli with DAPI counterstaining; mostly qualitative nuclear/nucleolar analysis	300 to 1000 nuclei per condition	Partial
Saffarian et al. (2025) [18]	Multiple fixed and permeabilised prostate cell lines including LNCaP and PNT1B	Confocal fluorescence microscopy	Fibrillarin and NPM1 used to assess nucleoli with DAPI counterstaining; qualitative analysis in cell lines	N not stated	No

Table 1.1: Scoped synthesis of fluorescence imaging studies quantifying nuclear/nucleolar descriptors relevant to this thesis. "Dist." indicates whether distributional analysis is treated as an explicit analysis object. Where studies report imaging design (e.g. fields/wells imaged) but do not provide effective cell counts per-condition, this is flagged as "N not stated". Crucially, none of the studies listed above gave any indication/justification of sample size sufficiency.

Across fluorescence imaging studies that interrogate nuclear and nucleolar compartments, the most consistently utilised measures include segmentation-derived nuclear morphology (including area and shape descriptors), nuclear intensity summaries (such as total intensity per nucleus and intensity per unit area), and chromatin-associated texture features from nuclear stains, alongside additional nucleolar descriptors (including nucleolar number, area, and intensity-derived measures) [139, 224, 243, 246, 252].

Several recurring patterns from the studies summarised in Table 1.1 are directly relevant to the evidential strength of the heterogeneity claims made in this study. First, the sample size used per condition (in terms of cells or nuclei) is frequently left ambiguous (either not given, or only implied approximately via imaging design such as cell seeding density and fields imaged per well), limiting reproducibility and preventing direct quantification of sampling error for distributional estimates. Second, even when sufficient single cell features are quantified, downstream workflows commonly collapse these data to per-well summaries (such as median based profiles), which while scientifically appropriate for some screening objectives, may suppress complex and informative distributional structure when heterogeneity itself is of interest [242–244]. Third, explicit convergence/sample size sufficiency testing is uncommon in the scoped literature: sampling rules are often implicit (i.e. "image x number of fields per well" or "use all segmented cells"), and estimate stability is very rarely demonstrated. Finally, where permeabilisation is performed (i.e. in immunofluorescence workflows), rarely is a matched non-permeabilised comparison included to ground results and ascertain whether observations may arise from preparation artefacts (though for many of the studies detailed in Table 1.1, the only analytical target is an intracellular/nuclear protein/structure, and so in these cases it may not have been feasible for them to perform this matched analysis) [247, 249, 250]. These observations motivate a conservative novelty framing for the present thesis: single-nucleus resolution distributions are retained as primary analysis targets, permeabilisation is treated as an experimental factor to quantify preparation-dependent

shifts, and explicit sample-size sufficiency testing is carried out for fluorescence-derived nuclear and nucleolar descriptors under the specific prostate cell line imaging workflow used here.

1.9 Study positioning and novelty

Having established the clinical, biological, and methodological context, this section states more explicitly how the thesis is positioned relative to existing fluorescence-imaging practice. As stated throughout, this thesis is positioned as a scope-controlled fluorescence-imaging study of prostate cell lines, in which individual nuclei are the primary observation unit, and nucleoli are defined as fibrillarin-positive subnuclear regions within each nuclear mask. The Introduction has thus far established three constraints that govern interpretation: first, claims are made at the level of LNCaP vs PNT2-C2 under controlled protocols, rather than directly reflecting any differences in clinical cases (i.e. primary cell cultures or in situ disease in patients) [5, 7]; second, intensity-derived descriptors are treated as explicitly dependent on preparation and acquisition conditions, and serve as representative proxies without explicit calibration; and finally, permeabilisation is treated as an experimental factor, the effect of which on measured descriptors is quantified explicitly, rather than assumed negligible [11, 16].

1.9.1 Positioning relative to fluorescence-imaging practice

The fluorescence imaging synthesis in Table 1.1 highlights two methodological features that are common in nucleus/nucleolus imaging pipelines, and directly motivate the present study design. First, while nuclear and nucleolar segmentation-based descriptor families are widely used (including morphometry, intensity summaries, texture/topology, and marker-defined nucleolar metrics), per-condition effective sample size (N) is frequently not stated explicitly, or is only indirectly inferrable from imaging design [243, 253]. Second, even when single-cell measurements exist at scale, analysis pipelines often emphasise aggregated per-well summaries (for example, median-based profiling), rather than examining the distributional structure [243, 244]. As such, this thesis adopts a distribution-first analysis philosophy and makes N explicit across key descriptors in each condition, going even further to directly assess the heterogeneity inherent to each descriptor by assessing the standard error convergence with increasing N.

Explicit sufficiency evaluation is more commonly performed in some optical single cell modalities, particularly vibrational spectroscopy, where convergence of mean spectra and dispersion summaries with increasing N is treated as a core methodological requirement when heterogeneity is a target [127, 168, 231]. The present thesis does not claim equivalence between Raman spectra and fluorescence descriptors; rather, these precedents motivate the transfer of an explicit sufficiency philosophy to fluorescence-based descriptors, where N and stability are often left implicit in practice.

1.9.2 Novelty pillars

This thesis seeks to contribute new knowledge and methodological value via four explicitly defined novelty pillars:

1. **Convergence-validated single-nucleus fluorescence analysis.** Rather than adopting a fixed "typical" per-condition N, the thesis treats sample-size sufficiency as descriptor and condition-specific, and quantifies convergence behaviour empirically for nuclear and nucleolar

descriptors. This addresses a recurrent limitation in fluorescence practice, where effective N and stability are frequently not explicit (Table 1.1) [243, 253].

2. **Distributional analysis with explicit subpopulation inference.** The thesis retains per-nucleus distributions as primary outputs, and uses KDE and mixture-modelling representations to support identification and interpretation of potential subpopulations where warranted [258, 259]. This is motivated by established biological and technical sources of mixture structure in cancer cell populations (including cell cycle variations from damaged checkpoint biology, genomic instability, and preparation/segmentation variability), and is implemented with explicit traceback from distribution regions to individual nuclei shown in representative images to audit artefacts [174].

3. **Preparation sensitivity quantified explicitly (permeabilisation as a measured factor).** The thesis incorporates a permeabilised DAPI-only control to quantify detergent-dependent shifts in nuclear descriptors, enabling separation of preparation-induced effects from effects attributable to introducing an additional antibody-labelled channel. This addresses the methodological reality that permeabilisation can perturb retention, accessibility, and apparent intensity distributions, and therefore cannot be assumed to be neutral for intensity-derived descriptors [11, 16].

4. **Nucleolar measurement beyond size: fibrillarin-attributed signal density and enrichment-like ratios.** Nucleolar state is quantified using anti-fibrillarin in defined subnuclear regions, including not only nucleolar area and nucleolar occupancy ratio, but also intensity-derived descriptors such as nucleolar signal density and nucleus-level fibrillarin summaries (Section 1.7.3). This extends beyond the common restriction to nucleolar size/number, and aligns with mechanistic prostate cancer evidence linking oncogenic programmes to nucleolar remodelling and fibrillarin-associated features [17, 214].

In summary, this thesis aims to establish and evaluate a fluorescence microscopy workflow for comparative nuclear and nucleolar analysis in prostate cell lines, using LNCaP cells as a malignant model and PNT2-C2 cells as a non-malignant comparator under tightly controlled preparation and imaging conditions. The central objective is to determine whether nuclear and nucleolar descriptor distributions at the single nucleus resolution contain reproducible differences between these cell lines, and whether such differences remain interpretable once the effects of permeabilisation are considered explicitly. Accordingly, the Results therefore proceed in a staged manner from non-permeabilised DAPI (nuclear) analysis, to permeabilised DAPI (nuclear) analysis, and then finally to combined permeabilised DAPI and anti-fibrillarin (nuclear and nucleolar) analysis, such that biological interpretation is grounded first in the least perturbed condition, and subsequently extended to nucleolar measurement under matched experimental controls.

Chapter 2

Methods

2.1 Fluorescence microscopy theory

Fluorescence microscopy is a powerful analytical technique and optical imaging modality that relies on the selective absorption and subsequent emission of electromagnetic radiation by fluorescent molecules known as fluorophores to generate image contrast. The fluorescence phenomenon was first reported by Sir John Herschel, who observed that an otherwise colourless quinine solution exhibited "an extremely vivid and beautiful celestial blue colour" under solar illumination [260]. George Stokes subsequently formalised the phenomenon, introducing the term fluorescence, and recognised that the emitted light is shifted to a longer wavelength than the excitation light [261]. As a microscopy technique, fluorescence imaging emerged in the early twentieth century with ultraviolet-excitation instruments developed by August Köhler and, more decisively, Oskar Heimstädt, before later advances in incident-light optics, immunofluorescent labelling, and fluorophore chemistry established fluorescence microscopy as a general method for structure-specific biological imaging [262–264]. In the present thesis, fluorescence microscopy is used to obtain measurements of the nucleus and nucleolus in fixed prostate cell lines (as stated in Section 1.7.2).

2.1.1 Photophysical basis of fluorescence contrast

At the level of single-photon absorption, the photon energy is

$$E_\gamma = h\nu = \frac{hc}{\lambda}, \tag{2.1}$$

where h is Planck's constant, c is the speed of light in vacuum, ν is the optical frequency, and λ is the wavelength. Absorption therefore preferentially occurs for wavelengths with photon energies which overlap with allowed electronic transitions and their associated vibronic (vibrationally resolved) manifolds [265, 266]. In condensed-phase (i.e. not gaseous) fluorophores, excitation to higher vibronic levels is typically followed by rapid non-radiative relaxation (internal conversion and vibrational relaxation) into the lowest vibrational level of the first excited singlet state prior to photon emission [267]; as such, fluorescence emission is commonly treated as originating from the lowest excited state, $S_1(v_0)$, even when excitation populates higher electronic/vibronic levels [265].

Explained in physical terms, incident light can promote a fluorophore from its electronic ground state

to an electronically excited state, and after rapid non-radiative relaxation within the excited-state manifold, the fluorophore may return radiatively towards the ground state by emitting a lower-energy photon [241]. In common practice, fluorophores are typically described by excitation and emission bands, summarised by their maxima, and the emitted photon is necessarily observed at a longer wavelength than the absorbed photon, because part of the absorbed energy is dissipated through vibrational relaxation and internal conversion before emission [233, 266, 268]. This separation in excitation and emission wavelengths, termed the Stokes shift, is central to fluorescence microscopy as it enables spectral separation of the relatively weak fluorescence signal from the much more intense excitation light [233, 261, 269, 270]. These processes, along with the origin of the Stokes shift, are illustrated schematically in Figure 2.1.

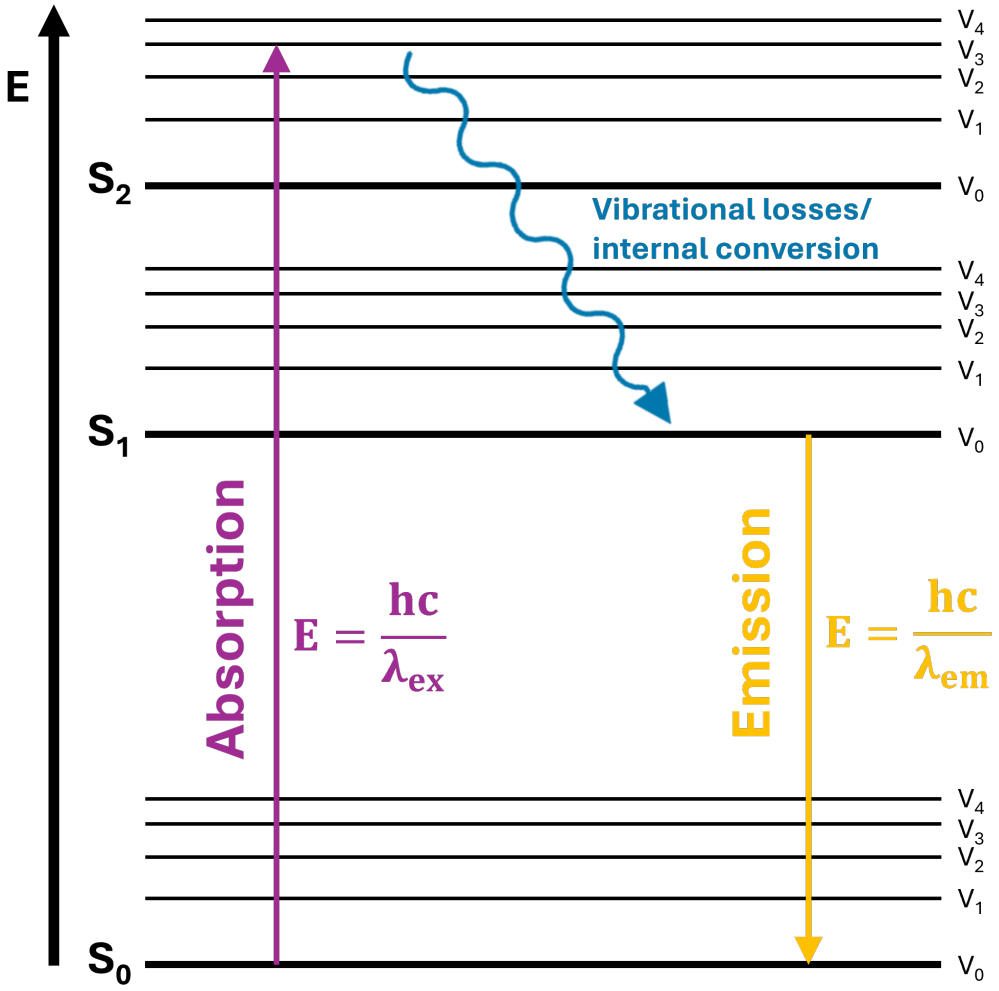


Figure 2.1: Schematic Jablonski diagram illustrating the core transitions relevant to fluorescence microscopy. S_0, S_1, and S_2 denote the electronic singlet ground and excited states, and $v_0, v_1, \ldots$ denote representative vibrational sublevels within each electronic manifold [241, 265]. Absorption of a photon at wavelength λ_{ex} promotes the molecule from S_0 to an excited vibronic level (here illustrated as population of S_2) [270]. Rapid non-radiative relaxation within/between singlet manifolds (vibrational relaxation and internal conversion respectively) then transfers population to the lowest vibrational level of S_1 [266]. Fluorescence emission occurs on relaxation towards S_0, producing a photon at wavelength $\lambda_{em} > \lambda_{ex}$, consistent with Stokes-shifted emission [261, 265]. Triplet-state pathways (e.g. intersystem crossing and phosphorescence) are possible, but omitted here for clarity [241, 265, 270].

A further practical implication of the vibronic manifold is that absorption and emission typically occur over bands, rather than at a single discrete wavelength. Broadening arises from vibronic structure (Franck-Condon factors), environmental heterogeneity (inhomogeneous broadening), and dynamical

interactions with the local environment (homogeneous broadening), such that the quantities typically reported by manufacturers and in protocol literature are usually the band maxima (or an effective band centre), rather than delta-function transition wavelengths [265, 266, 268]. In microscope configuration, the excitation wavelength (or excitation band) is therefore chosen to overlap the absorption band of the fluorophore, and the detection band is chosen to overlap the emission band while rejecting residual excitation light [233].

2.1.2 Rate model, quantum yield, and measured intensity

For a fluorophore cycling between ground and excited states under continuous illumination, fluorescence efficiency is commonly modelled via the fluorescence quantum yield Φ_{F}, defined as the probability that excitation ultimately yields a fluorescence photon:

$$\Phi_{\mathrm{F}} = \frac{N_{\mathrm{em}}}{N_{\mathrm{abs}}} = \frac{k_{\mathrm{F}}}{k_{\mathrm{F}} + k_{\mathrm{nr}}}, \tag{2.2}$$

where N_{abs} is the number of absorbed photons, N_{em} is the number of fluorescence photons emitted, k_{F} is the radiative decay rate, and k_{nr} is the effective non-radiative decay rate (incorporating internal conversion, quenching, and other loss channels) [265]. The fluorescence lifetime τ, defined as the average time a fluorophore remains in the excited state before returning to the ground state via either radiative or non-radiative decay, is

$$\tau = \frac{1}{k_{\mathrm{F}} + k_{\mathrm{nr}}}, \tag{2.3}$$

such that $\Phi_{\mathrm{F}} = {}^{\tau}/_{\tau_{\mathrm{r}}}$ where $\tau_{\mathrm{r}} = {}^{1}/_{k_{\mathrm{F}}}$ is the radiative lifetime, the hypothetical excited-state lifetime if radiative emission were the only available de-excitation pathway [271]. In the low-irradiance (non-saturating) regime relevant to conventional widefield epifluorescence imaging, the emitted fluorescence signal is approximately linear in excitation irradiance, as the steady-state excited-state population remains small [265].

For quantitative imaging interpretation, it is therefore useful to express the detected photon count during an exposure time t_{exp} as

$$N_{\mathrm{det}} \approx \eta_{\mathrm{sys}}\, \Phi_{\mathrm{F}}\, N_{\mathrm{abs}} \quad \text{with} \quad N_{\mathrm{abs}} \propto \sigma_{\mathrm{abs}}\, I_{\mathrm{ex}}\, t_{\mathrm{exp}}, \tag{2.4}$$

where I_{ex} is the excitation irradiance at the specimen, σ_{abs} is the absorption cross-section (or, equivalently in solution measurements, the molar extinction coefficient), and η_{sys} is a system efficiency factor incorporating collection efficiency (numerical aperture and refractive index), optical transmission of the emission path, filter transmission, and detector quantum efficiency [271, 272]. This expression makes explicit why intensity values in camera units are intrinsically instrument-dependent: even for identical fluorophore concentration, changes in η_{sys}, I_{ex}, or t_{exp} rescale measured intensities. As such, in this thesis, intensity-derived metrics are interpreted only under fixed acquisition settings (as described in Section 2.2) [9].

Finally, photon counting statistics provide a first-order noise model. If N_{s} is the detected signal photon count in a region of interest and N_{b} is the detected background (including autofluorescence, out-of-focus fluorescence, and residual excitation leakage), then a simple approximation for the signal-to-noise ratio

(SNR) is

$$\mathrm{SNR} \approx \frac{N_\mathrm{s}}{\sqrt{N_\mathrm{s} + N_\mathrm{b} + N_\mathrm{d} + \sigma_\mathrm{read}^2}}, \tag{2.5}$$

where N_d is the dark-count contribution (over the exposure) and σ_read is the detector read-noise expressed in electron-equivalent units [272]. Practically, maximising SNR requires both increasing N_s (within photobleaching and saturation limits) and suppressing N_b via spectral filtering and appropriate field selection.

2.1.3 Epifluorescence microscopy instrumentation

Epifluorescence microscopy is a widefield configuration in which the same objective lens is used both to deliver excitation light to the specimen and to collect emitted fluorescence [234]. This geometry is efficient for thin adherent samples on coverslips because high numerical aperture (NA) objectives simultaneously provide strong excitation irradiance at the focal plane as well as a high collection efficiency for isotropically emitted fluorescence [270, 272].

In a conventional epifluorescence instrument, a filter set is employed to separate excitation and emission into distinct paths. Conceptually, the system comprises: an excitation light source (such as a lamp or LED) whose spectrum is restricted by an excitation filter chosen to match the fluorophore absorption band, a dichroic beamsplitter that reflects the excitation light into the objective while transmitting the longer-wavelength emitted fluorescence, and an emission filter that transmits the desired fluorescence while blocking residual excitation light and unwanted background components [268]. Because fluorescence emission is typically orders of magnitude weaker than the excitation intensity at the specimen plane, the dichroic and emission filtering performance is crucial to obtaining a usable SNR in widefield imaging; without effective spectral separation, reflected/scattered excitation light dominates the detected signal [233, 270].

Two additional physical limitations are relevant for interpreting the images analysed in this thesis. First, lateral resolution (i.e. the ability to distinguish objects side by side) in diffraction-limited widefield imaging is bounded by the objective NA and emission wavelength. The Rayleigh criterion ($d_\perp \approx {}^{0.61\lambda_\mathrm{em}}/_{\mathrm{NA}}$) is commonly used to express this limit, where $d_\perp$ is the minimum resolvable separation between two points measured perpendicular to the excitation light. This emphasises that shorter wavelength channels (e.g. DAPI emission) can, in principle, support slightly finer spatial resolution than longer wavelength channels [241, 273, 274]. Second, widefield epifluorescence lacks intrinsic optical sectioning; it has a much lower axial resolution (i.e. the ability to distinguish objects stacked on top of one another) compared to the lateral resolution mentioned previously. As a result, fluorescence from planes both above and below the focal plane contribute background blur. In the present work, this is treated as a controlled and consistent property of the acquisition modality (held constant across groups by fixed imaging settings; Section 2.2), rather than being eliminated by confocal sectioning [10].

2.1.4 Fluorophores and labelling strategy

This study employs DAPI as a nuclear stain and anti-fibrillarin immunofluorescence as a nucleolar marker. DAPI provides a high-contrast nuclear channel for robust nuclear segmentation and nucleus-level morphological/intensity descriptors; fibrillarin labelling then provides an intranuclear spatially restricted signal enabling the extraction of nucleolar descriptors within each nuclear mask (as discussed in

Section 1.7.3).

DAPI (4',6-diamidino-2-phenylindole) is a DNA-binding fluorophore widely used for nuclear counterstaining. In fixed-cell preparations, DAPI preferentially binds AT-rich regions in the minor groove of double-stranded DNA and exhibits strong blue fluorescence upon UV/violet excitation [8]. Its practical strengths for the present pipeline are its high nuclear specificity and strong signal enabling reliable threshold-based nuclear segmentation across large image volumes. However, it is important to state explicitly that DAPI intensity is not a direct absolute measure of DNA mass without calibration; intensity depends on binding environment, optical path, acquisition settings, and potential preparation-dependent effects [8, 9].

Fibrillarin (FIB) is a nucleolar protein associated with small nucleolar RNP complexes and pre-rRNA processing, and is commonly used as an immunofluorescence marker for nucleoli and nucleolar reorganisation in multiple cell biology contexts [17, 18, 220, 254]. In this thesis, fibrillarin is labelled via a primary anti-fibrillarin antibody and a fluorescently labelled secondary antibody in the green channel (Alexa Fluor (AF) 488), chosen because Alexa dyes are widely used due to high brightness and photostability relative to earlier fluorophore generations [275].

The combination of a blue nuclear channel (DAPI) and a green nucleolar channel (AF488) was selected in part because it provides clear spectral separation that supports a dual-channel imaging workflow with low crosstalk under appropriate filter sets, while also providing a direct operational definition of nucleoli as fibrillarin-positive subnuclear regions within each nuclear mask (Section 1.7.3). In the present microscope configuration, the DAPI channel is excited in the near-UV/violet range and detected through a blue emission window, whereas the AF488 channel is excited in the blue range and detected through a green emission window [202, 276]. As such, efficient DAPI excitation occurs in a spectral region where AF488 absorption is negligible, while efficient AF488 detection occurs at wavelengths substantially longer than the dominant DAPI emission contribution transmitted by the DAPI filter set. Conversely, AF488 excitation is centred in the blue-green region where DAPI absorption is negligible, and AF488 emission is detected well above the DAPI emission band [202, 276]. The practical consequence is that the two channels remain well separated not only at the level of nominal dye spectra, but also after the full excitation and detection pathways are accounted for.

This is illustrated in Figure 2.2, which shows the modelled resultant detected emission spectra for the DAPI and AF488 channels. Here, the plotted spectra represent the effective detected signal after combination of the illumination band, excitation filtering, dichroic behaviour, dye excitation/emission response, and emission filtering. Presented in this way, the figure directly reflects the quantity most relevant to image formation and channel contamination in the present workflow. The model predicts that the detected DAPI signal is confined to the "blue" channel window, while the detected AF488 signal is confined to the "green" channel window, with negligible effective overlap between the two.

A simple spectral model based on these definitions predicts that direct optical bleed-through is effectively negligible under the present acquisition conditions. Using the current band definitions, the estimated AF488 contribution detected in the DAPI channel is approximately 4.76×10^{-7} % of the DAPI signal intensity, while the estimated DAPI contribution detected in the AF488 channel is approximately 5.66×10^{-9} % of the AF488 signal intensity. These values are exceedingly small, such that the modelled direct spectral crosstalk is negligible, and thus unlikely to be a material source of error in the present analysis. This does not imply that all apparent inter-channel correspondence is absent; residual

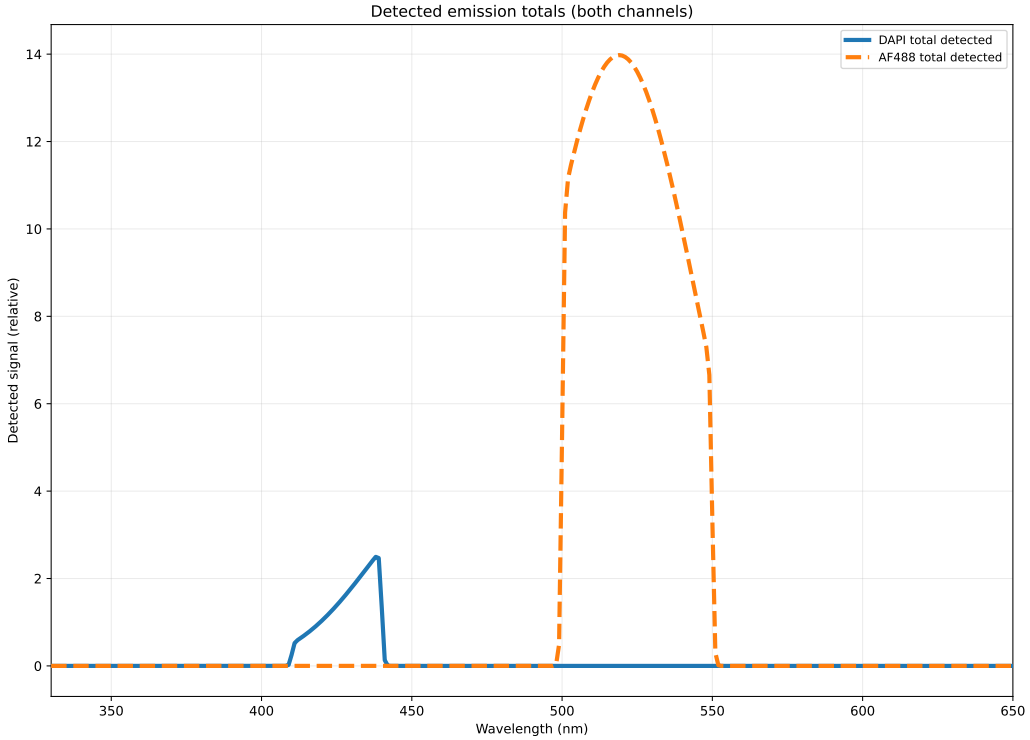


Figure 2.2: Modelled resultant detected emission spectra for the DAPI and AF488 channels under the filter/source definitions used in the present workflow. Each curve represents the effective detected signal after combination of the modelled illumination band, excitation filtering, dichroic splitting, dye excitation/emission response, and emission filtering. The detected DAPI and AF488 signals occupy well-separated spectral windows, and the modelled non-native contribution to each channel is negligible, supporting the use of this fluorophore/filter combination for sequential dual-channel imaging.

artefacts may still arise from non-spectral sources including out-of-focus widefield background, detector noise, uneven illumination, imperfect image registration, or segmentation and masking errors, however, the calculated spectral model supports the conclusion that the chosen DAPI/AF488 channel pair is (at least theoretically) well separated, such that any remaining ambiguity in nucleolar assignment is unlikely to arise primarily from direct excitation/emission bleed-through.

The combination of a blue nuclear channel (DAPI) and a green nucleolar channel (anti-fibrillarin/AF488) therefore supports a dual-channel imaging workflow in which nuclei and nucleoli are both optically and algorithmically distinguishable. This is advantageous for the present analysis because nuclei can first be segmented robustly using the DAPI channel, after which fibrillarin-positive subnuclear regions can be evaluated within each nuclear mask as an operational definition of nucleoli (Section 1.7.3).

2.2 Cell culture, sample preparation, and fluorescence imaging

Cell culture, fixation, and fluorescence staining were performed using standard handling approaches consistent with the established use of PNT2-C2 and LNCaP prostate cell lines in the literature, including studies associated with the Maitland group and related prostate cell culture work [2, 3, 151, 162, 166, 167, 277–284]. Immortalised cell lines were used rather than primary cultures because the present study was designed as a controlled, measurement focused fluorescence imaging workflow requiring tightly standardised preparation and imaging conditions. In this context, immortalised lines provide the practical advantages of culture stability and batch reproducibility, while still allowing biologically meaningful malignant vs. non-malignant contrasts to be examined [5, 144]. PNT2-C2 was used here as a non-malignant immortalised prostate epithelial comparator, and LNCaP as an established prostate

cancer line derived from lymph-node metastasis [2, 3, 166]. Three preparation regimes were treated as distinct analytical groups: non-permeabilised DAPI-only, permeabilised DAPI-only, and permeabilised DAPI plus anti-fibrillarin (FIB) immunofluorescence.

2.2.1 Cell culture and maintenance

Cryostocks were stored in liquid nitrogen for long-term storage, or at −80 °C for short-term storage, in freezing medium consisting of 10 % (v/v) dimethyl sulfoxide (DMSO; Fisher Chemical D/4120/PB08, batch 2354259) in fetal bovine serum (FBS; Gibco A5256701, batch 2575647), consistent with standard cryopreservation practice for mammalian cell lines [285]. On recovery, cryovials were warmed rapidly by hand and the thawed suspension transferred immediately to a T25 flask containing 9 mL of pre-warmed complete growth medium (R10). R10 comprised Roswell Park Memorial Institute (RPMI) 1640 (Gibco 31870-074, batch 2560983), supplemented with 10 % FBS and 2 mM L-glutamine (Gibco 25030-024, batch 2851935). No antibiotics or antifungals were used throughout the culture process to avoid any potential perturbation they may have caused [286]. Cells were then incubated at 37 °C in a humidified atmosphere of 5 % CO_2. After overnight recovery and adhesion, cultures were gently washed with phosphate-buffered saline (PBS; Gibco 14200-067, batch 2659578) and the medium was replaced with 6 mL fresh R10.

Routine culture was performed in T25 flasks under the same incubation conditions. Medium was replaced approximately every 48 h, with gentle PBS washing at each change, and mechanical shock was minimised throughout handling. Cultures were passaged at approximately 80 % confluence using trypsin (Gibco 25300-054, batch 2578371), which was neutralised immediately after detachment. Split ratios were no less than 1:10 for PNT2-C2 and 1:5 for LNCaP. Routine culture monitoring was performed using a Leica DM IL LED inverted microscope with a 10× phase-contrast dry objective [287]. Regular PCR-based mycoplasma testing was performed, including an additional test immediately prior to fixation. The gel electrophoresis image from the latest PCR test is shown in Figure 2.3.

Passage number was controlled throughout, with absolute upper passage limits set at < 50 for LNCaP and < 150 for PNT2-C2, consistent with the general need to limit the passage-related drift in observable cell-state characteristics in long-term cultured cancer and immortalised cell lines [153]. Experimental LNCaP cultures were all at passage 31 at the time of fixing, and experimental PNT2-C2 cultures were at passages 123-124. Cells were maintained for at least two passages after thawing prior to plating, and fixed and imaged within three passages of thawing.

2.2.2 Coverslip preparation, seeding, and serum starvation

Cells were plated onto 13 mm circular glass coverslips placed in 24-well plates. Coverslips were cleaned, UV-sterilised, and stored in 100 % ethanol prior to use. Cell number and viability were assessed using a haemocytometer with Trypan blue exclusion (Gibco 15250-061), and approximately 5×10^4 cells were seeded per well in 1 mL R10 [289, 290]. After seeding, plates were incubated for 24 h to permit attachment.

For experimental preparation, R10 was then replaced with an equal volume of serum-free RPMI 1640 (no added supplements), and cells were incubated for a further 24 h. This serum-starvation step was used to reduce variability in growth-state behaviour by synchronising cells to the G0 (resting) phase and impose a more standardised pre-fixation condition across groups, while recognising that such

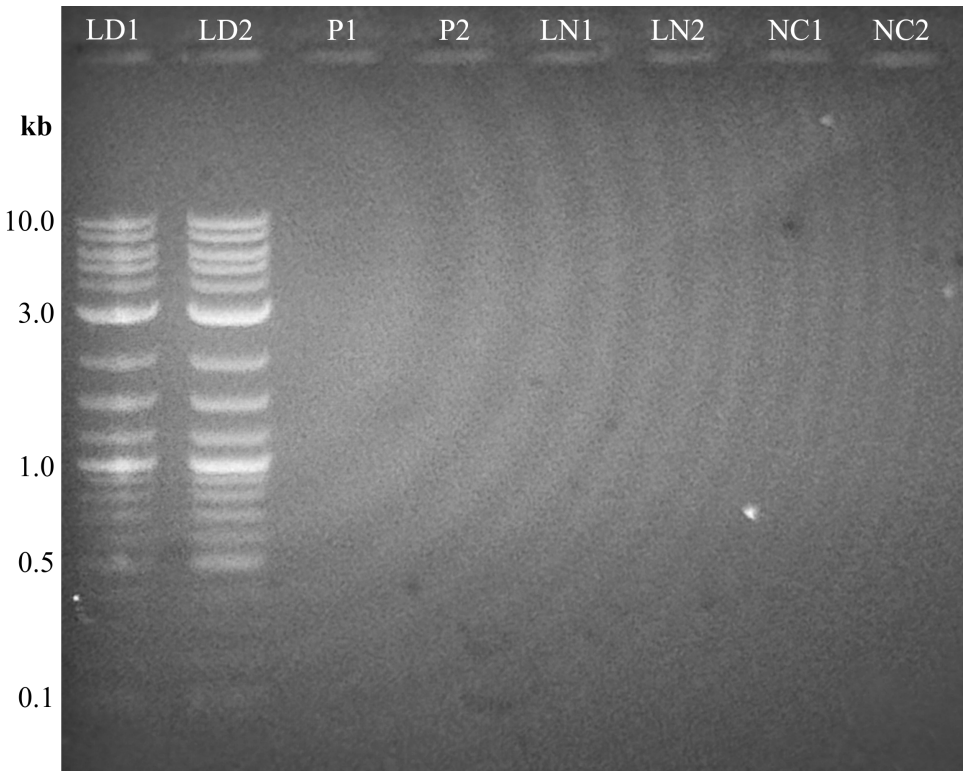


Figure 2.3: Gel electrophoresis image following PCR testing of LNCaP and PNT2-C2 cells immediately prior to fixation. From left to right, the lanes correspond to: LD1/LD2, 1kb Plus DNA ladder (NEB N3200S, batch 10222180); P1/P2, PNT2-C2 sample 1/2; LN1/LN2, LNCaP sample 1/2; NC1/2, negative control 1/2. The numbers on the left correspond to numbers of base pairs (in thousands), used to describe the size of detected DNA fragments. The negative control in this case utilised ultrapure, nuclease free water instead of cell culture supernatant. The absence of any detected signal in the tested sample lanes is consistent with the absence of any detectable mycoplasma contamination (also consistent with the evidence from fluorescence images, as DNA staining is an approved method for detecting mycoplasma recommended by the European Medicines Agency [288]), however as no positive control was used, the interpretation is somewhat limited.

synchronisation cannot be considered 100% absolute [291–293]. After serum starvation, medium was aspirated and wells were washed with PBS. Throughout the rest of this chapter, one "PBS wash" is defined as the addition of 500 µL PBS, gentle swirling, aspiration, and repetition for a total of three cycles.

2.2.3 Fixation, permeabilisation, and immunostaining

Cells were fixed with 4 % paraformaldehyde (PFA; Thermo Scientific Chemicals J19943-K2, batch 241250), using 400 µL per well for 20 min at room temperature. Gentle swirling was applied periodically during fixation to promote even exposure. Following fixation, wells were washed with PBS. PFA fixation was used to maximise preservation of cellular and nuclear architecture prior to fluorescence imaging (and to facilitate efficient DAPI staining), while acknowledging that chemical fixation and subsequent permeabilisation can influence molecular retention and measured fluorescence distributions [16, 294].

From the post-fixation wash, three endpoint-specific preparation branches were followed. For the non-permeabilised DAPI-only condition, cells proceeded directly to DAPI staining, thereby preserving the pre-detergent reference state. For the permeabilised DAPI-only condition, cells underwent blocking/permeabilisation prior to DAPI staining, but no antibody incubation was performed. For the permeabilised DAPI+FIB condition, cells underwent blocking/permeabilisation followed by primary and secondary antibody incubation prior to DAPI counterstaining. This branched design was used so

that detergent-associated effects could be isolated from the additional effects of antibody-based nucleolar labelling; more explicitly, comparison of non-permeabilised and permeabilised DAPI-only preparations provided a direct control for preparation-induced changes attributable to detergent exposure, which is known to alter molecular extraction and redistribution in fixed-cell protocols [16, 295, 296].

Blocking and permeabilisation were performed using a solution of 1 % (w/v) goat serum (Sigma Aldrich G6767, batch 23K345) and 0.1 % (v/v) Triton X-100 (Sigma Aldrich 9036-19-5) in PBS. Each well received 500 µL of this solution for 30 min at room temperature, with periodic gentle swirling. The 30 min duration for permeabilisation/blocking was directly supported by prior studies on adherent cultures. Closely comparable 0.1 % Triton X-100 permeabilisation conditions, including 30 min protocols in prostate cancer cells by van Oorschot et al. (2013), Inder et al. (2014), and Landers et al. (2008), provide direct precedent for the method used here [281, 297, 298]. More broadly, Triton X-100-based fixed-cell immunofluorescence workflows have also been used in other adherent cancer cell monolayers, including H1299 (lung cancer), and MIA PaCa-2 (pancreatic cancer), indicating that the present protocol lies within an established methodological range [283, 284]. For permeabilised DAPI+FIB samples, the blocking/permeabilisation solution was removed and, notably, no intervening wash was performed before primary antibody incubation. Primary immunostaining used anti-fibrillarin [38F3] (Abcam ab4566, batch 1093841-2), diluted 1:100 in 1 % (w/v) goat serum in PBS, with 400 µL added per well. Samples were incubated for 24 h at 4 °C with gentle agitation/swirling. From the end of this incubation onward, all remaining steps were performed protected from light. After removal of the primary antibody, wells were washed with PBS and incubated with goat anti-mouse IgG (H+L), highly cross-adsorbed, Alexa Fluor Plus 488 secondary antibody (Invitrogen A32723, batch AA402438), diluted 1:500 in 1 % (w/v) goat serum in PBS, again using 400 µL per well. Secondary incubation was performed for 1 h at room temperature in the dark with periodic gentle swirling, followed by PBS washing.

DAPI counterstaining was then performed for all preparation branches using DAPI (Sigma Aldrich D9542) at a working concentration of 400 µgmL^{-1} in PBS. Each well received 400 µL of staining solution for 5 min at room temperature in the dark with gentle swirling, after which stain was removed and a final PBS wash was performed.

2.2.4 Mounting, curing, and storage

After the final PBS aspiration, approximately 500 µL of reverse-osmosis water was added to each well and gently swirled. Coverslips were removed using a 12-gauge needle and forceps, touching only a single point at the edge. Each coverslip was then held perpendicular to the ground for several seconds to allow water to collect to one edge, and only the edge of the coverslip was touched to Whatman paper for several seconds to remove excess water by capillary action, ensuring that neither face contacted the paper. Coverslips were placed cell-side up on lint-free paper towel under a cover, protected from light, for no more than 5 min to permit limited evaporation without over-drying.

Glass slides were cleaned with isopropyl alcohol, rinsed with reverse-osmosis water, and stored in ethanol prior to mounting. Coverslips were mounted using approximately 5 µL of VECTASHIELD Vibrance® Antifade Mounting Medium (Vectorlabs H-1700-10, batch WOVUS32518). Mounted samples were allowed to cure for approximately 1 h at room temperature in the dark, and then overnight at 4 °C. The following morning, air-dry nail varnish was applied to the corners of the coverslip to secure it and

prevent movement during imaging. Slides were then stored in the dark at 4 °C until imaging, which was performed within ≤1 month of sample preparation.

2.2.5 Fluorescence image acquisition

Fluorescence imaging was performed on a Zeiss Observer 7 widefield epifluorescence microscope running Zeiss ZEN software, equipped with a Colibri 7 LED light source, a Plan-Apochromat 63×/1.40 Oil DIC M27 objective, and an Axiocam 807 camera [299, 300]. The same camera was used for both channels, and acquisition settings were held constant across all samples and conditions such that the intensity-derived measurements could be interpreted comparatively under matched optical conditions, as discussed in Section 2.1.3.

For the DAPI channel, excitation used the 385 nm LED module at 15 % intensity with an exposure time of 120 ms. For the fibrillarin (FIB) channel (recorded using the microscope GFP channel and corresponding to the Alexa Fluor 488 secondary), excitation used the 475 nm LED module at 20 % intensity with an exposure time of 250 ms. Nominal depth of focus was 0.72 µm for the DAPI channel and 0.80 µm for the fibrillarin channel. Where both channels were present, matched DAPI and FIB images were acquired sequentially for the same field of view. Images were exported as single-channel 16-bit greyscale TIFF files for downstream analysis.

Spatial calibration was 0.07 µm px^{-1}. If s denotes the calibrated pixel size and N_x, N_y the image width and height in pixels, then the physical field of view is

$$L_x = N_x s, \qquad L_y = N_y s. \tag{2.6}$$

With $N_x = 3120$, $N_y = 2208$, and $s = 0.07\,\mu\text{m px}^{-1}$, this corresponds to a field of view of approximately 218.4 µm × 154.6 µm.

Imaging was conducted over approximately one week, with the majority of data acquired over two consecutive days. Each coverslip was fully imaged within a single session. To reduce subjective field-selection bias, coverslips were traversed using a systematic snake-like raster. Starting at the top of the coverslip, fields were acquired from left to right until the edge; the stage was then moved down by one field of view, acquisition proceeded right to left, and this alternating pattern was repeated to the bottom of the coverslip. Where possible, fields with cells truncated at the image edges were minimised, and fields containing only severely damaged or clearly aberrant material (for example debris or overtly apoptotic cells) were excluded, whereas fields containing a mixture of acceptable and damaged cells were retained. This acquisition strategy was adopted with the aim of maximising coverage of each coverslip while limiting operator-driven bias as much as possible.

2.3 Automated computational analysis pipeline

2.3.1 Analytical rationale and data flow

The quantitative objectives of this study are primarily distributional rather than image-averaged: the individual nucleus, not the image, is treated as the primary sampling unit, and the principal analyses require relatively large pooled sets of nucleus-level measurements in order to assess both central tendency and heterogeneity with acceptable statistical stability [301]. Under these conditions,

manual selection of regions of interest is not methodologically appropriate; it is slow, difficult to standardise across large image sets, and can be liable to introduce observer-dependent inconsistency in both boundary placement and inclusion criteria [302]. An automated analysis pipeline was therefore developed so that nuclear identification, feature extraction, and downstream pooling/resampling could be applied reproducibly across all images using a single set of algorithmic rules.

A range of established platforms and modelling paradigms could, in principle, be used for automated high-content nuclear segmentation and measurement. These include modular GUI-driven workflows (most notably CellProfiler) and various supervised machine learning/deep learning segmentation models. The decision to implement a bespoke Python pipeline in this thesis was therefore not made on the basis that alternative tools are incapable of nuclear quantification; rather, it was made to satisfy a specific set of constraints implied by the present measurement model and downstream statistical design.

CellProfiler is a well-established and widely used platform for high-throughput microscopy analysis, built around modular pipelines that perform illumination correction, object identification, and per-object measurement [303, 304]. In particular, nucleolus-oriented workflows using CellProfiler have been published, including pipelines that explicitly use nucleolar markers (e.g. fibrillarin-based nucleolar segmentation as part of a high-throughput nucleolar assay) [221].

The analysis pathway adopted here, however, required several features that are not typical outputs of standard CellProfiler-based pipelines, and which are substantially more direct to implement and audit in code. First, a core design choice of this thesis is the explicit separation between the images which are used to generate segmentation masks and the images used for quantitative intensity measurements. Specifically, image normalisation was applied only during segmentation to improve the robustness of nucleus detection. Once the segmentation masks were generated, all fluorescence intensity measurements were extracted from the rolling-ball background-subtracted images in their original intensity scale, without normalisation, ensuring that the measured fluorescence values were not altered. Second, downstream stages required not only scalar per-nucleus summaries, but also the export of per-nucleus pixel intensity vectors ("pixel-block" files) so that the complete distribution of pixel intensities within each nucleus could be analysed. This allowed histogram generation and resampling-based convergence analysis to be performed directly using the original pixel values, rather than from summary statistics such as the mean or median. While CellProfiler readily exports per-object summary measurements, full per-object pixel lists are not provided as a standard CellProfiler output, and would therefore require additional custom tooling to generate. Third, the present workflow required deterministic batch processing with regular-expression-based pairing of DAPI and FIB fields, conditional transfer of DAPI-derived masks to a second channel, and stage-consistent provenance logging and quality control figure generation. While these functions could be reproduced within a modular image-analysis framework such as CellProfiler, doing so would require significant pipeline-specific scripting. Implementing the complete workflow within a single Python script ensured that every stage of the analysis including image pairing, segmentation, filtering, feature extraction, and output generation was fully specified, reproducible and version controlled.

Supervised machine learning and deep learning segmentation models represent a second major alternative. Deep learning has enabled strong performance in cell and nucleus segmentation across many imaging modalities [147, 190, 305], with widely adopted architectures and packages including U-Net-type models [306], Cellpose [307], and StarDist [308]. These approaches can outperform classical

thresholding/watershed pipelines in challenging datasets, particularly when shape variability, low SNR, or dense packing make hand-designed rules brittle. However, their practical use typically requires either image distributions that are sufficiently similar to those used for training a "generalist" model, or fine-tuning using representative labelled ground truth for the specific dataset and segmentation style required. Even for generalist models, performance can be suboptimal when test images differ meaningfully from the training distribution, motivating manual/supervised re-training or model adaptation [309]. In the present thesis, the preparation design intentionally includes multiple protocol regimes (non-permeabilised vs. permeabilised, with and without secondary-channel immunofluorescence), which can alter background structure and intensity statistics. Constructing and validating a training set that adequately spans these regimes, and demonstrating that the resulting model does not introduce systematic biases into the nucleus-level distributions that are central to the subsequent statistical analyses, would require a dedicated benchmarking and annotation effort beyond the intended scope of this project. Moreover, because nuclear morphology and DAPI intensity heterogeneity are themselves outcome variables, a rule-based, parameter-auditable method was preferred so that the segmentation assumptions remain explicit and more easily understood by readers, and can be interrogated directly via quality control overlays and parameter sweeps.

Finally, the specific segmentation approach adopted here was selected in preference to highly specialised methods such as that of Molnar et al. (2016) for reasons of model compatibility, morphological neutrality, and downstream measurement integration [310]. Molnar et al. proposed an active-contour/phase-field framework (the "multi-layer gas of near-circles") intended to segment overlapping nuclei in high-confluency settings containing substantial overlap by exploiting an explicit fluorescence image-formation assumption; namely, that under low numerical-aperture widefield imaging, the measured intensity in regions of superposition is approximately additive, such that two overlapping nuclei contribute roughly twice the signal of one [310]. This is a principled and elegant solution for dense overlap regimes, but is also a stronger modelling commitment than the one made in the present workflow. The images analysed in this thesis were acquired using a high-NA oil-immersion objective, and the principal analytical requirement was not merely to recover plausible object outlines, but to do so in a way that preserved compatibility with subsequent per-nucleus intensity extraction, full distributional analysis, and export of raw intranuclear pixel vectors. Under those conditions, adopting a segmentation model whose data term is explicitly tied to an additive overlap-intensity formalism would have required additional assumptions about how the measured DAPI signal should behave in partially superposed or poorly separated regions, whereas the present watershed-based segmentation pipeline instead relies on local topographic separation of the DAPI foreground and then measures intensity directly from the retained nuclear mask without embedding an explicit generative overlap model.

A second consideration is that the Molnar framework incorporates a geometric prior favouring near-circular nuclei [310]. This is advantageous when the target objects are highly circular and overlap is the dominant ambiguity, but is less desirable when nuclear elongation is itself a biological quantity of interest. Molnar et al. explicitly report progressive degradation with increasing ellipticity and do not recommend the method when the major-to-minor axis ratio (i.e. aspect ratio) exceeds approximately 1.75, a result which was observed regularly for LNCaP nuclei in this study [310]. By contrast, nuclear shape variation in this study was not a nuisance to be regularised away, but a measured descriptor that was preserved and quantified through ellipse fitting and aspect-ratio extraction. As such, a segmentation approach with intrinsically weaker built-in shape preference was methodologically preferable, even if

this meant accepting that some difficult cases of under-segmentation would remain. More explicitly, the present pipeline was designed so that any contraction, merging, or separation behaviour arises from transparent operations: background correction, thresholding, distance-transform peak selection, watershed separation, and post hoc filtering, rather than from a strong prior which may systematically bias the recovered morphology towards near-circular forms.

A third practical reason concerns initialisation, scalability, and auditability. Molnar et al. showed that watershed segmentation initialised from user-defined seed points (i.e. manually specified markers indicating the approximate centre of each nucleus) improved segmentation accuracy relative to neutral initialisation, in which no such prior markers were provided [310]. The present study analysed large batches of images using a single, fixed analysis pipeline. Consequently, segmentation was required to operate automatically and reproducibly across all fields without manual intervention or image-specific parameter adjustment. Methods whose optimal performance depends on manual seed placement or comparable image-specific tuning were therefore less suitable than a fully automated thresholding and watershed approach using fixed parameter values. The bespoke Python workflow also needed to integrate directly with later analysis stages by exporting per-nucleus CSV summaries, raw DAPI pixel vectors, and mask-referenced paired-channel measurements for subsequent nucleolar analysis. In other words, the required output of Stage 1 was not only a visually plausible segmentation, but a complete, auditable, batch-stable measurement object for each nucleus.

Overall, the chosen DAPI-driven pipeline provided a practical balance between segmentation accuracy, automation, and reproducibility for the objectives of this thesis. Although more sophisticated overlap-resolving methods may achieve higher segmentation accuracy in some circumstances, the chosen approach was better suited to the automated, large-scale quantitative analysis undertaken in this thesis.

2.3.2 Implementation overview

The computational workflow was implemented in Python 3.12.10, and split across several distinct scripts [311]. The Stage 1 segmentation/extraction script extended/built upon an earlier DAPI analysis pipeline created within the research group [312]. The analytically important modifications for the present work were: enabling batch processing of complete image sets, regular-expression-based channel pairing, rolling-ball background subtraction, explicit separation of segmentation pre-processing from intensity quantification, ellipse-based nuclear shape estimation with propagated uncertainty, parallel execution across fields, and conditional handling of paired fibrillarin (FIB) images. Software dependencies are therefore reported here only at the level needed to define the method: PyImageJ/Fiji for background subtraction, NumPy/SciPy and scikit-image for numerical and segmentation operations, OpenCV for contour handling and ellipse fitting, tifffile for image input/output, and Matplotlib for quality-control visualisation [313–320].

The full analysis pipeline was organised as four linked stages. Stage 1 comprised DAPI-driven nuclear segmentation and per-nucleus feature extraction, with optional transfer of the DAPI-derived nuclear masks to paired FIB images. Stage 2 used the Stage 1 exports to generate pooled descriptive outputs and, separately, repeated subsampling-derived convergence series data. Stage 3 fitted distributional models to the exported raw-value summaries using kernel density estimation (KDE) and identified the best fitting model(s) using information criteria and integrated classification likelihood (ICL) [258, 259].

Stage 4 performed convergence testing and fitted power law decay models to the stage 2 convergence outputs. This subsection describes Stage 1 in detail. Its role in the overall workflow is therefore foundational: all downstream analyses inherit the nuclear definitions established here. The Stage 1 algorithm is summarised in Figure 2.4, while some of the downstream consumption of its outputs is shown schematically in Figure 2.8.

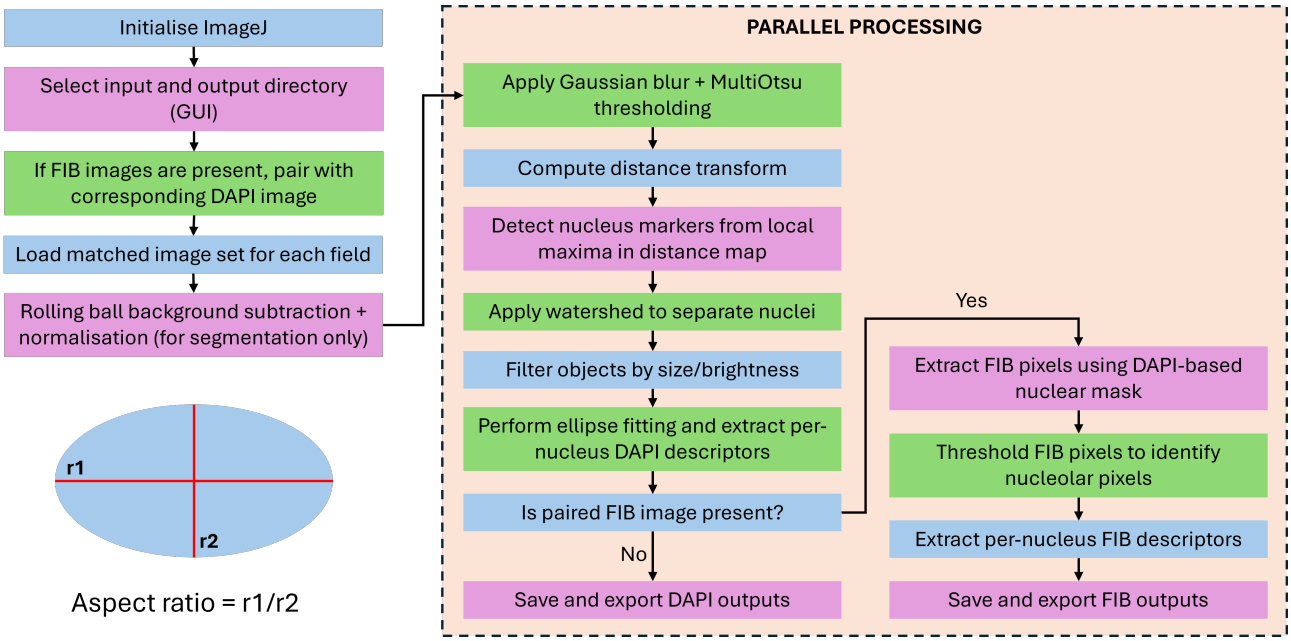


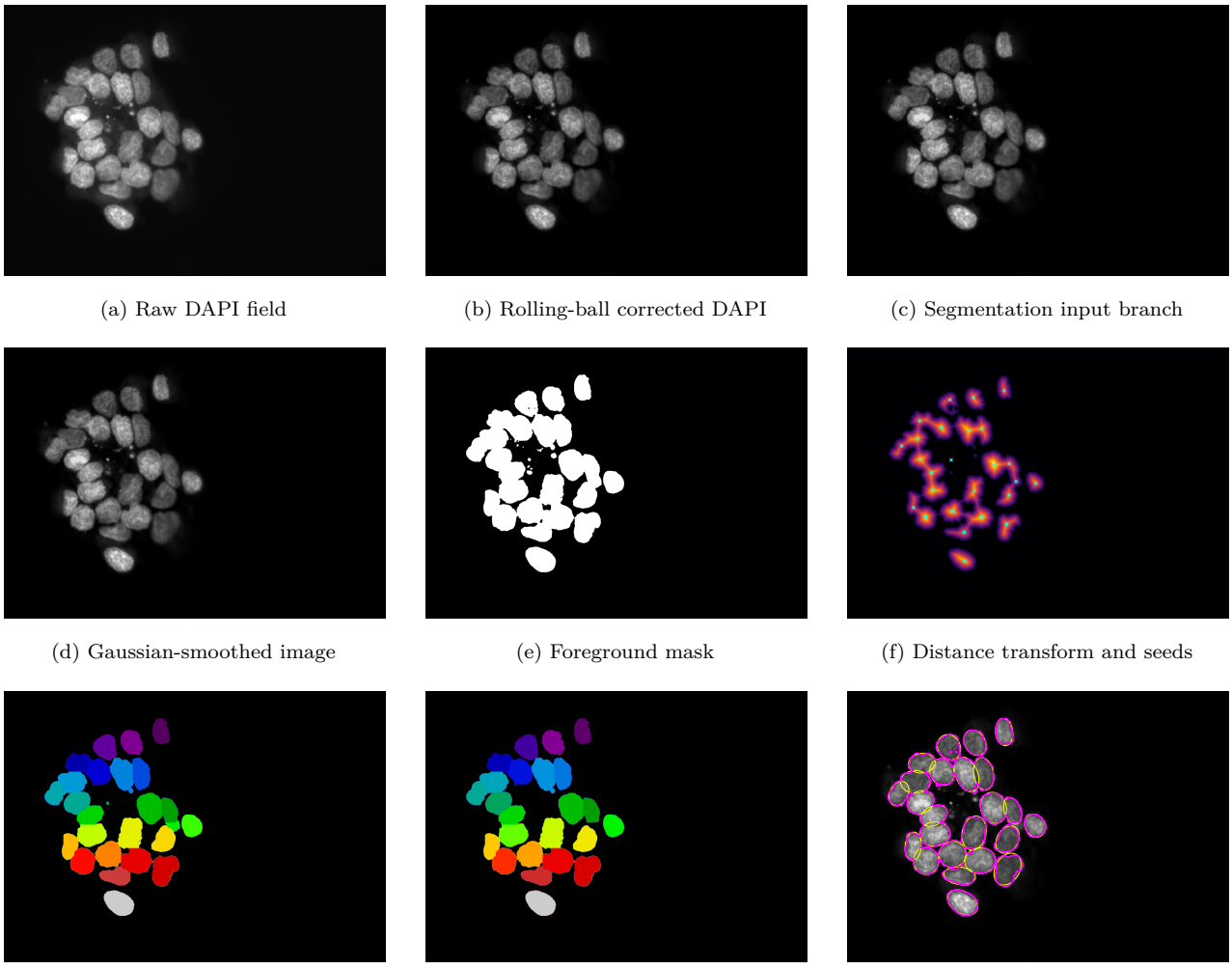
Figure 2.4: Stage 1 automated segmentation and feature-extraction workflow used to process DAPI images, with optional transfer of retained DAPI-derived nuclear masks to paired fibrillarin (FIB) images. The flowchart summarises the algorithmic logic from image pairing and background correction through foreground detection, watershed-based instance separation, post-segmentation filtering, ellipse fitting, and descriptor extraction, followed by conditional FIB-based nucleolar assignment where a matched second channel was available. A representative image-level implementation of the DAPI-led segmentation branch is shown in Figure 2.5, while the downstream transfer of retained DAPI-derived masks to the paired FIB image is shown in Figure 2.6. Exact implementation parameters are provided in Section A.1.

2.3.3 Stage 1: DAPI-driven nuclear segmentation and per-nucleus feature extraction

Inputs, pairing, and execution structure

The Stage 1 segmentation script operated on single-channel greyscale TIFF images exported from the microscope workflow described in Section 2.2. DAPI images were mandatory, and where a matched FIB image was present for the same field, it was paired to the corresponding DAPI image using filename-based regular-expression rules. Unpaired DAPI images were treated as DAPI-only inputs. Processing then proceeded independently on a field-by-field basis, allowing parallel execution across images while preserving nucleus-level outputs for each field, consistent with the initialisation, pairing, and per-field execution structure shown in Figure 2.4.

This design reflects the measurement model used throughout the chapter: nuclear segmentation was always driven by the DAPI channel, because DAPI provided the highest-contrast and most spatially complete nuclear signal. Paired FIB images were never used to define nuclear boundaries independently. Instead, where present, they were interrogated only within the DAPI-derived nuclear masks, ensuring that nucleus-level measurements remained grounded to a single segmentation basis across all preparation groups. This logic is illustrated in Figure 2.4, with each the effect of each step shown in Figures 2.5



(a) Raw DAPI field

(b) Rolling-ball corrected DAPI

(c) Segmentation input branch

(d) Gaussian-smoothed image

(e) Foreground mask

(f) Distance transform and seeds

(g) Watershed labels before filtering

(h) Final retained nuclear masks

(i) Final retained nuclei and ellipses

Figure 2.5: Representative step-by-step implementation of the DAPI-led Stage 1 nuclear-segmentation pipeline on a single field of view, corresponding to the DAPI-processing branch of Figure 2.4. The same permeabilised PNT2-C2 DAPI+FIB field is shown throughout so that the effect of each operation is visually traceable. (a) shows the raw DAPI image, as exported from the microscope. (b) shows the same field after rolling-ball background subtraction. (c) shows the DAPI image branch following min-max normalisation, used strictly for segmentation. (d) shows the Gaussian-blurred segmentation image. (e) shows the tightened multi-Otsu foreground mask used as the provisional nuclear foreground. (f) shows the corresponding Euclidean distance transform with the marker seeds used to initialise separation of touching nuclei. (g) shows the watershed-labelled nuclei before post-segmentation filtering. (h) shows the final retained nuclear masks after size- and brightness-based filtering. (i) shows those same retained nuclei with fitted ellipses overlaid to connect the segmentation output directly to the downstream aspect-ratio measurement. Transfer of these retained DAPI-derived nuclear masks to the paired FIB image is shown separately in Figure 2.6. Exact implementation parameters are provided in Section A.1.

and 2.6.

Pre-processing

Pre-processing was designed around an explicit separation between segmentation robustness and quantitative measurement. A representative field carried through these operations is shown in Figure 2.5, beginning with the raw DAPI image in Figure 2.5(a). First, uneven low-frequency background staining was reduced by using rolling-ball subtraction [313, 321], yielding the background-corrected image shown in Figure 2.5(b). This corrected DAPI image was retained as the quantitative image used for downstream intensity extraction. A separate segmentation copy was then min-max normalised for mask generation only, as illustrated in Figure 2.5(c). This separation was methodologically deliberate: it improved threshold stability without artificially rescaling the reported per-nucleus intensity descriptors. Exact implementation details are provided in Section A.1.

Nuclear foreground detection

Initial nuclear foreground detection was performed on the segmentation branch after Gaussian smoothing followed by multi-Otsu thresholding [317, 322, 323], as illustrated in Figures 2.5(d) and 2.5(e). The blur acted to suppress fine intranuclear texture and local shot-noise fluctuations, while thresholding produced a provisional binary nuclear mask. The working threshold was tightened slightly relative to the first Otsu boundary with the aim of reducing false bridges between neighbouring nuclei in crowded fields; the exact threshold rule is detailed in Section A.1. This stage corresponds to the foreground-detection branch of Figure 2.4.

Separation of touching nuclei

Because adherent-cell preparations frequently contain nuclei that are close together or partially touching in projection, foreground detection alone is insufficient to define nuclear instances robustly. Nuclei separation was therefore achieved by a marker-controlled watershed procedure. The Euclidean distance transform of the binary foreground mask was first computed, and local maxima in this distance map were then used as candidate nucleus markers, as shown in Figure 2.5(f) [324]. Watershed was subsequently applied to the negative distance map, constrained to the foreground mask, thereby yielding the labelled nuclear instances shown before filtering in Figure 2.5(g) [325]. This stage corresponds to the seeded-separation branch of Figure 2.4. The exact details of this stage used to limit over-segmentation are given in Section A.1.

Post-segmentation filtering

After watershed separation, labelled objects were filtered to remove regions unlikely to represent valid nuclei, corresponding to the post-segmentation filtering stage shown schematically in Figure 2.4. The transition from the pre-filter result in Figure 2.5(g) to the retained masks in Figure 2.5(h) illustrates this step directly. The practical aim was to suppress debris, grossly merged regions, and saturated or otherwise aberrant objects before descriptor extraction, without being too aggressive so as to not exclude genuine biological signal. Using the microscope calibration $s = 0.07\,\mu\text{m}\,\text{px}^{-1}$ from Section 2.2, projected nuclear area was calculated as

$$\text{Area}_i = n_i s^2, \tag{2.7}$$

where n_i is the number of pixels in nuclear mask i. The retained nuclei were then relabelled consecutively, and the numbers of detected and removed objects in each dataset were recorded for auditability. Exact size and brightness filtering rules are given in Section A.1. The final retained masks, and their downstream geometric interpretation via ellipse fitting, are shown in Figures 2.5(h) and 2.5(i).

Ellipse fitting and per-nucleus DAPI descriptor extraction

For each retained nucleus, the contour of the DAPI-derived mask was extracted and fitted with an ellipse using OpenCV-based contour geometry with uncertainty propagation handled separately [318, 326, 327]. Ellipse fitting was introduced because axis-aligned bounding boxes do not provide a rotation-invariant description of nuclear shape: an elongated nucleus can appear artificially circular if its major axis lies obliquely to the image axes. The fitted ellipse instead yields a geometric descriptor referenced to the principal axes of the nuclear outline itself, and this operation is illustrated directly by the transition from the retained masks in Figure 2.5(h) to the ellipse-overlay panel in Figure 2.5(i), as well as schematically by the aspect-ratio inset in Figure 2.4.

If r_1 and r_2 denote the major and minor axes of the ellipse fit to nucleus i, with $r_1 \geq r_2$, the nuclear aspect ratio was defined as

$$\mathrm{AR}_i = \frac{r_1}{r_2}. \tag{2.8}$$

By definition, $\mathrm{AR}_i \geq 1$, with larger values indicating greater elongation/eccentricity. An associated ellipse-fit uncertainty was propagated and exported as an aspect-ratio error term. Where ellipse fitting failed, the raw export retained a placeholder value so that such cases could be identified explicitly during downstream handling rather than being silently discarded.

Per-nucleus DAPI intensity descriptors were then extracted from the background-subtracted measurement image using the nuclear mask $\mathcal{N}_i$ defined by Stage 1 segmentation. Let $D(p)$ denote the DAPI intensity at pixel $p \in \mathcal{N}_i$, and let $n_i = |\mathcal{N}_i|$. The total nuclear DAPI intensity was

$$I_{\mathrm{DAPI},i} = \sum_{p \in \mathcal{N}_i} D(p), \tag{2.9}$$

and the corresponding DAPI density was

$$\rho_{\mathrm{DAPI},i} = \frac{I_{\mathrm{DAPI},i}}{\mathrm{Area}_i}, \tag{2.10}$$

reported in a.u. $\mu\mathrm{m}^{-2}$. In addition, the within-nucleus mean, standard deviation, and coefficient of variation were calculated from the nuclear pixel vector:

$$\mu_i = \frac{1}{n_i} \sum_{p \in \mathcal{N}_i} D(p), \tag{2.11}$$

$$\sigma_i = \sqrt{\frac{1}{n_i - 1} \sum_{p \in \mathcal{N}_i} (D(p) - \mu_i)^2}, \tag{2.12}$$

$$\mathrm{CV}_{\mathrm{DAPI},i} = 100 \frac{\sigma_i}{\mu_i}. \tag{2.13}$$

These descriptors respectively quantify total signal, area-normalised signal, mean intranuclear intensity,

intranuclear dispersion, and relative heterogeneity of the DAPI distribution within each nucleus.

Two additional descriptors were also exported to capture the relative contribution of very bright and very dark DAPI pixels within each nucleus. In the default fixed-threshold path, global 16-bit thresholds of $T_{\text{bright}} = 22\,000$ a.u. and $T_{\text{dark}} = 10\,000$ a.u. were used. The corresponding intensity-weighted ratios were

$$R_{\text{bright},i} = \frac{\sum_{p\in\mathcal{N}_i,\, D(p)>T_{\text{bright}}} D(p)}{\sum_{p\in\mathcal{N}_i} D(p)}, \tag{2.14}$$

$$R_{\text{dark},i} = \frac{\sum_{p\in\mathcal{N}_i,\, D(p)<T_{\text{dark}}} D(p)}{\sum_{p\in\mathcal{N}_i} D(p)}. \tag{2.15}$$

These are not simply counts of bright or dark pixels; rather, they are fractions of the total nuclear DAPI signal contained within the corresponding intensity tails. As such, they are sensitive both to the extent and magnitude of tail-valued signal within each nucleus.

Transfer of DAPI-derived masks to paired FIB images and nucleolar extraction

Where a matched FIB image was present, the DAPI-derived nuclear masks were carried forward unchanged and used as the sole spatial reference for all paired-channel measurements; no independent nuclear segmentation was performed from the FIB channel, consistent with the conditional paired-FIB branch shown in Figure 2.4. This design choice is methodologically important. The DAPI channel provides a spatially complete nuclear signal, whereas fibrillarin is an intranuclear marker whose signal is spatially restricted to nucleolar regions and therefore is not, by itself, an appropriate basis for defining the full nuclear boundary. While there was some dim background nuclear staining within the FIB channel, it was not generally strong enough to be a reliable segmentation target, and as such, the FIB channel was treated here not as a second segmentation channel, but as a conditional intranuclear measurement channel interrogated strictly within the bounds of the DAPI-derived masks.

More explicitly, let $\mathcal{N}_j$ denote the DAPI-derived mask for nucleus j, and let $I_{\text{FIB}}(p)$ denote the FIB intensity at pixel $p \in \mathcal{N}_j$. The full intranuclear FIB pixel vector was first extracted for each nucleus:

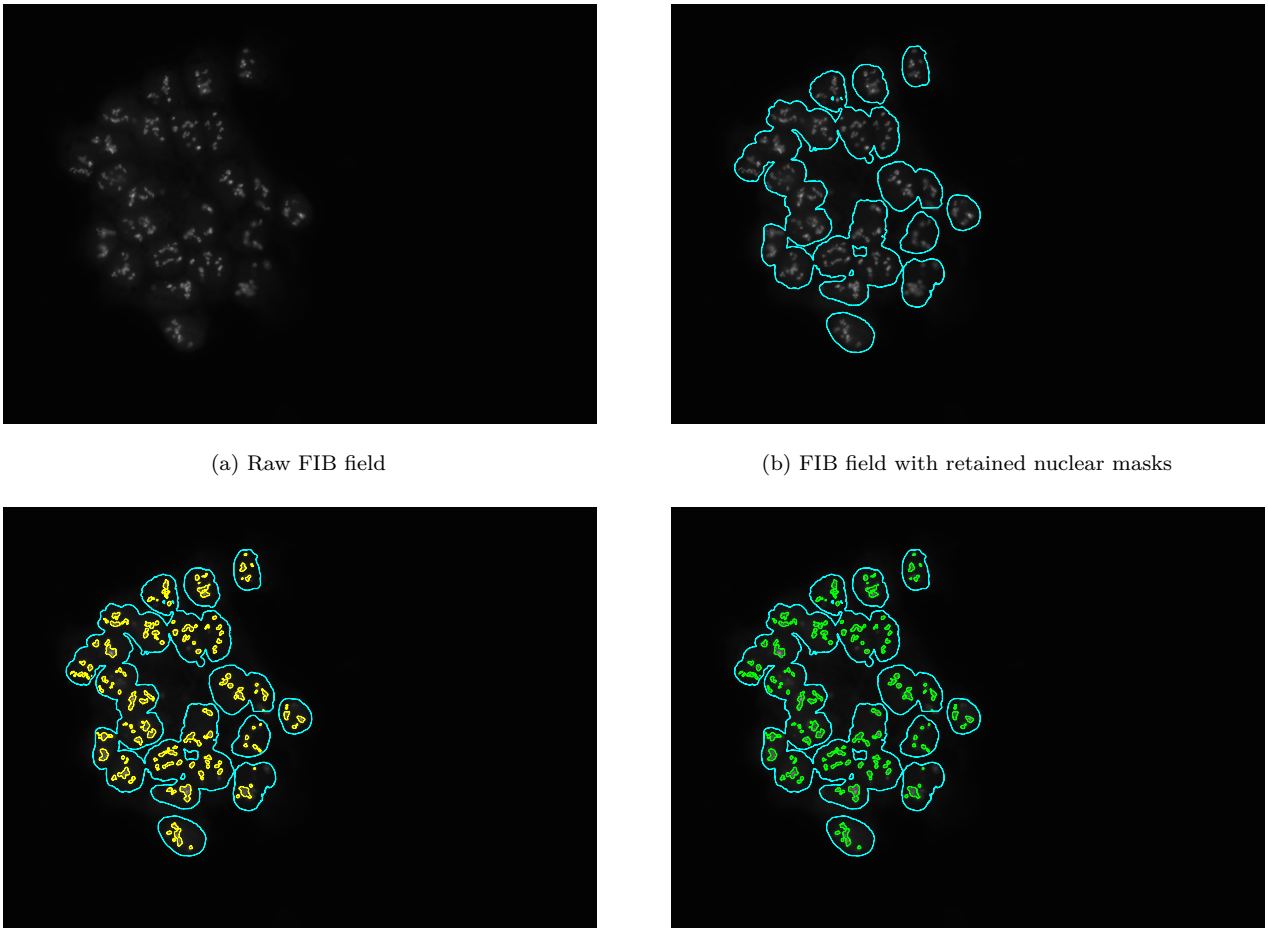
$$\mathcal{F}_j = \{I_{\text{FIB}}(p) : p \in \mathcal{N}_j\}, \tag{2.16}$$

from which the corresponding total intranuclear FIB intensity was calculated as

$$I_{\text{FIB},j} = \sum_{p\in\mathcal{N}_j} I_{\text{FIB}}(p). \tag{2.17}$$

Nucleolar extraction was then performed conditionally from the pooled intranuclear FIB intensities in each image. First, all FIB pixels lying inside all DAPI-derived nuclear masks for that image were pooled, consistent with the sequence illustrated in Figure 2.6. Specifically, all FIB pixels lying inside all retained DAPI-derived nuclear masks for that image were pooled, and a three-class multi-Otsu threshold was computed from that pooled intranuclear distribution [317]. If the resulting thresholds are denoted T_1 and T_2, with $T_1 < T_2$, then the implemented global nucleolar threshold was not taken as the upper Otsu threshold itself, but as a relaxed intermediate threshold,

$$T_{\text{global}} = T_1 + 0.73\,(T_2 - T_1). \tag{2.18}$$



(a) Raw FIB field

(b) FIB field with retained nuclear masks

(c) Global FIB nucleolar-candidate mask

(d) Final per-nucleus nucleolar assignment

Figure 2.6: Continuation of the representative Stage 1 processing sequence shown in Figure 2.5, using the same permeabilised PNT2-C2 DAPI+FIB field of view and the same retained DAPI-derived nuclear masks. This figure corresponds to the conditional paired-FIB branch of Figure 2.4. (a) shows the raw paired fibrillarin (FIB) image. (b) shows the retained DAPI-derived nuclear masks overlaid onto that field, defining the intranuclear search region for nucleolar assignment. (c) shows the nucleolar-candidate pixels obtained from a single global intranuclear FIB threshold. (d) shows the final per-nucleus nucleolar assignment following application of the nucleus-level guardrail rule. Where no nucleus required threshold replacement, panels (c) and (d) may appear very similar. Exact implementation parameters are provided in Section A.1.

Pixels satisfying $I_{\mathrm{FIB}}(p) > T_{\mathrm{global}}$ were classified as nucleolar-candidate pixels, giving the global candidate-assignment stage shown in Figure 2.6(c). This relaxation step was used because a threshold set exactly at the upper Otsu boundary was found to be excessively restrictive for the present fibrillarin signal, isolating only the most extreme bright tail of the intranuclear distribution rather than the broader nucleolar compartment of interest. If multi-Otsu thresholding failed for a given image, a percentile-based fallback threshold was used, as detailed in Section A.1. While the relaxed threshold used here showed marked improvement over the unmodified upper Otsu threshold, some tendency remained to underestimate the full extent of nucleolar-associated pixels because thresholding was still performed on a per-image basis. A further nucleus-level guardrail reassignment step was therefore applied after this global thresholding stage, as shown in Figure 2.6(d) and detailed in Section A.1.

Denoting the final threshold used for nucleus j by T_j, the nucleolar pixel set was therefore defined as

$$\mathcal{U}_j = \{p \in \mathcal{N}_j : I_{\mathrm{FIB}}(p) > T_j\}. \tag{2.19}$$

Nucleolar area was then calculated as

$$\mathrm{Area}_{\mathrm{nucleoli},j} = |\mathcal{U}_j|\, s^2, \tag{2.20}$$

where s is the calibrated pixel size from Section 2.2, and nucleolar FIB intensity was calculated as

$$I_{\mathrm{nucleoli},j} = \sum_{p \in \mathcal{U}_j} I_{\mathrm{FIB}}(p). \tag{2.21}$$

This procedure should be interpreted as an intensity-based subnuclear partitioning step, rather than a fully independent segmentation procedure on the FIB channel. In other words, the code does not attempt to identify nucleoli as freestanding image objects across the whole field of view. Instead, it asks a more constrained question that is better matched to the experimental design: given a nucleus already defined from DAPI, which subset of the intranuclear fibrillarin signal lies in the bright nucleolar regime? This approach preserves a common nucleus identity across channels, avoids circularity in whole-nucleus definition, and yields per-nucleus FIB and nucleolar pixel exports that can be consumed directly by the downstream pooled and repeated random subsampling analyses.

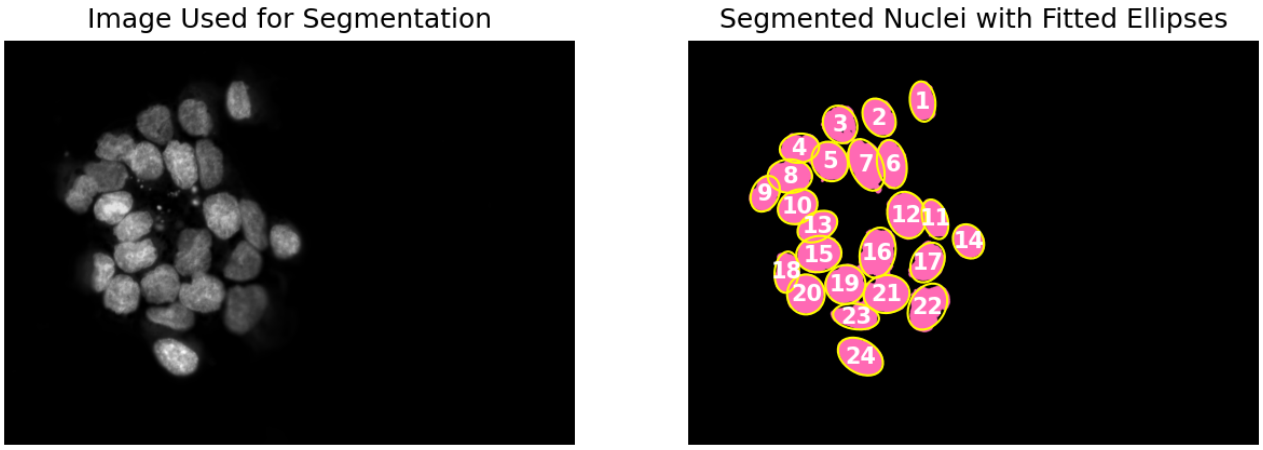
Output generation and quality control

The Stage 1 segmentation script produced a set of both scalar and pixel-resolved outputs, matching the terminal export blocks shown in Figure 2.4. For each processed image, a per-nucleus CSV table was written containing the principal scalar descriptors required at the segmentation stage, namely: projected nuclear area, aspect ratio, aspect-ratio error, DAPI density, and nucleolar area where a paired FIB image was available. In parallel, per-nucleus pixel-block text files were written for the raw DAPI nuclear pixel vectors, and where FIB data were present, two additional pixel-block exports were generated: the full intranuclear FIB pixel vector for each DAPI-defined nucleus, and the nucleolar-only FIB pixel vector obtained after thresholding. This distinction is important for the downstream analysis, because it permits later stages to analyse both the full nuclear FIB distribution and the restricted nucleolar compartment separately, rather than collapsing both into a single summary value at the segmentation stage.

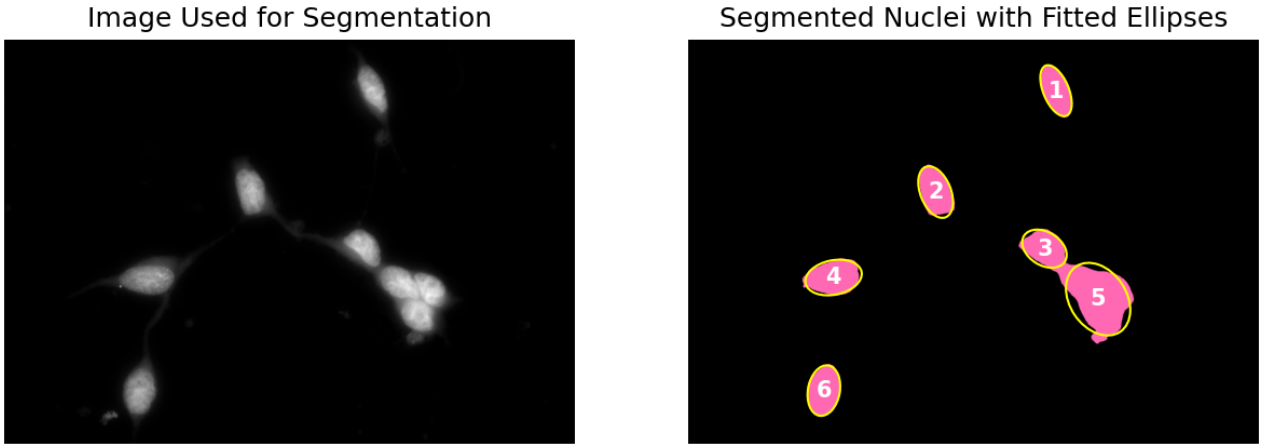
Bright and dark DAPI intensity ratio exports were also written with the threshold values included in the file header for reproducibility. In addition, threshold diagnostics for the FIB/nucleolar pathway were appended to a shared CSV log, including the threshold value used for each nucleus and the decision path by which it was obtained (global relaxed multi-Otsu threshold or percentile-based guardrail/fallback), and was included so that threshold behaviour could be audited explicitly, particularly in nuclei with weak, unusually diffuse, or unusually extensive fibrillarin signal.

Quality-control outputs were also generated for each image as two-panel figures showing the pre-processed field on the left, alongside the labelled nuclear instances with fitted ellipses and nucleus identifiers overlaid on the right. These QC images were used to verify that the segmentation, watershed separation, and ellipse fitting behaved plausibly across different fields and preparation groups, two examples are shown in Figure 2.7. The principal QC burden at this stage remained on the DAPI-derived masks, because all subsequent FIB and nucleolar measurements inherit those nuclear definitions. However, the FIB pathway was also constrained by the saved threshold diagnostics and by the availability of full FIB and nucleolar-only pixel exports, allowing atypical thresholding behaviour to be

traced back and checked where required.



(a) Representative successful Stage 1 DAPI-driven segmentation in a PNT2-C2 DAPI+FIB field. The left panel shows the pre-processed DAPI image used for segmentation; the right panel shows the corresponding quality-control overlay, with each retained nucleus labelled in pink (and numbered in white), and its fitted ellipse superimposed in yellow. In this example, the nuclei are sufficiently well separated in intensity/topography that the foreground thresholding, distance-transform marker detection, and marker-controlled watershed steps recover one nucleus instance per visually identifiable nucleus, with ellipse fits that closely follow the projected nuclear outlines. This image is therefore representative of the typical intended operating regime of the pipeline.



(b) Representative under-segmentation failure in a permeabilised LNCaP DAPI-only field. The left panel shows the pre-processed DAPI image used for segmentation; the right panel shows the corresponding quality-control overlay. In the lower-right region, diffuse extranuclear DAPI signal and reduced contrast at the inter-nuclear boundaries obscure the true separation between three neighbouring nuclei. As a result, the watershed stage returned these nuclei as a single merged object rather than three distinct instances. This example illustrates a principal failure mode observed in the present workflow: local boundary obfuscation, whether from overlap, leakage, or weak boundary contrast, leading predominantly to under-segmentation.

Figure 2.7: Representative quality-control outputs from the Stage 1 DAPI-derived nuclear segmentation pipeline, showing both typical successful behaviour and a principal observed failure mode. For each field, the left image is the DAPI image used for segmentation after pre-processing, and the right image is the corresponding QC overlay showing the retained labelled nuclei and fitted ellipses. Panel (a) illustrates a field in which the pipeline recovers the visually apparent nuclei with good one-to-one correspondence. Panel (b) illustrates a challenging case in which obscured inter-nuclear boundaries lead to merging of adjacent nuclei into a single detected object. Together, these examples show that the pipeline performs well in typical fields, but remains susceptible to under-segmentation in regions of overlap, diffuse extranuclear staining, or low local boundary contrast.

In addition to routine visual inspection of the automatically generated QC outputs, a pragmatic manual review was performed on a random subset of 25 fields drawn across all datasets. These 25 images contained 234 visually identifiable nuclei in total, while the algorithm returned 230 detected nuclear objects. 15 of the 25 images were segmented with full apparent accuracy. The remaining images accounted for the observed discrepancies, which occurred predominantly in challenging cases including closely packed or overlapping nuclei, diffuse extranuclear staining, and very low-intensity nuclei. When summarised as the number of poorly segmented detected objects divided by the total

number of detected objects, the approximate segmentation error rate in this reviewed subset was $^{14}/_{230} \approx 6\%$. This should be interpreted as a practical audit of segmentation performance, rather than as a formal benchmark study. Nonetheless, it provides an empirical check that the dominant behaviour of the pipeline is acceptable for large-batch nucleus-level analysis, and in fact performance is similar to many alternative segmentation pipelines [328]. Note however, that due to the small size of the manually tested subset in this study, the true error rate is likely not fully represented by this manual assessment, so this assessment is tentative at best.

The representative examples shown in Figure 2.7 illustrate the intended operation of the segmentation code, as well as a common failure mode. Figure 2.7(a) shows a successful example from a PNT2-C2 DAPI+FIB field, in which the nuclei were recovered with the expected one-to-one correspondence, and with ellipse fits that closely follow the true nuclear outlines. By contrast, Figure 2.7(b) shows a representative failure from a permeabilised LNCaP DAPI-only field, in which diffuse extranuclear staining partially obscured the true nuclear boundaries, causing the three adjacent nuclei to be returned as a single merged object. The dominant observed failure mode in this instance was therefore under-segmentation, rather than spurious fragmentation of single nuclei into multiple smaller objects (over-segmentation), however that is not to say that over-segmentation did not occur; the dominant failure mode depends on the exact nature of each field of view, with over-segmentation more commonly occurring in very dim nuclei.

This distinction is methodologically relevant for the interpretation of the downstream statistics. Under-segmentation of adjacent nuclei tends to reduce the apparent nucleus count locally while inflating the measured area and integrated intensity of the merged object, and can therefore introduce a small number of outlier measurements into the pooled distributions. The reviewed subset, however, suggests that such failures were infrequent relative to the total number of analysed nuclei, and occurred primarily in visually difficult fields, rather than as a systematic/global behaviour across the dataset. Since the principal analyses in this thesis were based on large pooled nucleus counts, the nuclear segmentation errors are not expected to dominate the overall distributional results, though they remain an explicitly recognised source of residual uncertainty and a limitation of the method here.

Qualitatively, segmentation failures appeared somewhat more common in LNCaP than in PNT2-C2 fields. One plausible explanation is that LNCaP nuclei, particularly in permeabilised preparations, were more prone to boundary degradation or leakage-like diffuse staining, thereby reducing the local contrast needed for reliable watershed separation [295]. While this interpretation is supported by observed alterations in nuclear membranes and lipid membrane composition in prostate cancer, this interpretation was not independently validated, and is therefore presented cautiously as an observational tendency rather than a demonstrated mechanistic conclusion [329–331].

The outputs of Stage 1 thus comprised:

1. per-nucleus scalar descriptor tables
2. per-nucleus DAPI pixel vectors
3. DAPI quality control images

Where applicable, the workflow also generated:

4. per-nucleus full-FIB pixel vectors

5. nucleolar-only FIB pixel vectors

6. threshold-diagnostic logs

These files provided the sole inputs for subsequent pooled descriptive, repeated random subsampling, distributional, and convergence analyses described in later parts of Section 2.3.

2.3.4 Stage 2: pooled descriptive analysis and repeated random subsampling convergence-series generation

Stage 2 utilised the per-nucleus outputs generated by Stage 1 and transformed them into two distinct classes of downstream product: pooled descriptive summaries of nucleus-level distributions, and resampling series with an increasing number of nuclei (N) used for later convergence modelling, with both sets being constrained within each biological group. Stated clearly, Stage 1 defined the nuclei and exported their scalar and pixel-level measurements; Stage 2 then regrouped those nuclei by cell line and preparation condition, constructed the descriptive metrics required for subsequent interpretation, and generated the files required by the distribution fitting and convergence testing stages. The second workflow therefore begins where the first ends: the Stage 1 segmentation/extraction outputs form the direct inputs to the Stage 2 pooled-analysis and resampling procedures. The algorithmic representation of this analysis step (Figure 2.8) emphasises two branches: collective mode and statistics mode, each serving a specific purpose.

Inputs, parsing, and per-nucleus alignment

The second script consumed the Stage 1 outputs (Section 2.3.3) on a per-sample basis. The minimum required inputs included a per-nucleus CSV table containing the scalar morphological/intensity descriptors exported by the segmentation script, a per-nucleus DAPI pixel-block text file, and a bright/dark ratio text file containing the intensity-weighted outputs and associated threshold metadata. If FIB analysis had been performed in Stage 1, the corresponding per-nucleus full-FIB and nucleolar-FIB pixel-block text files were also loaded. Pixel-block files were parsed as nuclei separated by blank lines. Alignment between scalar tables and pixel-block files was preserved by nucleus order within each sample, which was strictly enforced using a global count for all outputs generated from a given image. More explicitly, if sample s contained nuclei $j = 1, \ldots, n_s$, then all scalar and pixel-array objects were indexed by the same within-sample nucleus ordering, allowing full sample provenance (s, j) to be retained throughout later pooling and resampling operations.

A small number of derived per-nucleus quantities were constructed in Stage 2 rather than Stage 1, because they depended on the pixel-block exports rather than the segmentation mask alone, including the mean, standard deviation, and coefficient of variation of the DAPI nuclear pixel vector, and, where applicable, the corresponding statistics for the full FIB and nucleolar FIB pixel vectors. Nuclei with fewer than two valid pixels were excluded from statistics requiring a variance estimate, and such exclusions were tracked explicitly rather than being dropped silently.

Stage 2 also constructed several per-nucleus ratio metrics used later in the descriptive and convergence analyses. If Area_j is the nuclear area, $\text{Area}_{\text{nucleoli},j}$ the nucleolar area, $I_{\text{DAPI},j}$ the total nuclear DAPI intensity, $I_{\text{FIB},j}$ the total nuclear FIB intensity, and $I_{\text{nucleoli},j}$ the total nucleolar FIB intensity, then

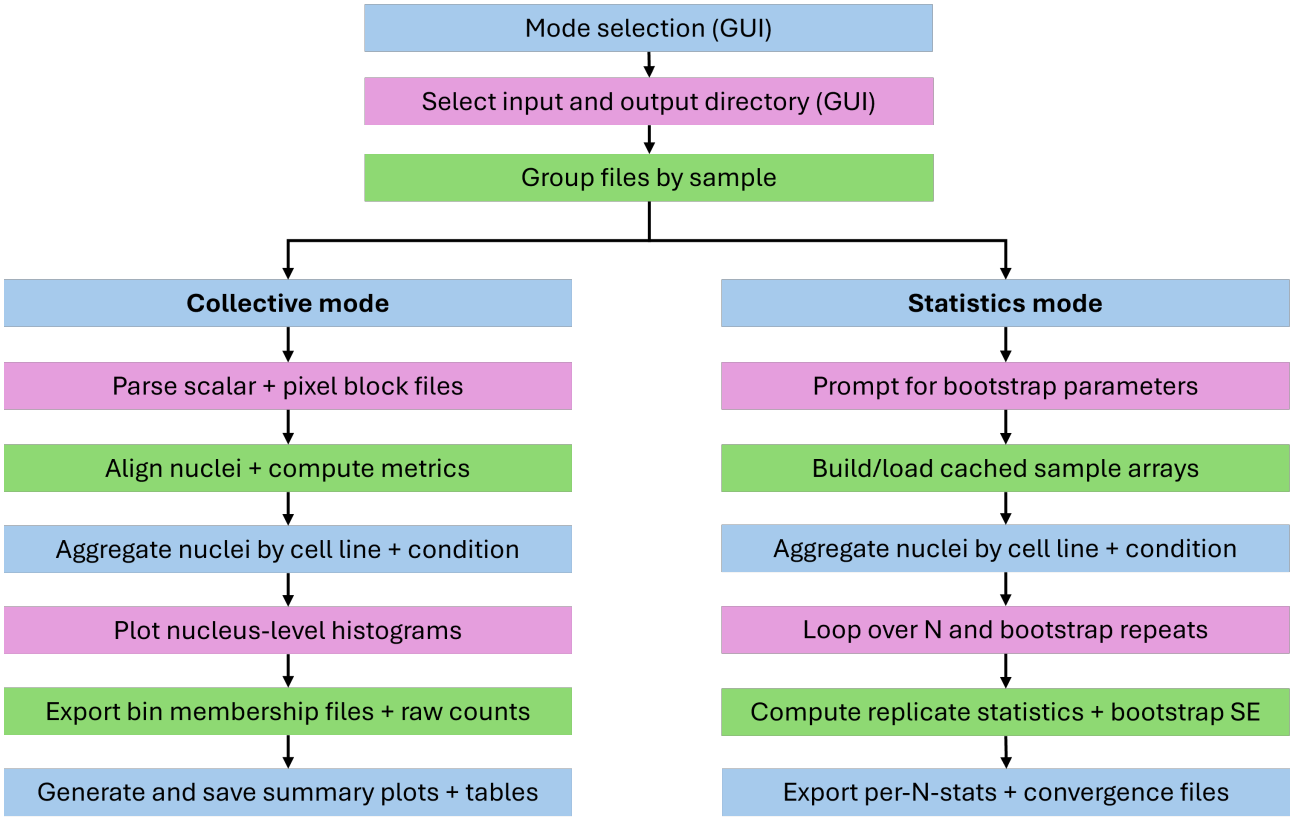


Figure 2.8: Workflow for the Stage 2 analysis script, which consumes the Stage 1 (Section 2.3.3) segmentation/extraction outputs and converts them either into pooled descriptive summaries (Collective mode) or convergence series (Statistics mode). After selecting the mode of operation and input files/output directory using a GUI, files are grouped by sample. In collective mode, the script parses the scalar and pixel-block exports from Stage 1, aligns nuclei across the different file types (as pixel-block files were created per image, with pixels from individual nuclei in each file separated by a blank line), computes derived per-nucleus metrics, aggregates nuclei by cell line and preparation condition, generates pooled histograms and summary plots, and finally exports both bin-membership files and raw pooled-data tables (i.e. the raw data from all loaded nuclei in a given cell line and condition) for downstream distributional modelling. In statistics mode, the script prompts for the resampling parameters, builds or loads cache files to reduce overhead from repeated parsing, aggregates nuclei by group, loops over nucleus count N and resampling repeats, computes replicate-level statistics and their across-replicate standard errors, and exports per N summary files together with the convergence-series files used for subsequent convergence testing.

the principal ratios were

$$R_{\mathrm{N:NA},j} = \frac{\mathrm{Area}_{\mathrm{nucleoli},j}}{\mathrm{Area}_j}, \tag{2.22}$$

$$R_{\mathrm{D:F},j} = \frac{I_{\mathrm{DAPI},j}}{I_{\mathrm{FIB},j}}, \tag{2.23}$$

$$R_{\mathrm{N:D},j} = \frac{I_{\mathrm{nucleoli},j}}{I_{\mathrm{DAPI},j}}. \tag{2.24}$$

The bright and dark ratios imported from Stage 1 (Equations (2.14) and (2.15)) were retained unchanged.

Collective mode

Collective mode was the principal descriptive branch used for pooled nucleus-level summaries. Following the left branch of Figure 2.8, the script first parsed the scalar and pixel-block files produced by Stage 1, then aligned nuclei across the different file types, computed the derived per-nucleus metrics as

described above, and finally linked nuclei across all samples belonging to the same biological group (i.e. cell line and preparation condition). In this mode, the sampling unit remained the individual nucleus: images and samples served only as containers by which to trace nuclei back to their origin, while the pooled descriptive distributions were formed directly from the nucleus-level values.

Histograms were generated primarily for visual inspection of pooled distributions, with the full distributional analysis being performed through a separate script later in Stage 3 (Section 2.3.5). For each histogram, a corresponding raw text export of unique values and associated counts was written. This export format is methodologically important because it allowed Stage 3 to operate on weighted value-count tables rather than requiring computationally expensive reconstruction of the full pooled nucleus list. Bin-membership files were also available, explicitly mapping each nucleus (sample, nucleus ID) to its histogram bin index (as well as giving the exact value). These files were used for auditing tails, outliers, or any subpopulations identified in the pooled plots.

Collective-mode summary plots included pooled histograms, box plots, and error-bar plots. Error-bar plots reported group means, KDE modes (defined formally in Section 2.3.5), and CVs, with the standard error plotted as uncertainty, which describes the precision of the estimated statistic. If $x_1, \ldots, x_n$ denotes the nucleus-level values of a metric in a pooled group, with mean $\bar{x}$, then

$$\mathrm{SD} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}, \tag{2.25}$$

$$\mathrm{SEM} = \frac{SD}{\sqrt{n}}, \tag{2.26}$$

$$\mathrm{CV} = 100 \times \frac{SD}{\bar{x}}. \tag{2.27}$$

For derived statistics where standard errors could not be calculated directly, they were approximated using the first-order delta method [332, 333]. If $T = g(\hat{\boldsymbol{\theta}})$, where $\hat{\boldsymbol{\theta}}$ has estimated covariance matrix $\widehat{\Sigma}_{\hat{\boldsymbol{\theta}}}$, then

$$\mathrm{SE}(T) \approx \sqrt{\nabla g(\hat{\boldsymbol{\theta}})^{\mathsf{T}}\, \widehat{\Sigma}_{\hat{\boldsymbol{\theta}}}\, \nabla g(\hat{\boldsymbol{\theta}})}. \tag{2.28}$$

This approximation was used for statistics derived from pooled nucleus-level measurements, including the coefficient of variation across nuclei and KDE-based mode estimates. For CV expressed as a percentage (Equation (2.27)), the approximation used here was

$$\mathrm{SE}(\mathrm{CV}) \approx \mathrm{CV}\sqrt{\frac{1 + 2\,(\mathrm{CV}/100)^2}{2n}}. \tag{2.29}$$

For KDE modes, the standard error was approximated by applying the same first-order argument to the estimated mode location using the local derivative structure of the fitted KDE.

For box plots, the interquartile range (IQR) was reported as

$$\mathrm{IQR} = Q_3 - Q_1, \tag{2.30}$$

where Q_1 and Q_3 are the first and third quartiles, respectively. These quantities describe different aspects of the pooled nuclear distributions: SD reflects nucleus-level spread, SEM reflects mean-estimation precision, CV is a normalised, dimensionless measure of the relative spread in a given metric (and as a

result often more directly interpretable than SD), and IQR provides a robust summary of the central half of the data.

Summary tables generated in collective mode contained the group mean, median, SD, SEM, CV, and IQR for each metric. It is important at this point to distinguish the role of histograms in Stage 2 from the role of Stage 3 modelling. In collective mode, histograms were used primarily for diagnostic visualisation, to generate the bin-membership files, and for exporting value-count tables for use in subsequent analysis. They were not treated as the inferential target of the later distribution fitting. Instead, the raw weighted exports from collective mode served as the starting point for the kernel density based modelling described in Stage 3. Collective mode also included an optional fixed-N equalisation procedure, using a single reproducible subsample to control for unequal nucleus counts across groups (typically, more PNT2-C2 nuclei were available than LNCaP in a given condition, and so this functionality was used to determine if any observed changes were simply an artefact of this sampling inequality).

For all histogram-based analyses, the pooled distributions were interrogated at the level of individual nuclei by tracing nuclei back from defined regions of the histograms or KDEs to the source image data, in order to inspect obvious segmentation failures, merged objects, and visually coherent subpopulations. Representative traceback-image panels were not shown exhaustively for every histogram in the final thesis, and instead were included selectively for cases in which direct visual inspection materially informed interpretation of tails, multimodality, or artefactual structure.

Statistics mode

Statistics mode was the right-hand branch of Figure 2.8 and generated the convergence-series files later consumed by Stage 4 (Section 2.3.6). Its purpose was to estimate how the uncertainty of key summary statistics evolved as an increasing number of nuclei were included in their calculation. The user specified the maximum nucleus count $N_{\max}$, the increment ΔN, the number of resampling repeats R, and an optional seed for reproducibility (which was randomly generated if unspecified). The script then constructed the sequence

$$N \in \{\Delta N,\, 2\Delta N,\, 3\Delta N, \ldots, N_{\max}\},$$

subject to the availability of at least N nuclei in the pooled group. Where the requested $N_{\max}$ exceeded the total available nuclei in a given dataset (N_{total}), $N_{\max}$ was automatically set to N_{total}.

To reduce repeated parsing and RAM overhead, statistics mode could first build or load cache files containing the precomputed scalar arrays and associated sidecar metadata needed to fully map all available nuclei, and reliably reconstruct nucleus-level values without necessitating keeping the full dataset in memory. This caching layer was important because statistics mode was the most computationally demanding branch of the second script by a large margin. After aggregation by group, the script then looped over the requested N values and performed repeated resampling from the pooled nuclei at each N. Within each replicate, N nuclei were drawn without replacement from the available pool for that group, and the required summary statistics were computed from those N nuclei. Sampling without replacement was chosen deliberately, because the objective here was to quantify within-dataset convergence of the estimated statistics as increasing numbers of the available nuclei were included, rather than to approximate repeated sampling from an inferred superpopulation

[334, 335]. Given this, selection without replacement was the most appropriate choice; it preserved the finite-pool structure of each dataset, prevented duplicate inclusion of the same nucleus within a given replicate, and ensured that resampling variability necessarily approached zero as $N \to N_{\text{total}}$ [334]. By contrast, sampling with replacement would have permitted repeated selection of the same nucleus and omission of others, even when $N = N_{\text{total}}$, which would not have matched the intended interpretation of the analysis [333]. Repeated random subsampling across all requested N values therefore provided an empirical assessment of how rapidly each statistic stabilised as a function of nucleus count within the realised dataset, and gave more reliable and repeatable inference than simply using one random selection of nuclei by reducing the dependence of results on any single arbitrary selection (which may give especially low or high standard error values purely by random chance) [334, 336].

For each replicate and each metric, the script computed quantities including the mean, SD, SEM, median, IQR, and coefficient of variation (CV). Across the R replicates at fixed N, the script stored both the mean of the replicate statistic and the standard deviation of that statistic across replicates. The latter quantity was interpreted operationally as a "subsampling standard error" for the statistic. For resampling repeat $r = 1, \ldots, R$ at fixed nucleus count N, let

$$\mathrm{SEM}_r(N) = \frac{\sigma_r(N)}{\sqrt{N}}, \tag{2.31}$$

where $\sigma_r(N)$ is the sample standard deviation within repeat r. The convergence series was then

$$\overline{\mathrm{SEM}}(N) = \frac{1}{R} \sum_{r=1}^{R} \mathrm{SEM}_r(N), \tag{2.32}$$

with the repeat-to-repeat spread (or subsampling standard error)

$$\sigma_{\mathrm{SEM}}(N) = \sqrt{\frac{1}{R-1} \sum_{r=1}^{R} \left[\mathrm{SEM}_r(N) - \overline{\mathrm{SEM}}(N)\right]^2} \tag{2.33}$$

being retained separately and used as error bars for the mean standard error convergence plots created later in Stage 4 convergence modelling.

Outputs from statistics mode were written both per N and as aggregated convergence series. At the individual N level, each N directory contained per-group CSV summaries listing each statistic, and their associated mean and SE across replicates. After all N values had been processed, master convergence files containing $\overline{\mathrm{SEM}}(N)$ and $\sigma_{\mathrm{SEM}}(N)$ for each metric were exported with one entry for each N step. These files provided the sole numerical inputs to Stage 4. Statistics mode therefore acted as the bridge between pooled nucleus-level measurements and the later formal convergence analysis: it converted the static pooled distributions of collective mode into explicit standard error decay trajectories for each descriptor.

2.3.5 Stage 3: KDE estimation and parametric distribution fitting

Stage 3 was implemented as a distinct downstream modelling step and did not operate directly on the raw microscope images or segmentation masks, rather it consumed the raw weighted histogram exports generated by collective mode in Stage 2, with each file containing unique values x_i and associated counts w_i. The data were filtered to retain only finite x_i and strictly positive weights, and the total

weight

$$W = \sum_i w_i$$

was used as the effective pooled nucleus count for the distribution under analysis.

The first modelling layer was weighted kernel density estimation (KDE), used to obtain a smooth non-parametric estimate of the underlying distribution [258]. With Gaussian kernel K and bandwidth h, the weighted KDE was

$$\hat{f}_h(x) = \frac{1}{Wh} \sum_{i=1}^{m} w_i \, K\left(\frac{x - x_i}{h}\right), \tag{2.34}$$

where m is the number of unique exported values. In the implemented workflow, the KDE was evaluated on a uniform grid, and the default bandwidth followed the Scott-rule implementation [316, 337]. Peak finding on the evaluated KDE grid yielded the number and positions of visually distinctive modes. The principal KDE-derived modal descriptor reported was the mode, defined as the grid location maximising $\hat{f}_h(x)$. Numerical integration of the KDE was also carried out for percentile or interval-mass calculations where required.

The second modelling layer comprised parametric mixture fitting. Candidate distribution families were Gaussian, skew-normal, log-normal, Gamma, log-logistic (Fisk), and inverse Gaussian, with mixture sizes $k = 1, \ldots, K_{\max}$ [338–341]. "Mixture sizes" here refers to the total number of components needed to properly characterise the data (for example if 3 Gaussian distributions best described the data, then $k = 3$). Fitting was performed to the KDE grid rather than directly to the raw weighted value-count pairs or histogram, such that the smooth KDE estimate served as the empirical target density. Initial component placement used weighted pseudo-sampling and/or peak-informed initialisation, with quantile-based fallback initialisation when peak information was insufficient. Parameter estimation then proceeded through an expectation-maximisation (EM) based mixture-fitting strategy, with bounded numerical optimisation of the component parameters, weighted responsibility updates, and explicit rescue of dead or numerically degenerate components [342, 343]. The central aim of this stage was not to enforce any single canonical family, but to directly study the distribution of the various descriptors reported here, and to provide a reproducible and auditable model-selection framework.

Model comparison used three information criteria: the AIC defined earlier, the Bayesian information criterion (BIC), and the integrated completed likelihood (ICL) [259, 344]. In the present workflow, ICL was used as the definitive selection method, because it penalises not only parameter count (i.e. complexity), but also uncertain component assignment, and therefore tends to favour more conservative mixture structures than AIC or BIC when component overlap is substantial. RMSE values against both the KDE and histogram were also recorded as diagnostic measures of curve-tracking performance, but the primary model recommendation logic was based on the information criteria rather than on visual fit quality alone. Where several candidate models lay within a small Δ ICL range of the best-scoring model, the smaller k model was preferred as a conservative recommendation.

Stage 3 outputs consisted of a JSON report containing all candidate models and diagnostics, a summary CSV reporting the best and recommended models under the different information criteria, a best-fit component table giving component parameters, standard errors, and mixture weights π, and a diagnostic PDF plot overlaying the histogram, KDE, fitted mixture sum, and individual component curves. Overall, this stage converted the pooled descriptive exports from Stage 2 into

explicit descriptions (both parametric, captured by the distribution fitting, and non-parametric, by the empirical KDE) of the observed nucleus-level distributions, while keeping the modelling assumptions and model-selection criteria transparent and auditable.

2.3.6 Stage 4: convergence modelling and convergence-point estimation

Stage 4 consumed the convergence-series files exported by statistics mode and fitted explicit models to the SEM vs. N trajectories. The rationale behind the analysis carried out in this stage was twofold. First, it provided a formal way to determine whether the number of analysed nuclei was sufficient to reach a given level of precision for a chosen descriptor. Second, it supplied a descriptor-specific summary of relative heterogeneity, realised through the rate at which uncertainty on the estimated mean decayed as additional nuclei (N) were included (note that here and throughout this study, "rate" is strictly in terms of N, not time, i.e. a greater reduction in SEM per increase in N would give a greater "rate of decay") [127, 231, 345]. The central assumption underpinning this interpretation is that nuclei behave as approximately independent draws from a broader group-specific distribution, at least well enough for the scaling of SEM to be informative [346, 347]. This assumption is imperfect, and treated cautiously; however, it provides a useful framework for comparing descriptors and groups within the present study design.

For approximately independent observations with finite variance σ_x^2, the classical scaling relation for the sample mean is

$$\mathrm{Var}(\bar{x}(N)) = \frac{\sigma_x^2}{N}, \tag{2.35}$$

$$\mathrm{SEM}(N) = \frac{\sigma_x}{\sqrt{N}}, \tag{2.36}$$

(where Equation (2.36) is rewritten in a slightly simpler form from Equation (2.31)) such that SEM decays proportionally to $N^{-1/2}$ [346–350]. In the workflow implemented here, the primary convergence series carried forward from Stage 2 (Section 2.3.4) were not analytic SEM values computed from a single sample of size N, but empirical estimates of SEM obtained by repeated random subsampling of the pooled nucleus set at increasing N (Equation (2.32)). While convergence was also assessed using a single random draw (i.e. only sampling once at each N), the replicated resampling results were favoured in this study for the reasons outlined in Section 2.3.4. More explicitly, for each resampling repeat r at nucleus count N, the SEM of the descriptor was calculated within that resample, and the quantity propagated to the convergence analysis was the mean of these SEM values across repeats. The repeat-to-repeat spread was retained separately, and used as an uncertainty measure on the estimated mean-SEM curve (Equation (2.33)). Under approximate independence, these empirical mean SEM values play the same operational role as an uncertainty estimate on the sample mean, and were therefore modelled within the same monotonically decreasing power-law framework:

$$\mathrm{SEM}(N) = AN^{pow}, \tag{2.37}$$

where $\mathrm{SEM}(N)$ denotes the uncertainty series, A is the amplitude, and pow is the fitted scaling exponent. Under the classical independent finite-variance expectation, $pow = -\frac{1}{2}$ [346, 348–350]. The amplitude A is proportional to the underlying descriptor dispersion, and, when $pow \approx -\frac{1}{2}$, A is proportional to the underlying nucleus-level dispersion, σ_x (i.e. the standard deviation, according to

Equation (2.36)). The inverse of the fitted amplitude A,

$$\gamma = \frac{1}{A}, \tag{2.38}$$

was used as a practical convergence-rate proxy, such that larger γ implies faster decay of uncertainty with increasing N and therefore lower effective dispersion for that descriptor. Note again that "rate", and "faster" are grounded strictly in terms of N, and only serve as an intuitive description of the SEM decay; "faster" convergence/decay means a greater reduction in SEM per increase in N. It must also be stated that this interpretation is approximate rather than absolute: deviations of pow from $-\frac{1}{2}$ may result from a variety of factors including inter-nuclear dependence/correlation, non-stationarity across the pooled sample, or residual imaging/processing artefacts, as well as genuine biological heterogeneity.

The convergence criterion used in the thesis was based on the percentage standard error of the mean,

$$\%\mathrm{SEM}(N) = 100 \times \frac{\overline{\mathrm{SEM}}(N)}{\bar{x}(N)}, \tag{2.39}$$

where $\overline{\mathrm{SEM}}(N)$ is the mean resampled SEM estimate for descriptor x at nucleus count N, and $\bar{x}(N)$ is the corresponding mean. A descriptor was considered converged at the first N for which $\%\mathrm{SEM}(N) \leq 2.5\%$, and remained below this threshold for all subsequent sampled N. This criterion was chosen because it provides a direct, dimensionless statement of the relative estimate precision, therefore allowing descriptors to be compared even across different absolute scales [301]. An alternative stability-based criterion based on successive percentage changes in the uncertainty series was also considered, however the convergence point defined in Equation (2.39) was chosen as the definitive rule for use in this thesis.

Fitting in Stage 4 was performed in three principal ways, with the primary fit modality using weighted non-linear least squares in linear space to estimate A and *pow* in Equation (2.37). If $y_i = \overline{\mathrm{SEM}}(N_i)$ denotes the mean SEM value at sampled nucleus count N_i, with associated uncertainty $\sigma_{y,i} = \sigma_{\mathrm{SEM}}(N_i)$, then the fitted parameter vector $\boldsymbol{\theta} = (A,\ pow)^{\mathsf{T}}$ was obtained by minimising

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \sum_{i=1}^{N_{\mathrm{pts}}} \frac{[y_i - AN_i^{pow}]^2}{\sigma_{y,i}^2}. \tag{2.40}$$

Parameter uncertainties were then obtained from the fitted covariance matrix, and uncertainties in fit-derived quantities were propagated using first-order covariance-aware error propagation[351, 352]. A constrained variant with *pow* fixed to -0.5 was also used, allowing the amplitude to be estimated under the ideal theoretical scaling exponent expected for independent averaging [353]. The final fitting method was an additional log-log regression calculated by taking logarithms of Equation (2.37),

$$\log_{10}\mathrm{SEM}(N) = pow \log_{10} N + \log_{10} A, \tag{2.41}$$

where $\log_{10} A$ is the intercept in log space. This transformation was used primarily for visualisation and robustness assessment [354]. A power-law relationship curved in linear space becomes linear in log-log space, such that the fitted slope corresponds directly to the scaling exponent *pow* [354]. This makes it simpler to assess whether the empirical convergence follows an approximately power-law form

across the full sampled range of N, and to identify systematic deviations from linearity in log space that may indicate finite-sample effects, departures from ideal averaging behaviour, or the onset of a measurement floor [354, 355]. Plotting in log-log space also compresses the dynamic range of both axes, improving visual interpretability when SEM(N) changes substantially over the sampled N range. However, because the logarithmic transformation changes the error structure and effectively emphasises relative rather than absolute deviations, the log-log regression was not treated as the primary fit from which the main parameter estimates were taken, but rather as a complementary check on the stability and interpretability of the fit across fitting domains [354].

Goodness of fit metrics were reported in each case, but interpreted descriptively rather than as formal hypothesis tests. If $y_i = \overline{\mathrm{SEM}}(N_i)$ denotes the observed mean SEM value at sampled nucleus count N_i, $\sigma_{y,i} = \sigma_{\mathrm{SEM}}(N_i)$ its associated uncertainty, $\hat{y}_i$ the fitted value, and $r_i = y_i - \hat{y}_i$ the residuals, then

$$SS_{\mathrm{res}} = \sum_i r_i^2, \tag{2.42}$$

$$SS_{\mathrm{tot}} = \sum_i (y_i - \bar{y})^2, \tag{2.43}$$

where SS denotes the sum of squares, and

$$R^2 = 1 - \frac{SS_{\mathrm{res}}}{SS_{\mathrm{tot}}}, \tag{2.44}$$

$$\mathrm{RMSE} = \sqrt{\frac{SS_{\mathrm{res}}}{N_{\mathrm{pts}}}}, \tag{2.45}$$

$$\mathrm{NRMSE} = \frac{\mathrm{RMSE}}{|y_{\mathrm{last}}|}, \tag{2.46}$$

$$\chi^2 = \sum_{i=1}^{N_{\mathrm{pts}}} \left(\frac{y_i - \hat{y}_i}{\sigma_{y,i}} \right)^2, \tag{2.47}$$

$$\chi^2_\nu = \frac{\chi^2}{N_{\mathrm{pts}} - p} \tag{2.48}$$

where N_{pts} is the number of sampled N values, y_{last} is the mean-SEM value at the largest evaluated N, p is the number of fitted parameters, and $\sigma_{y,i}$ is the uncertainty associated with y_i.

Within this study, the primary measure of fit quality is the normalised root mean square error (NRMSE), which provides a direct measure of the typical deviation between the fitted curve and the observed SEM trajectory, expressed relative to the terminal SEM value. Here, the terminal SEM denotes the SEM at the largest evaluated N, and because N was incremented in fixed steps, this terminal N could be slightly smaller than the total number of available nuclei. This normalisation makes NRMSE straightforward to interpret and facilitates comparison across cell lines and metrics defined on different absolute scales.

By contrast, reduced chi-squared (χ^2_ν) is not treated as a primary statistic in this setting. Formal interpretation of χ^2_ν depends on assumptions regarding independence of datapoints, correct specification of the effective degrees of freedom, and an appropriate residual model. These assumptions are not fully satisfied here, because the SEM series were generated by repeated resampling from overlapping nucleus

pools and then fitted with a non-linear model [356]. As such, SEM points may exhibit correlation, and χ^2_ν can therefore be misleading if interpreted as a strict test of model adequacy. In this thesis, it is therefore reported only as a descriptive, weighted indicator of residual magnitude; smaller χ^2_ν values indicate that the fitted curve follows the mean SEM trajectory more closely relative to the estimated pointwise variability, whereas larger values indicate greater deviation from the observed trend.

Similarly, the coefficient of determination, R^2, is not uniquely defined for weighted and/or non-linear regression settings [357]. It is therefore treated as supplementary for the power-law fits performed in linear space. Log-log plots of SEM against N were used separately as an explicit diagnostic for power-law behaviour, since linearity in log space corresponds to power-law scaling in linear space, and in that diagnostic representation, R^2 is more readily interpretable as a measure of linear association.

Uncertainties on the goodness of fit metrics themselves were not propagated or reported; only the pointwise summary statistics of the SEM curve were retained, and the goodness of fit metrics were therefore treated as deterministic summaries computed from the mean SEM trajectory. Overall, these statistics were used only to describe how closely the fitted curve tracked the observed convergence behaviour, rather than to support any formal hypothesis-testing claims regarding model adequacy.

In summary, Stage 4 outputs comprised descriptor-specific plots of standard error vs. N, with fitted power-law curves, and tabulated summaries containing the fitted A, pow, γ, convergence-point estimates, and descriptive goodness of fit metrics. Taken collectively, Stages 2-4 thus provided a complete post-segmentation analysis chain: Stage 2 reorganised the Stage 1 nucleus-level outputs into pooled descriptive and resampled convergence series, Stage 3 characterised the pooled distributions, and Stage 4 quantified how rapidly the uncertainty on those descriptors decayed as progressively larger numbers of nuclei were included.

Quality control and validation procedures have for the most part already been described at the point at which they arise, particularly for Stage 1 segmentation and Stage 4 convergence fitting. This final section therefore serves primarily to consolidate the assumptions and limitations that remain most relevant to interpretation of the reported results.

At the segmentation stage, automated quality-control outputs were generated for each processed field, showing the DAPI image used for segmentation together with the retained labelled nuclei, fitted ellipses, and nucleus identifiers (Figure 2.7). These outputs were reviewed visually to check that foreground detection, watershed separation, and ellipse fitting behaved plausibly across different image types and preparation groups. In addition, a manually reviewed subset of 25 randomly selected fields was used as a pragmatic validation check, as described in Section 2.3.3. This manual review was by no means exhaustive, but it does provide direct evidence that the dominant behaviour of the pipeline was largely acceptable for the intended nucleus-level analyses, especially since the number of nuclei sampled, N, was large [301].

Several working assumptions also regulate the downstream statistical stages. First, nuclei were treated as approximately independent observations when interpreting SEM scaling and fitting the convergence curves in Stage 4 [346–350]. This assumption is not exact, but is sufficient for the present stated purpose of comparing how rapidly descriptor uncertainty decayed with increasing nucleus count. Second, intensity-derived quantities were interpreted comparatively under strictly matched preparation and acquisition settings, rather than as absolute molecular quantities [8, 9]. Third, DAPI was treated as the

sole channel used to define nuclei, while fibrillarin was treated as a conditional intranuclear (nucleolar) measurement channel, explicitly referenced back to the DAPI-defined geometry. Finally, histograms in Stage 2 were used for brief visualisation and export convenience; the distributional inference was based on the exported raw value-count tables, KDE estimates, and parametric models fitted downstream [258, 259, 344].

Only a small number of limitations materially affect interpretation of the computational workflow. The code takes greyscale TIFF inputs rather than proprietary microscope container formats, such that image metadata are not interrogated directly during analysis. Spatial calibration was therefore treated explicitly and applied using the known instrument settings, rather than being read automatically from the source files. In addition, paired DAPI and FIB images were assumed to be sufficiently well registered/aligned for mask transfer, though sequential acquisition means that small cross-channel misregistration cannot be excluded with 100% confidence, however, the motorised imaging system meant that it was not necessary to swap filter cubes or adjust the instrument by hand, and the time between capturing images was almost negligible, so while this is mentioned for completeness, misregistration between channels was almost certainly not a significant issue within this workflow [300]. The most significant impact likely comes from the segmentation procedure, which remained vulnerable in fields containing heavily overlapping nuclei, markedly obscured boundaries, diffuse extranuclear signal, or very low nuclear intensities [310]. The fitted ellipse used to extract aspect ratio is also only an approximation to true nuclear shape in strongly irregular cases [326]. Finally, the statistics mode pipeline detailed in Stage 2 (Section 2.3.4) is computationally demanding, particularly for large resampling runs, and therefore relies on caching and controlled execution in order to remain tractable.

The likely impact of these limitations should, however, be interpreted proportionately. Because the principal analyses in this thesis were based on large pooled nucleus counts and nucleus-level summary distributions, while isolated segmentation imperfections may have some small effects, they are unlikely to dominate the reported results unless they occur systematically/preferentially in one group or metric [301]. Their most plausible effect is instead on the tails of the distributions, on a small number of outlier nuclei, or on descriptors which are particularly sensitive to merged masks or weak boundaries. This is precisely why the workflow retained explicit filtering rules, source tracking, threshold logs, and visual QC figures; not because the pipeline is assumed to be perfect, but because its residual failure modes can thus be inspected and constrained transparently.

Chapter 3

Results: Analysis of non-permeabilised cells

This chapter focuses on quantitative morphometric analysis of nuclei in the non-permeabilised condition, treated as the minimally perturbed fixed-cell baseline for this study. This baseline is used both as a standalone characterisation of cell line nuclear morphology and as the reference against which later permeabilised results are interpreted, such that preparation-induced shifts can be distinguished from real biological differences (Chapter 4). Under this preparation, the analysis is intentionally anchored to the DAPI channel; DAPI provides a robust DNA-associated nuclear signal in intact/fixed cells, whereas antibody-based labels targeting intranuclear components are not expected to access their targets reliably without permeabilisation [8, 11, 12, 202]. This focus is also biologically motivated. In prostate cancer, altered nuclear size, shape, and chromatin organisation comprise long-standing diagnostic and prognostic criteria (often collectively referred to as nuclear atypia/pleomorphism), and are mechanistically linked to genome instability, altered nuclear architecture and envelope mechanics, and chromatin reorganisation [131, 132, 135, 184]. As such, the nucleus is treated here not merely as a convenient segmentation target, but as a biologically informative single-cell characteristic that can be measured reproducibly.

Three main contributions arise from the analysis in this chapter. First, it establishes statistical sufficiency of the nucleus count prior to interpretation, rather than treating adequate sample size as an implicit assumption [301]; this addresses a practical limitation of many quantitative microscopy studies, as sample sizes are often not reported explicitly, and even less commonly is the choice of sample size justified (as shown in Table 1.1). Second, it uses standard error convergence behaviour as a metric-specific proxy for relative heterogeneity as described in Section 2.3.6, which is complementary to more conventional variance-based reporting. Third, it analyses full nucleus-level distributions using kernel density estimation (as outlined in Section 2.3.5) rather than collapsing data to mean values alone, allowing central tendency, dispersion, and distributional structure (including the presence of any potential subpopulations) to be assessed [174].

Operationally, each segmented nucleus contributes descriptors detailing size, shape, and DAPI-derived intensity structure. These descriptors were defined previously in Section 2.3, namely projected nuclear area (Equation (2.7)), nuclear aspect ratio (Equation (2.8)), total DAPI intensity (Equation (2.9)), DAPI density (Equation (2.10)), threshold-based bright and dark pixel ratios (Equations (2.14) and (2.15)), and scalar CV reporting (Equation (2.27)). Intranuclear texture is additionally assessed

using the per-nucleus CV of DAPI pixel intensities, supported by bright and dark pixel fractions that summarise occupation of the high and low intensity tails of the DAPI distribution [9, 312, 358]. This separation is method-aware; it distinguishes descriptors expected to track genome content and/or nuclear scaling from descriptors that more directly reflect spatial organisation of DNA-associated signal under fixed staining and imaging conditions [8, 9, 134].

Only a summary of the analytical roadmap is given here; detailed definitions, thresholds, KDE construction and fitting details were introduced previously through the Methods (Sections 2.3 and 2.3.4 to 2.3.6). The present chapter first evaluates convergence behaviour and sampling sufficiency (Sections 3.1 and 3.1.3), then compares nuclear morphology between cell lines using both summary statistics and full distributions (Section 3.2.1 and Figures 3.8 to 3.10, 3.12 and 3.13), and finally analyses DAPI-derived intensity structure (Section 3.2.2 and Figures 3.14 and 3.16 to 3.19).

The first section applies the convergence framework defined previously in Section 2.3.6. The subsampling procedure, SEM formalism, power law model, convergence-rate proxy γ, and convergence criterion were introduced in Sections 2.3.4 and 2.3.6 and Equations (2.37) to (2.39) and are therefore not re-derived here. Instead, the focus is on whether the non-permeabilised dataset satisfies the predefined convergence criterion, and on how the fitted parameters compare between cell lines (Figures 3.1 to 3.3 and Table 3.1).

A subset of metrics is then selected for heterogeneity comparison. From the candidate descriptor set, three statistically well-behaved and biologically interpretable metrics are identified as primary convergence probes for this chapter: nuclear area, nuclear aspect ratio, and DAPI density (Figure 3.5 and Table 3.1). The fitted exponents for these quantities were consistent with the expected $N^{-1/2}$ scaling, supporting the use of SEM-convergence modelling as a valid framework for this dataset. Applying the same convergence criterion across cell lines also yields a key biological contrast: the malignant model required more nuclei to achieve the same relative precision, with $N_{\mathrm{conv}} = 365$ vs. 225 for nuclear area, 170 vs. 70 for aspect ratio, and 190 vs. 145 for DAPI density in LNCaP and PNT2-C2 respectively, consistent with increased heterogeneity in nuclear descriptors within the disease state model (Table 3.1). More explicitly, for nuclear area, the convergence threshold was reached at a higher N in LNCaP than in PNT2-C2, and the derived convergence rate was slower ($\gamma = 0.0114 \pm 0.0005$ vs. 0.0215 ± 0.0003, an approximately $2\times$ reduction), providing a quantitative, metric-specific statement of increased heterogeneity that is distinct from (and complementary to) conventional CV reporting(Table 3.1 and Figure 3.2).

After convergence validation, nuclear morphology is compared between LNCaP and PNT2-C2 using both bulk summary statistics and explicit distributional modelling (Figure 3.8 and Section 3.2.1). Summary statistics including mean, mode, and CV are reported to separate a shift in typical nuclear morphology from an enlargement of within-population variability, both of which are encompassed by nuclear pleomorphism in prostate cancer [132, 135, 359]. Mode here refers to the KDE peak position throughout, rather than to the tallest histogram bin (Equation (2.34), Figure 3.8, and Section 3.2.1). In the non-permeabilised condition, LNCaP exhibited larger mean nuclear area and higher mean aspect ratio than PNT2-C2, indicating nucleomegaly and reduced nuclear circularity in the malignant model; the magnitude of the mean shift was substantial (area increased by $\sim 45\%$, from $(136 \pm 1)\,\mu\mathrm{m}^2$ to $(195 \pm 3)\,\mu\mathrm{m}^2$, and aspect ratio by $\sim 20\%$, from 1.443 ± 0.006 to 1.72 ± 0.02), and the dispersion was also increased (CV increased by $\sim 30\%$ for area and $\sim 60\%$ for aspect ratio relative to PNT2-C2;

Figure 3.8). These findings align with the established role of nuclear morphometry in prostate cancer progression while providing a higher-resolution characterisation than mean-only comparisons [131, 132, 360].

To ensure that these contrasts were not artefacts of mean sensitivity to rare extremes, the full nucleus-level distributions for nuclear area and aspect ratio were explicitly analysed using the KDE and ICL-based framework introduced previously in Section 2.3.5 and shown here in Section 3.2.1 and Figures 3.9, 3.10, 3.12 and 3.13. This distributional analysis yields three distinct forms of information that are typically unavailable from mean-based summaries alone. First, the relative shift in modal (typical) nuclear morphology, which was directionally concordant with the mean-level comparison, with LNCaP showing an approximately 25% higher modal nuclear area (from $(120 \pm 9)\,\mu m^2$ to $(152 \pm 4)\,\mu m^2$), and an approximately 10% higher modal aspect ratio (from 1.443 ± 0.006 to 1.72 ± 0.02). Second, tail probabilities and percentile-level differences that quantify the frequency of extreme nuclei (for example, $\sim$ 10% of LNCaP vs. $\sim$ 1% of PNT2-C2 nuclei exceeded $300\,\mu m^2$, and $\sim$ 25% vs. $\sim$ 5% exceeded aspect ratio 2). Third, negative evidence for clearly separable subpopulations when conservative mixture selection favours a single component. A further internal robustness check rarely carried out in the wider literature is included by downsampling the larger PNT2-C2 dataset to match the LNCaP nucleus count and repeating the distributional comparisons, thereby isolating the qualitative interpretation from simple sample-size imbalance (Figure 3.11).

The second section investigates DAPI-derived intensity descriptors to separate absolute DNA-associated signal differences from nuclear-size effects and from intranuclear organisation (Figure 3.14 and Section 3.2.2). Total DAPI intensity per nucleus was increased in LNCaP relative to PNT2-C2 by approximately 40%, consistent with the known aneuploidy and genomic complexity of LNCaP and with the association between aneuploidy and clinically aggressive prostate cancer (Figure 3.14; [2, 361–363]). However, because total intensity is expected to vary with nuclear size, DAPI density is also evaluated as an area-normalised measure. Here, another key distributional result emerged: while mean DAPI density values overlapped, the modal density in LNCaP exceeded that of PNT2-C2 by $\sim$ 15%, and the distributional shapes were distinct between cell lines, suggesting that the "typical" density can shift, even when the mean appears similar, by virtue of differing skewness and tail structure, though it is important to interpret this carefully, as extreme tails especially were sensitive to segmentation failures/artefacts (Figures 3.14, 3.16 and 3.17). Concretely, the chapter demonstrates that distributional modelling can reveal biologically interpretable differences that are not visible from mean comparison alone, with the caveat that DAPI intensity is influenced by staining chemistry, accessibility, and optical factors, not solely DNA mass [8, 9, 358].

Most importantly for the new biological knowledge gained through this chapter, intranuclear heterogeneity is quantified directly and in a distributionally explicit manner. Each nucleus is assigned a DAPI-CV, which measures intranuclear intensity contrast and has previously been used as a practical descriptor of chromatin heterogeneity under controlled conditions [9, 358]. LNCaP exhibited a markedly higher DAPI-CV than PNT2-C2 ($\sim$ 40% relative increase), and threshold-based bright and dark-pixel fractions corroborated that LNCaP nuclei occupied larger fractions of both high and low-intensity extremes (median $R_{\text{bright}} = 70.7\%$ vs. 55.6%, and median $R_{\text{dark}} = 4.2\%$ vs. 0.5%; Figures 3.14, 3.15, 3.18 and 3.19). Distributional modelling then provided an additional, distinct result: the LNCaP DAPI-CV distribution was best represented by a two-component mixture structure, whereas the PNT2-C2 distribution was adequately represented as unimodal (Figures 3.18 and 3.19). This constitutes a direct

measure of relative intranuclear heterogeneity that cannot be recovered from single mean/CV reporting alone.

Finally, the DAPI-derived heterogeneity results are used to motivate the permeabilised-cell analyses that follow. In DAPI images, nucleoli are DNA-poor and can therefore appear as DAPI-dim voids within nuclei, enabling nucleolar inference via negative contrast under appropriate conditions [364]. In parallel, nucleolar enlargement and prominence are established correlates of prostate malignancy in diagnostic pathology [72], and MYC-driven programmes can drive nucleolar structural alterations in prostatic neoplasia [17]. As such, the observed elevation in intranuclear intensity heterogeneity in the malignant model plausibly reflects, at least in part, altered nucleolar architecture and/or altered chromatin compartmentalisation (Figures 3.14, 3.15, 3.18 and 3.19). This chapter therefore establishes a convergence-verified and distributionally complete baseline for nuclear morphology and DAPI-derived heterogeneity in the non-permeabilised condition, and defines the quantitative questions subsequently addressed under permeabilised conditions, where intranuclear components can be assessed directly (Chapter 4).

3.1 Convergence testing

3.1.1 Convergence behaviour and robustness checks

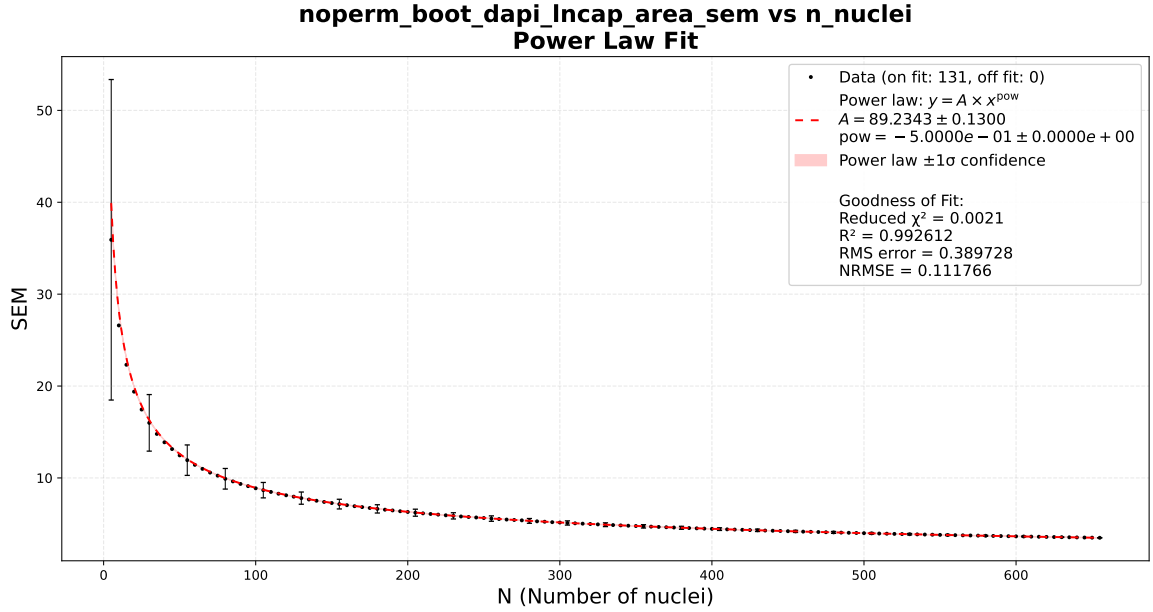
Convergence testing was performed with two key aims in mind. First, it assessed sampling sufficiency, i.e. whether the number of analysed nuclei (N) was large enough for stable estimation of the reported descriptors. Second, it provided a metric-specific proxy for relative heterogeneity, because slower decay of $\mathrm{SEM}_x(N)$ with increasing N implies greater nucleus-level dispersion for descriptor x under the approximately independent-nuclei assumption (Section 2.3.6, Equations (2.37) and (2.38), and Table 3.1; [346]).

Consistent with these aims, the SEM decreased systematically with increasing nucleus count N for all reported descriptors in both cell lines. Convergence was summarised using the percentage standard error defined in Equation (2.39), and data were deemed to be converged when this quantity crossed below the chosen threshold of 2.5%, and remained below that threshold for all subsequent sampled N, with the convergence point being referred to as N_{conv}, providing an explicit criterion for determining sample-size sufficiency (Section 2.3.6) [301]. For nuclear area (used as an illustrative example in Figures 3.1 to 3.4), the SEM decay closely followed the expected $N^{-1/2}$ scaling and was thus modelled accurately using the power law framework defined in Equation (2.37) [346, 348–350]. Departures from $pow \approx -\frac{1}{2}$ are interpreted diagnostically, as they may indicate inter-nuclear dependence/correlation, non-stationarity across sampled nuclei, and/or artefacts introduced during imaging or processing.

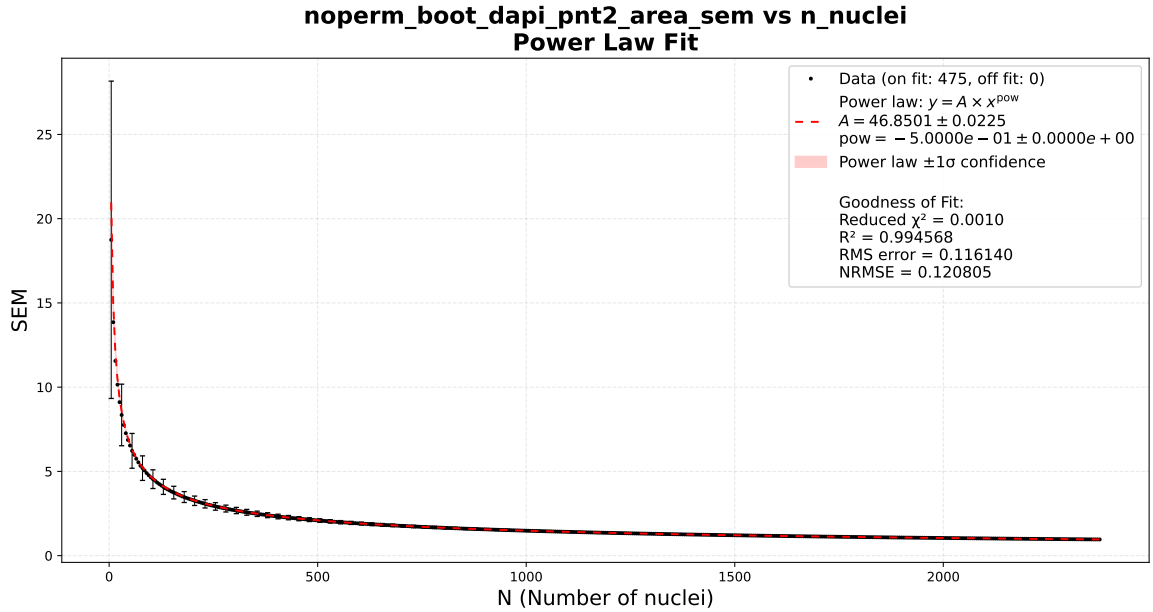
Unlike conventional microscopy studies which report a fixed cell number without quantitative justification, the present work explicitly models the decay of the sampling uncertainty as a function of nucleus count, and defines a reproducible convergence criterion (Figures 3.1 and 3.2 and Table 3.1). This establishes the minimum N required for the stable estimation for each metric independently, and therefore converts the choice of sample size into a quantitatively defensible, metric-specific determination. In addition, by fitting the SEM decay with a theoretically motivated power law and extracting the amplitude parameter (A) and convergence rate ($\gamma = 1/A$), heterogeneity is leveraged as a directly measurable quantity rather than inferred qualitatively [365].

Within the framework laid out above, the observed differences in N required to reach the chosen convergence threshold and differences in the convergence rates reported in Table 3.1 are interpreted as differences in relative dispersion.

First, a power law with *pow* fixed to $-\frac{1}{2}$ was fit to the data to show the ideal relationship:



(a) The average SEM of nuclear area in LNCaP cells calculated from 10000 repeated random subsamples plotted against N. A power law fit is applied, with fit parameters of $A = 89.2 \pm 0.1$, and *pow* fixed at -0.5. Goodness of fit metrics were NRMSE $= 0.112$, $\chi^2_\nu = 0.002\,07$, and $R^2 = 0.993$.



(b) The average SEM of nuclear area in PNT2-C2 cells calculated from 10000 repeated random subsamples plotted against N. A power law fit is applied, with fit parameters of $A = 46.85 \pm 0.02$, and *pow* fixed at -0.5. Goodness of fit metrics were NRMSE $= 0.121$, $\chi^2_\nu = 9.65 \times 10^{-4}$, and $R^2 = 0.995$.

Figure 3.1: Graphs showing the standard error of the mean (SEM) convergence of LNCaP (a) and PNT2-C2 (b) nuclear area as a function of N, the number of nuclei included in the calculation Equation (2.37). Error bars are drawn every five points, and indicate the subsampling standard error of each datapoint Equation (2.33). NRMSE refers to normalised root mean square error (Equation (2.46)), χ^2_ν refers to reduced chi squared (Equation (2.48)), and R^2 refers to the coefficient of determination (Equation (2.44)). Fits use weighted least squares regression with inverse variance weighting applied, and uncertainties were calculated using covariance aware error propagation (Equation (2.40)).

Goodness of fit in Figure 3.1 is summarised using the normalised root mean square error (NRMSE)

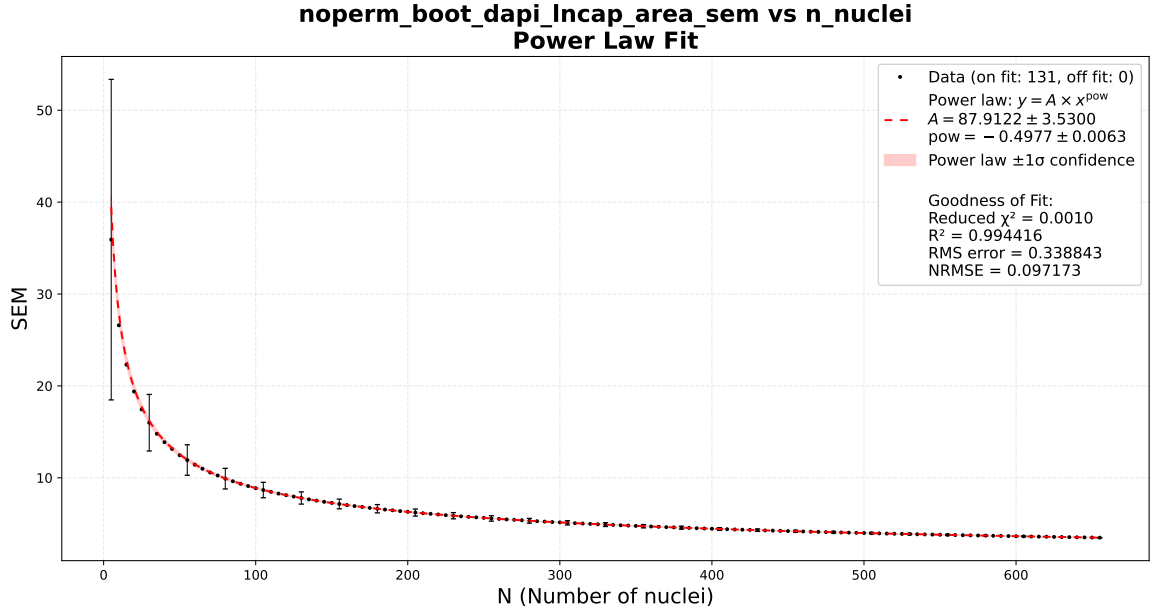
as described in Equation (2.46) so that deviations are interpretable on a relative scale across metrics with different units. Reduced chi squared (χ^2_ν) and the coefficient of determination (R^2) are calculated as in Equations (2.44) and (2.48) and treated as supplementary descriptive summaries only; SEM points derived from repeated random subsampling are not necessarily strictly independent, and the regressions are weighted and non-linear, which can limit the formal interpretation of these quantities [356, 357].

Considering this, the power law model with the scaling exponent *pow* fixed to the theoretical value of $-\frac{1}{2}$ provides an excellent description of the observed SEM decay for nuclear area in both cell lines (Figure 3.1). In both cases, the fit tracks the data well across the full N range, with very low NRMSE and χ^2_ν values (~ 0.11 to 0.12 and $\sim 0.002\,07$ to 9.65×10^{-4} respectively), as well as R^2 very close to 1, collectively indicating that the data closely follows the expected $\frac{1}{\sqrt{N}}$ averaging behaviour for this metric under this constraint [346, 347]. The convergence rate, γ, differs substantially between cell lines, being approximately two times smaller in LNCaP than PNT2-C2, consistent with slower convergence and greater variability in this metric in LNCaP. Having established that the fixed-exponent model is adequate, *pow* was next released as a free parameter to test whether the observed scaling exponent is intrinsically consistent with $-\frac{1}{2}$, rather than being enforced by construction, and to provide a sensitive diagnostic for any departures from ideal averaging behaviour.

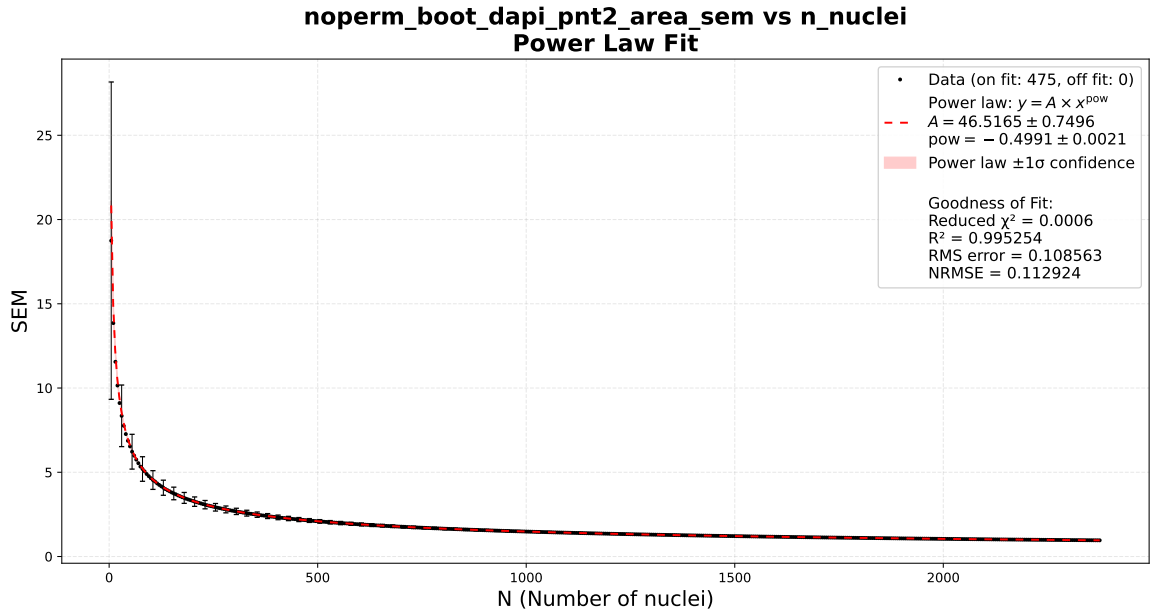
Allowing *pow* to vary freely (Figure 3.2) produces fits that are essentially unchanged in shape and interpretation relative to the fixed model (Figure 3.1). In both cell lines, the fitted exponent remained tightly constrained to within 0.5% of the theoretical value ($pow = -0.498 \pm 0.006$ and -0.499 ± 0.002 for LNCaP and PNT2-C2 respectively), indicating that the observed SEM decay is in fact consistent with the expected $\frac{1}{\sqrt{N}}$ scaling. Correspondingly, the convergence rates in free-*pow* fits were stable to within $\sim 1\%$ relative to the fixed-exponent fits, supporting the interpretation that LNCaP exhibits slower convergence and greater dispersion for nuclear area. Having established clear agreement in both exponent and amplitude under the unconstrained model, the analysis was next repeated in log space to further verify power law scaling, where the fitted slope provides an independent estimate of the exponent and offers a linear-regression analogue for assessing scaling consistency.

The log-log graphs in Figure 3.3 provide an explicit test of power law scaling using the relation defined in Equation (2.41), where the fitted slope corresponds to *pow* and the intercept to $\log_{10} A$. In both cell lines, the data lie close to a straight line across the sampled range; all points fall well within uncertainty of the fitted relationship. Weighted linear regressions also show excellent agreement: $R^2 = 0.9995$, RMSE $= 0.004\,36$, and $\chi^2_\nu = 9.66 \times 10^{-4}$ for LNCaP, and $R^2 = 0.9998$, RMSE $= 0.002\,82$, and $\chi^2_\nu = 5.32 \times 10^{-4}$ for PNT2-C2. Note that RMSE, rather than NRMSE is reported for log-log graphs. This is because in log space, residuals already represent relative deviations; as such, the corresponding RMSE is inherently dimensionless, and so does not require normalisation to convert it into a relative measure in the same way as RMSE in linear space.

The fitted slope on the log-log graph, m, is equivalent to *pow*, and the y-axis intercept, c, is equivalent to $\log_{10}(A)$. It can be seen that the fit parameters (and derived convergence rates) from the log-log graphs are consistent with those from the power law fits in linear space. From Figure 3.3 (log space), LNCaP gave $pow = -0.498 \pm 0.006$ and $\gamma = 0.0114 \pm 0.0005$, with PNT2-C2 giving $pow = -0.499 \pm 0.002$ and $\gamma = 0.0215 \pm 0.0003$. In Figure 3.2 (linear space), the corresponding power law fits gave $pow = -0.498 \pm 0.006$ and $\gamma = 0.0114 \pm 0.0005$ for LNCaP, and $pow = -0.499 \pm 0.002$ and $\gamma =$

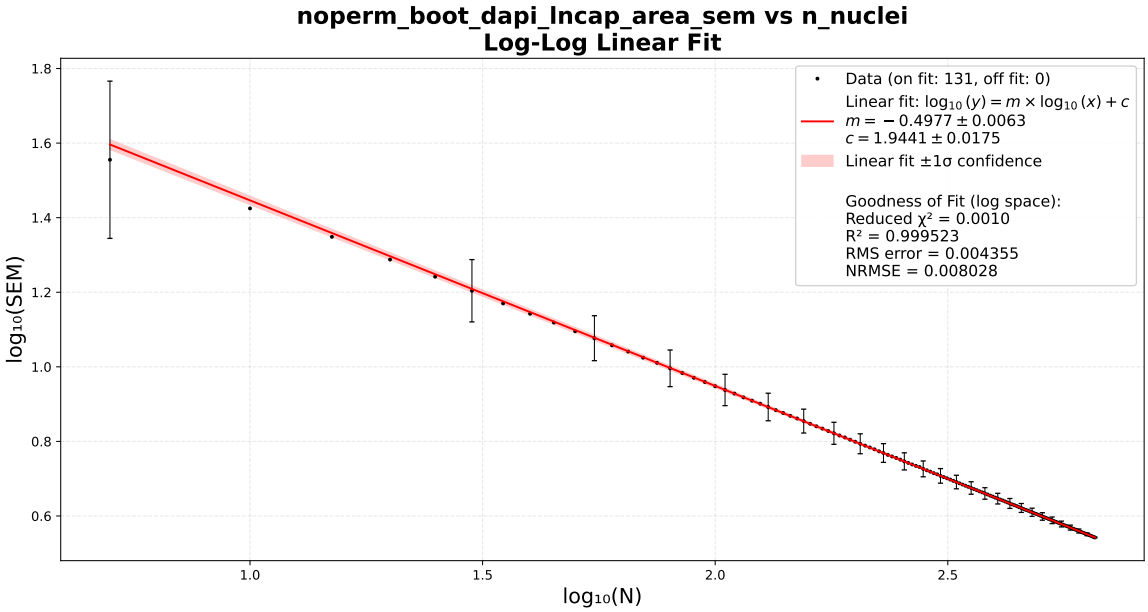


(a) The average SEM of nuclear area in LNCaP cells calculated from 10000 repeated random subsamples plotted against N. A power law fit is applied, with fit parameters of $A = 88 \pm 4$, and $pow = -0.498 \pm 0.006$. Goodness of fit metrics were NRMSE = 0.0972, $\chi^2_\nu = 0.001\,01$, and $R^2 = 0.994$.

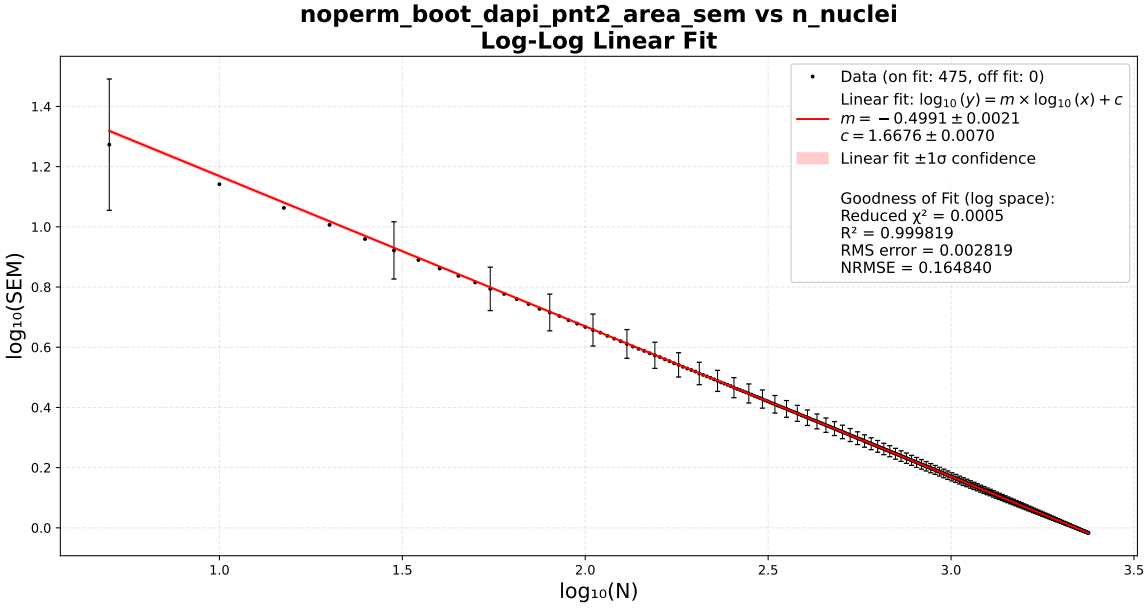


(b) The average SEM of nuclear area in PNT2-C2 cells calculated from 10000 repeated random subsamples plotted against N. A power law fit is applied, with fit parameters of $A = 46.5 \pm 0.7$, and $pow = -0.499 \pm 0.002$. Goodness of fit metrics were NRMSE = 0.113, $\chi^2_\nu = 5.50 \times 10^{-4}$, and $R^2 = 0.995$.

Figure 3.2: Graphs showing the standard error of the mean (SEM) convergence of LNCaP (a) and PNT2-C2 (b) nuclear area as a function of N, the number of nuclei included in the calculation Equation (2.37). Error bars are drawn every five points, and indicate the subsampling standard error of each datapoint Equation (2.33). NRMSE refers to normalised root mean square error (Equation (2.46)), χ^2_ν refers to reduced chi squared (Equation (2.48)), and R^2 refers to the coefficient of determination (Equation (2.44)). Fits use weighted least squares regression with inverse variance weighting applied, and uncertainties were calculated using covariance aware error propagation (Equation (2.40)).



(a) The log of average SEM of nuclear area in LNCaP cells calculated from 10000 repeated random subsamples plotted against log N. A linear fit is applied with fit parameters $m = -0.498 \pm 0.006$ and $c = 1.94 \pm 0.02$. Goodness of fit metrics were RMSE $= 0.004\,36$, $\chi^2_\nu = 9.66 \times 10^{-4}$, and $R^2 = 0.9995$.



(b) The log of average SEM of nuclear area in PNT2-C2 cells calculated from 10000 repeated random subsamples plotted against log N. A linear fit is applied with fit parameters $m = -0.499 \pm 0.002$ and $c = 1.668 \pm 0.007$. Goodness of fit metrics were RMSE $= 0.002\,82$, $\chi^2_\nu = 5.32 \times 10^{-4}$, and $R^2 = 0.9998$.

Figure 3.3: Graphs showing the log of standard error of the mean (SEM) convergence of LNCaP (a) and PNT2-C2 (b) nuclear area as a function of log N, the number of nuclei included in the calculation Equation (2.41). Error bars are drawn every five points, and indicate the subsampling standard error of each datapoint Equation (2.33). RMSE refers to root mean square error (Equation (2.45)), χ^2_ν refers to reduced chi squared (Equation (2.48)), and R^2 refers to the coefficient of determination (Equation (2.44)). Fits use weighted least squares regression with inverse variance weighting applied, and uncertainties were calculated using covariance aware error propagation (Equation (2.40)).

0.0215 ± 0.0003 for PNT2-C2. All values are exactly congruent, and so together with the strong linearity of the log-log plots, this further supports the use of the power law model as the primary description of SEM decay.

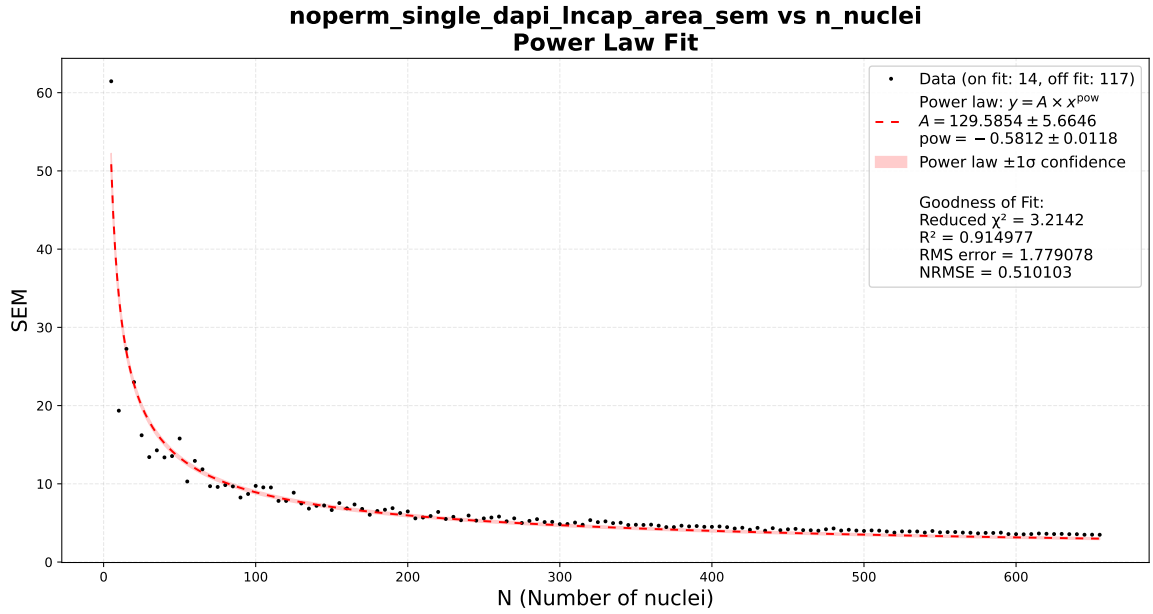
To motivate the use of repeated random subsampling for convergence assessment, convergence curves were also computed using an unreplicated procedure, in which a single random sample was taken at each N, and the SEM was evaluated once for that draw. This produced one SEM estimate per N without averaging across resamples, and therefore provided a direct illustration of the level of stochastic variation introduced by resampling [334, 335]. The resulting single draw curves are shown in Figure 3.4 to allow direct comparison with the corresponding repeated subsampling results.

Repeated random subsampling was used as the primary approach because it enables estimation of the expected SEM associated with each sample size N by aggregating over many resamples, rather than basing analysis on an arbitrary single draw, as detailed in Sections 2.3.4 and 2.3.6 [334, 335]. This reduced the random variability introduced by the resampling procedure, thereby reducing the influence of stochastic variations on any threshold based convergence rules or fitted power law parameters; the effect is especially visible when comparing Figures 3.2 and 3.4. In the single draw case, datapoints exhibit notable point-to-point fluctuations, and *pow* deviates from the expected value by $\sim 15\%$. Convergence rates also vary substantially between the repeated subsample and single draw figures (0.0114 ± 0.0005 and 0.0215 ± 0.0003 vs. 0.0077 ± 0.0003 and 0.035 ± 0.001 for LNCaP and PNT2-C2 in the repeated subsampling and single draw cases respectively; Figures 3.2 and 3.4), and goodness of fit measures clearly show that the fits to single draw data are less exact than for data obtained via repeated random subsampling. As such, the replicated subsampling curves are used for all primary convergence analyses, with single draw results only serving as an illustrative comparison.

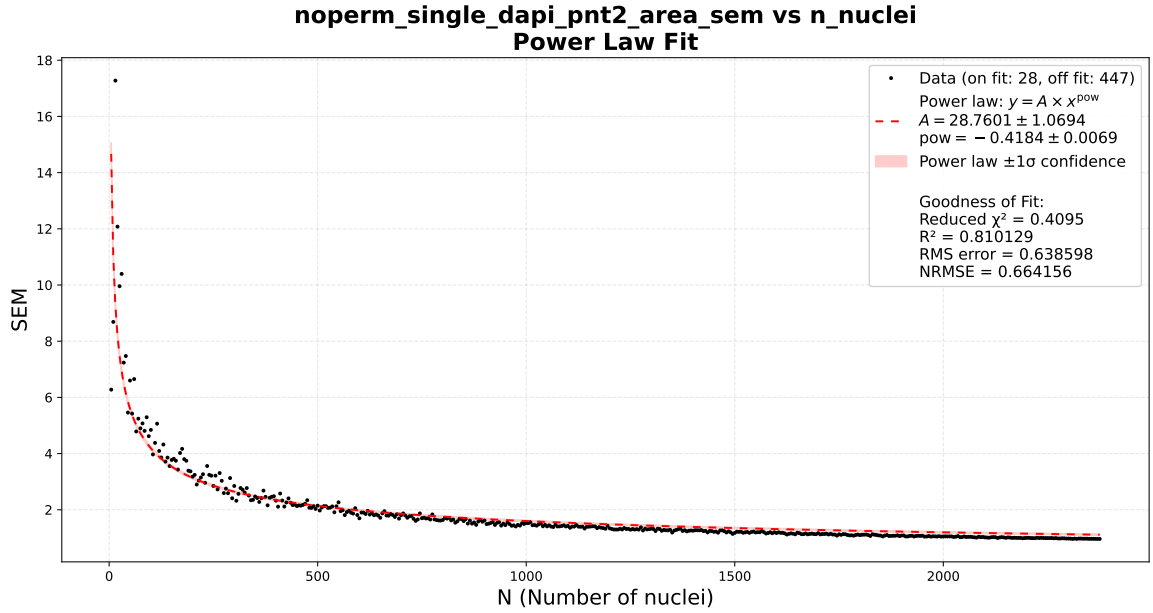
3.1.2 Metric selection

Across the full set of quantities calculated from the non-permeabilised dataset, convergence testing was performed to identify a subset of metrics that are both statistically well-behaved and biologically interpretable. Metric selection was guided by prioritising good statistical behaviour under the SEM power law model and biological interpretability. Metrics were retained based on proximity of the fitted exponent to the theoretical SEM scaling of $pow = -0.5$ (with an inclusion criterion of $\pm 1\%$), the power law providing a low-residual description (primarily assessed using NRMSE, with R^2 and χ^2_ν treated as supplementary), and agreement of linear space power law fits with the corresponding log space linear regressions [346, 347, 350]. The final selection also required that each metric be readily interpretable in the context of nuclear morphology and DAPI staining [8, 134, 193, 203, 366].

To support this selection, Figure 3.5 summarises the fitted *pow* values obtained from the linear space power law fits and the corresponding log space linear regressions for each candidate metric and cell line. Both replicated and single draw data are shown, as deviations from the ideal *pow* are much more pronounced in the single draw case. On the basis of the above-mentioned criteria, three metrics were selected for detailed convergence assessment and subsequent comparison of heterogeneity between cell lines: nuclear area, nuclear aspect ratio, and DAPI density (defined as mean DAPI intensity per unit area). Conversely, aspect ratio error, the ratio of dark pixels to total pixels, and DAPI intensity show the largest deviations from the expected *pow*. The ratio of bright pixels to total pixels was excluded from the chosen metrics despite good agreement with the theoretical *pow* and promising goodness of

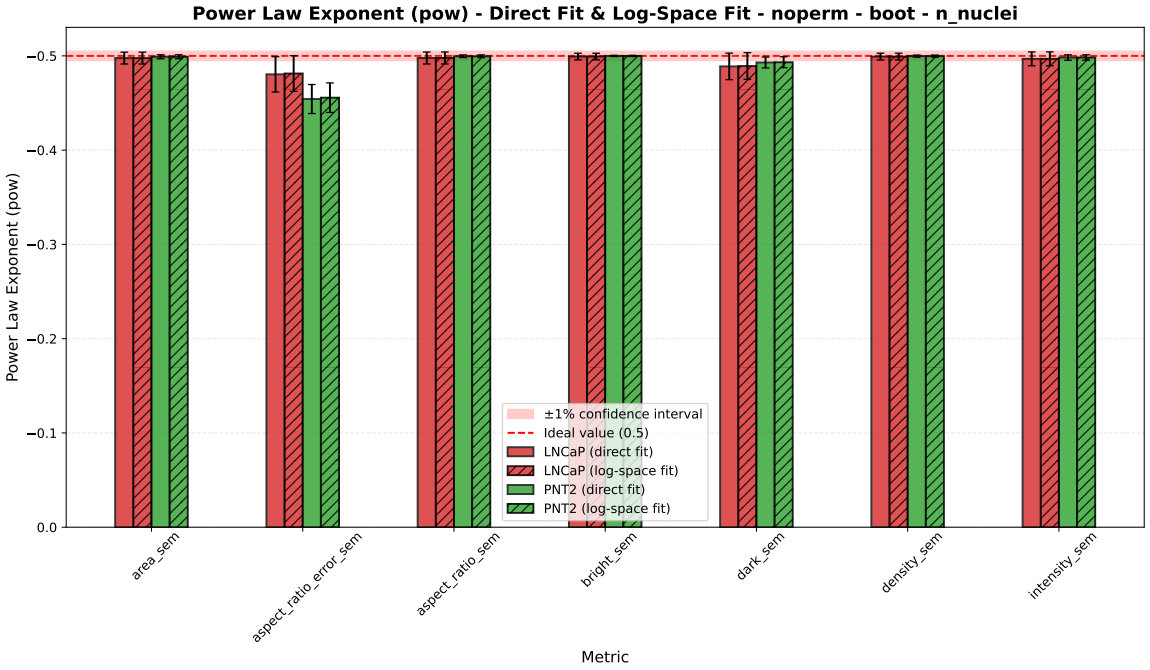


(a) SEM of nuclear area in LNCaP cells plotted against N. A power law fit is applied, with fit parameters of $A = 130 \pm 6$, and $pow = -0.58 \pm 0.01$. Goodness of fit metrics were NRMSE $= 0.510$, $\chi^2_\nu = 3.21$, and $R^2 = 0.915$.

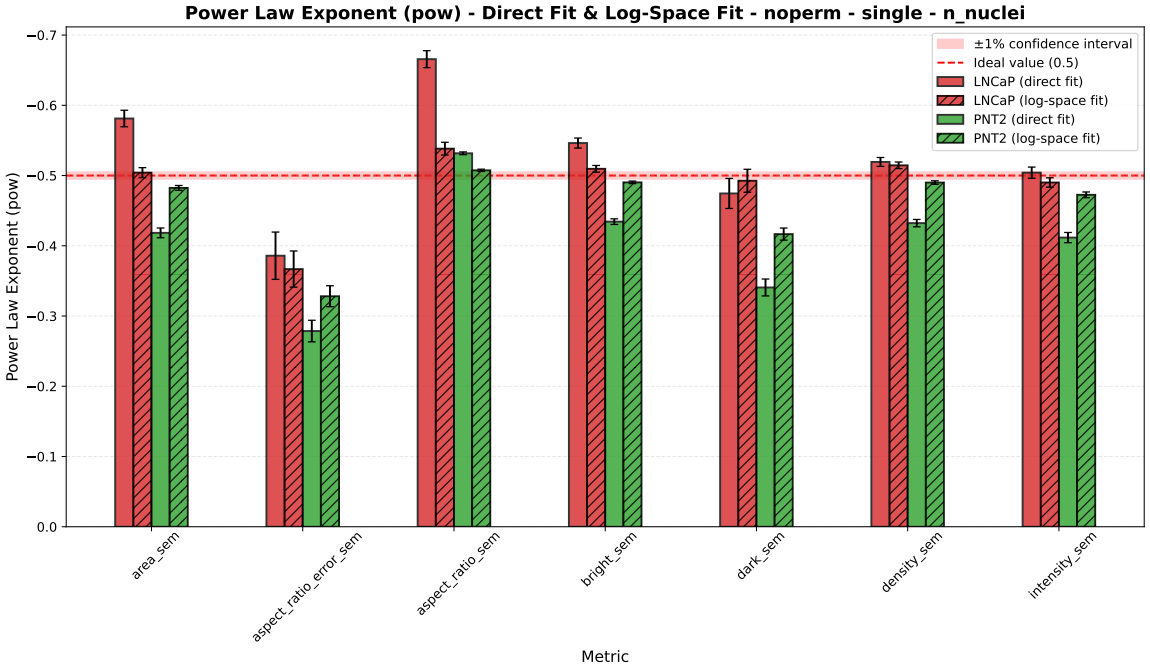


(b) SEM of nuclear area in PNT2-C2 cells plotted against N. A power law fit is applied, with fit parameters of $A = 29 \pm 1$, and $pow = -0.418 \pm 0.007$. Goodness of fit metrics were NRMSE $= 0.664$, $\chi^2_\nu = 0.41$, and $R^2 = 0.810$.

Figure 3.4: Graphs showing the standard error of the mean (SEM) convergence of LNCaP (a) and PNT2-C2 (b) nuclear area as a function of N, the number of nuclei included in the calculation Equation (2.37). NRMSE refers to normalised root mean square error (Equation (2.46)), χ^2_ν refers to reduced chi squared (Equation (2.48)), and R^2 refers to the coefficient of determination (Equation (2.44)). Fits used equally weighted least squares regression, and uncertainties were calculated using covariance aware error propagation (Equation (2.40)).



(a) *pow* values from the repeated randomly subsampled power law fit in linear space (solid fill), and the linear fit in log space (hatched fill) for each metric and cell line. LNCaP is shown in red, and PNT2-C2 in green. The red dotted line indicates the ideal value of -0.5, with the shaded region representing the allowed 1% variation.



(b) *pow* values from the single draw power law fit in linear space (solid fill), and the linear fit in log space (hatched fill) for each metric and cell line. LNCaP is shown in red, and PNT2-C2 in green. The red dotted line indicates the ideal value of -0.5, with the shaded region representing the allowed 1% variation.

Figure 3.5: Graphs showing *pow* values from repeated random subsampling (a) and single draw (b) fits for SEM of each calculated metric: nuclear area, aspect ratio error, aspect ratio, the ratio of bright pixels to total pixels, the ratio of dark pixels to total pixels, DAPI density, and DAPI intensity.

fit measures, as it is not necessarily straightforward/intuitive in its interpretation. Nuclear area, aspect ratio, and DAPI density are thus treated as representative convergence probes for the non-permeabilised dataset. Demonstrating stable SEM behaviour for these descriptors provides evidence that the available sample size is sufficient for robust estimation of nuclear morphology and DAPI-intensity-derived measures, and thus supports the use of this dataset for subsequent analyses.

3.1.3 Summary and implications

For clarity, a quantitative summary of the key fit and convergence outputs for all candidate metrics is provided in Table 3.1. This table reports convergence rates, power law scaling exponents, goodness of fit metrics, and convergence points defined as the point at which the percentage standard error (%SEM) fell below 2.5%, and remained below that threshold for all subsequent points (Equation (2.39)).

Additional convergence data are included in Section B.1.

Key quantitative outcomes

Table 3.1 summarises the power law fit parameters, goodness of fit statistics, and convergence points for the three selected metrics (nuclear area, nuclear aspect ratio, and DAPI density). Across these metrics, the fitted exponents were consistently very close to the theoretical expectation of $pow = -0.5$, supporting the interpretation that SEM decays approximately as $\mathrm{SEM}(N) \propto N^{-1/2}$ over the sampled range [346, 347, 350].

Using the specified threshold of %SEM $< 2.5\%$, convergence was achieved at $N = 365$ and $N = 225$ nuclei for nuclear area in LNCaP and PNT2-C2 respectively, at $N = 170$ and $N = 70$ for aspect ratio, and at $N = 190$ and $N = 145$ for DAPI density (Table 3.1). To visualise the relative convergence points, the SEM decay for nuclear aspect ratio in both cell lines is shown below in Figure 3.6, with PNT2-C2 data truncated to the same $N_{\max}$ as LNCaP to enable direct comparison between the two cell lines.

All convergence points lay well below the total available nucleus counts ($N_{\max} = 657$ for LNCaP and $N_{\max} = 2378$ for PNT2-C2), demonstrating that the dataset was large enough for stable estimation under the chosen precision criterion (Table 3.1). Quantitatively, the %SEM $< 2.5\%$ convergence threshold corresponded to 25 % to 55 % of the available LNCaP nuclei vs. only 3 % to 10 % of the available PNT2-C2 nuclei (Table 3.1). This contrast was also visible in Figure 3.6: even when PNT2-C2 was cropped to the same $N_{\max}$ as LNCaP, it still showed faster convergence, smaller absolute SEM values, and a smaller N_{conv}. This is consistent with an intrinsic reduction in dispersion rather than an artefact of unequal $N_{\max}$ between datasets.

Part of the reason for the significant disparity in $N_{\max}$ between cell lines is the difficulty of maintaining LNCaP as a relatively sparse, single layer suitable for the imaging carried out in this study. LNCaP is known for being weakly adherent, such that LNCaP cells are very easily dislodged during the culture process, and often grow in clusters, rather than forming a single monolayer [163, 367].

Because the fitted exponents remain consistent with the theoretical $N^{-1/2}$ scaling, the observed differences in γ and N_{conv} arise primarily from amplitude differences rather than anomalous averaging behaviour (Table 3.1 and Figures 3.2 and 3.6). As such, the heterogeneity contrast between LNCaP and PNT2-C2 is quantitatively encoded in the SEM amplitude itself. This transforms heterogeneity

Cell line	Metric	γ	*pow*	NRMSE	χ^2_ν	N_{conv}
LNCaP ($N_{\max} = 657$)	Area	0.0114 ± 0.0005	-0.498 ± 0.006	0.0972	0.001 01	365
	Aspect ratio	1.94 ± 0.08	-0.498 ± 0.006	0.104	9.61×10^{-4}	170
	DAPI density	$(7.1 \pm 0.2) \times 10^{-7}$	-0.499 ± 0.004	0.0512	5.02×10^{-4}	190
PNT2-C2 ($N_{\max} = 2378$)	Area	0.0215 ± 0.0003	-0.499 ± 0.002	0.113	5.50×10^{-4}	225
	Aspect ratio	3.72 ± 0.04	-0.500 ± 0.001	0.0838	3.94×10^{-4}	70
	DAPI density	$(8.36 \pm 0.07) \times 10^{-7}$	-0.500 ± 0.001	0.0604	2.82×10^{-4}	145

Table 3.1: Summary of convergence analysis for non-permeabilised data. Power law fit parameters and goodness of fit statistics refer to weighted fits performed to the mean SEM calculated using repeated random subsampling as a function of N, where γ is the convergence-rate proxy defined in Equation (2.38), and *pow* is the fitted exponent. The convergence point, N_{conv}, is defined using the percentage standard error criterion in Equation (2.39), and is reported as a discrete value without uncertainty. NRMSE refers to normalised root mean square error, normalised to terminal SEM, and χ^2_ν refers to reduced chi squared. $N_{\max}$ is the total number of nuclei available in each cell line.

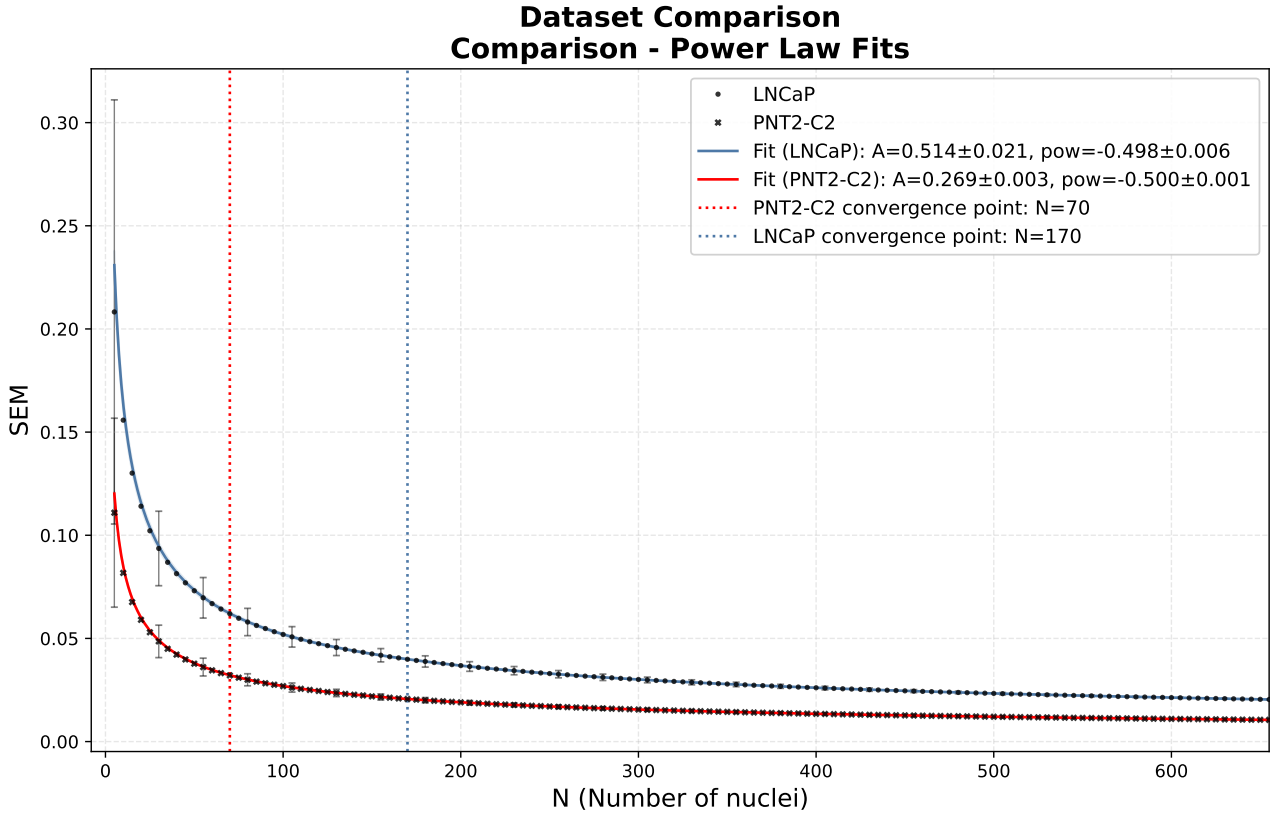


Figure 3.6: The standard error of the mean (SEM) convergence of LNCaP and PNT2-C2 nuclear aspect ratio calculated from 10000 repeated random subsamples as a function of N, the number of nuclei included in the calculation. The data range of PNT2-C2 has been cropped to $N \leq N_{\max,\text{LNCaP}} = 657$ to enable direct comparison on a common N range, and N_{conv} for both cell lines are indicated with dashed vertical lines. Error bars are drawn every five points, and indicate the subsampling standard error of each datapoint Equation (2.33). A power law fit is applied in both cases (Equation (2.37)), with fit parameters of $A = 0.51 \pm 0.02$ and $pow = -0.498 \pm 0.006$ for LNCaP, and $A = 0.269 \pm 0.003$ and $pow = -0.500 \pm 0.001$ for PNT2-C2. Fits use weighted least squares regression with inverse variance weighting applied, and uncertainties were calculated using covariance aware error propagation (Equation (2.40)).

from a descriptive label into a parameterised quantity, enabling formal comparison between metrics and cell lines. The convergence analysis therefore functions not merely as a validation step, but as a primary analytical output in its own right.

Relative cell line variability and heterogeneity

Under the empirically supported $pow \approx -\frac{1}{2}$ scaling (Figures 3.1 to 3.3), $\mathrm{SEM}(N) \approx \sigma/\sqrt{N}$, and thus the amplitude A is proportional to the underlying nucleus-level dispersion, σ. As such, the number of nuclei required to satisfy the convergence threshold N_{conv} is directly aligned with the argument that the minimum number of nuclei measured should be chosen to reduce averaging error to a level commensurate with the biological question under study [301]. Because $\gamma = \frac{1}{A}$ (Equation (2.38)), it may be interpreted as a "convergence rate", i.e. if the reduction in SEM is greater for a given step in N, so too is γ ([365]). Consequently, the consistently smaller convergence rates and larger N_{conv} observed in LNCaP across all three selected metrics indicate greater cell-to-cell variability, and thus greater heterogeneity in the metastatic cancer model relative to the non-malignant epithelial model (Table 3.1).

Quantitatively, γ in PNT2-C2 was almost double that in LNCaP for area and aspect ratio, and increased by almost 20% for DAPI density (Table 3.1). Similarly, N_{conv} was larger by $\sim 50\%$ for area, $\sim 85\%$ for aspect ratio, and $\sim 25\%$ for DAPI density in LNCaP compared with PNT2-C2 (Table 3.1). The same ordering is visible in the convergence-point comparison for aspect ratio in Figure 3.6, where N_{conv} was approximately 2.5× greater in LNCaP than in PNT2-C2. These contrasts are not driven by anomalous scaling: *pow* remained close to -0.5 in both cell lines, so the observed differences arise predominantly from the fitted amplitude, or equivalently γ, rather than from departures from the expected SEM scaling [346, 347].

The convergence framework presented here yields two explicit pieces of new knowledge. First, it demonstrates that the present dataset is statistically sufficient for robust estimation of morphology and DAPI-derived metrics under a predefined precision criterion (Table 3.1). Second, it establishes that the malignant model exhibits systematically slower SEM decay, and therefore greater nucleus-level dispersion, across multiple independent descriptors (Table 3.1 and Figure 3.6). Importantly, this conclusion does not depend on unequal N_{max}, nor on post hoc interpretation of broad histograms, but instead follows directly from fitted convergence parameters. These parameters provide a reproducible basis for within-study comparison and, with appropriate caution, for qualitative comparison across studies. More explicitly, the absolute fitted convergence rate γ is unlikely to be directly comparable between separate studies, because it may depend on study-specific factors such as imaging conditions, segmentation choices, and sampling design. However, the relative ordering of γ across cell lines or preparation states within a given study is more likely to be robust, and that qualitative relationship may therefore be meaningfully compared with the corresponding within-study pattern reported elsewhere. Such an explicit linkage between convergence rate and biological heterogeneity does not appear to have been implemented previously for nucleus-level morphometric descriptors in prostate cancer cell-line systems.

Interpretation and contextualisation

The observed slower convergence, and therefore greater heterogeneity, in LNCaP, as summarised in Table 3.1 and illustrated for aspect ratio in Figure 3.6, is consistent with independent evidence that

LNCaP is genomically unstable and exhibits aneuploidy and multi-subpopulation structure [14, 15, 368]. In particular, early characterisation of LNCaP already reported an aneuploid karyotype [2], and later cytogenetic profiling documented complex, unstable karyotypes, consistent with chromosomal instability [164, 361, 369]. More recent genome-scale analyses further support non-integral bulk copy-number states and multiple co-existing clones, consistent with clonal heterogeneity and ongoing instability [14, 368]. Such subpopulation structure provides a direct mechanistic basis for broader nucleus-to-nucleus variation in DNA-associated and morphology-linked imaging descriptors, and therefore for a smaller γ and increased N_{conv} in LNCaP relative to PNT2-C2 (Table 3.1). At the transcriptomic level, single-cell RNA sequencing of LNCaP identifies multiple subpopulations with heterogeneous androgen responsiveness, including subsets with attenuated AR dependence and altered cell-cycle programmes [172, 173], and AR-expression heterogeneity has been linked to distinct treatment responses in prostate cancer [370]. The existence of such subpopulations, both genetic and non-genetic, is precisely the type of underlying variability expected to manifest as broader distributions of nuclear size, shape, and chromatin organisation.

Though not the first study to leverage convergence behaviour as a measure of heterogeneity [168, 231], the single-cell Raman spectroscopy study by West et al. (2025) sets a methodologically aligned precedent [127]. They explicitly performed convergence testing to justify representativeness, used converged SE envelopes to quantify molecular heterogeneity at the single-cell level, and reported that LNCaP partitioned into multiple unsupervised molecular clusters, whereas PNT2-C2 formed a single principal cluster [127]. Importantly, this supports not only the interpretation of greater heterogeneity in the cancer state, but also the relevance of reporting explicit convergence sample sizes. The present microscopy-based convergence results are therefore concordant with that broader expectation of greater cell-to-cell variability in the metastatic model when estimating mean morphology and DAPI-derived descriptors (Table 3.1 and Figure 3.6).

Conversely, PNT2-C2 is an SV40-immortalised line (derived from PNT2 (RRID:CVCL_2164) [4]), and is best interpreted as a non-malignant comparator rather than a true primary epithelium surrogate [3, 151, 162, 166, 168]. This caveat limits absolute biological generalisation; however, it does not negate the contrast observed under matched culture, imaging, and analysis conditions (Table 3.1 and Figure 3.6). Moving beyond the specific case of PNT2-C2 and LNCaP, increased heterogeneity is expected in cancer models, as genome instability generates genetic diversity on which selection can act [13, 371, 372]. In this context, chromosomal instability and aneuploidy are widely recognised routes to intercellular heterogeneity, and thus provide a biological rationale for the consistently lower convergence rates observed in LNCaP (Table 3.1; [373–375]).

Microscopy-based nuclear descriptors integrate several sources of biological variability that can broaden nucleus-to-nucleus distributions and therefore increase the sample size required for a given precision. The DAPI-density convergence difference summarised in Table 3.1 may therefore reflect multiple overlapping contributors rather than a single mechanism. DNA-associated measures are particularly sensitive because integrated DNA-dye intensity is widely used to stage the cell cycle in imaging, and asynchronous cultures necessarily contain residual variation in DNA content and nuclear organisation phases, with an approximate doubling of DNA content from G1 to G2 [193, 194, 203, 376]. In cancer, replication stress and checkpoint dysregulation can increase atypical cycling states, and together with polyploid or aneuploid subpopulations provide a plausible route to increased dispersion in DAPI density [377–379]. In addition, DAPI and Hoechst readouts are affected by chromatin condensation; prostate

cell-line comparisons have previously used DNA-dye intensity to assess condensation differences [206, 380]. While serum starvation was applied here to reduce cell-cycle asynchrony [127, 291, 292], full synchrony cannot be assumed, and residual cell-cycle-related dispersion remains a plausible, though unlikely dominant, contributor.

Separately, malignancy-associated nuclear architecture abnormalities can broaden morphology distributions. The area and aspect-ratio convergence differences summarised in Table 3.1 are therefore also compatible with a broader spectrum of nuclear architectural states in LNCaP. Nuclear size and shape irregularities are diagnostically used across cancers and are mechanistically linked with altered chromatin mechanics, lamina composition, and nucleocytoskeletal coupling [131, 381]. In PCa, nuclear envelope defects and micronuclear abnormalities have been linked to chromosomal instability and aggressive behaviour, with reported alterations in micronuclear structures and nuclear envelope component localisation across normal vs. prostate cancer cell lines including LNCaP [382–384]. Although such features were not quantified here, they motivate the interpretation that cancer nuclei may occupy a broader spectrum of architectural states, reducing effective convergence rates.

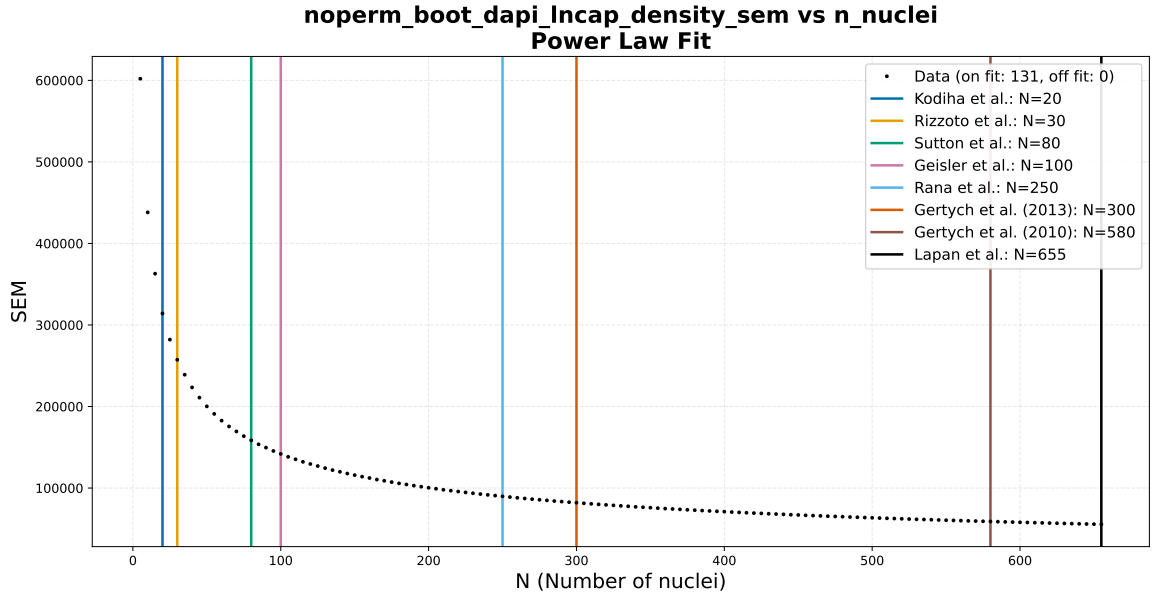
Finally, heterogeneity has practical consequences for reproducibility: cancer cell lines can harbour clonal diversity and evolve in culture, affecting the observed biological characteristics across repeats [385]. The convergence-based assessment used here operationalises this by estimating the nucleus count required for stable mean estimates at a specified precision, implying that slower-converging cell lines such as LNCaP are intrinsically more vulnerable to unstable estimates and apparent inconsistencies when quantified with limited cell counts (Table 3.1).

Comparison to cell numbers used in other studies

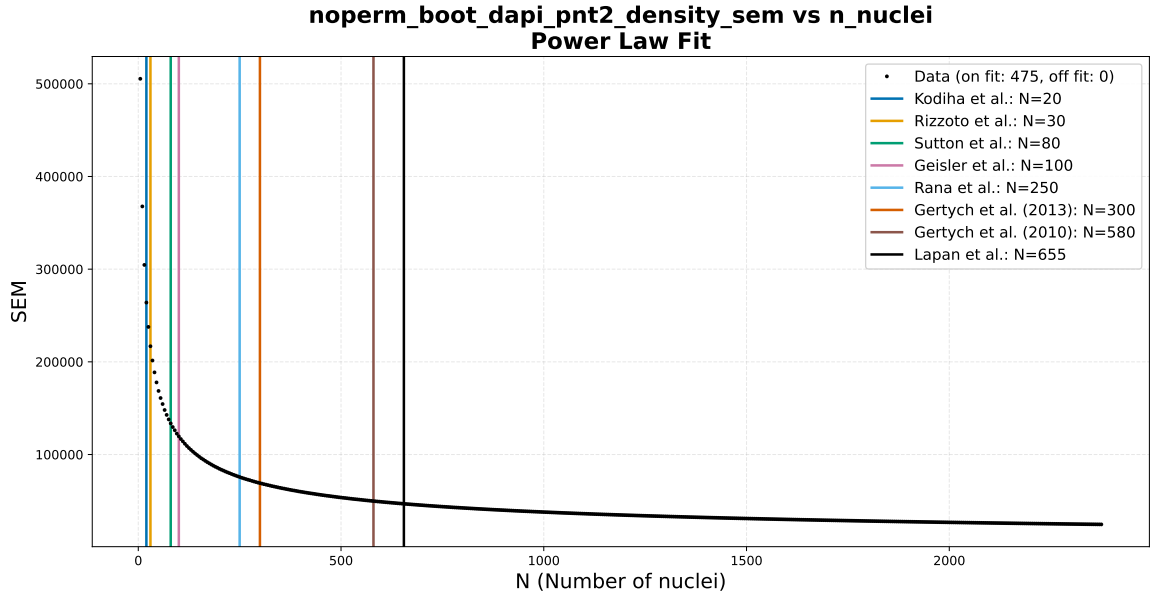
To place the present convergence results in context with other studies from the wider literature, the comparison here is restricted to fluorescence-imaging studies. This is important because Raman and FTIR-based studies do not generate directly analogous observation units, and therefore do not provide a like-for-like reference for microscopy-derived nucleus counts. Even within fluorescence microscopy, however, reported N values are not universally equivalent: some studies quantify manually curated confocal fields, whereas others use automated high-content pipelines specifically designed to segment very large numbers of cells imaged directly in wells, for example. Without some explicit method of testing, sample size also cannot be stated to be sufficient with absolute certainty. As such, literature-reported N should not be interpreted as a universal sufficiency threshold or target, but rather as context for commonly used workflows within the field.

The overlaid reference lines in Figure 3.7 make clear that, at the lower end of fluorescence practice, several studies still operate in the tens to low-hundreds regime. Kodiha et al. (2011) reported using ≥ 20 cells per condition to study nucleolar biology using computer-based image analysis, and Rizzotto et al. (2024), Sutton et al. (2021), and Geisler et al. (2026) reported ≥ 30, ≥ 80, and ≥ 100 cells per condition respectively [139, 192, 255, 257]. In both Figures 3.7(a) and 3.7(b), these reference points lie well to the left of the present $N_{\max}$, indicating that the current dataset extends substantially beyond many commonly used sample sizes in fluorescence experiments, particularly where acquisition is confocal, z-stack based, or marker-rich.

A second group of fluorescence studies lies in the few-hundred-cell regime, which is more directly comparable to the present work. As shown in Figure 3.7, the ~ 250 cell and ~ 300 nucleus sample-size



(a) The average SEM of DAPI density in LNCaP cells calculated from 10000 repeated random subsamples plotted against N. $N_{\max} = 657$ for this cell line.



(b) The average SEM of DAPI density in PNT2-C2 cells calculated from 10000 repeated random subsamples plotted against N. $N_{\max} = 2378$ for this cell line.

Figure 3.7: Literature-reported fluorescence-microscopy sample sizes overlaid on the DAPI-density convergence curves for contextual comparison. Panels show LNCaP (a) and PNT2-C2 (b). For visual clarity, the overlaid reference lines correspond only to representative fluorescence studies with explicit N values, or a relatively tight range of reported values that lay within (or close to) the plotted range in this study: Rizzotto et al. (≥ 30 cells), Sutton et al. (≥ 80 cells), Geisler et al. (≥ 100 cells), Gertych et al. (2013; ~ 300 nuclei), Gertych et al. (2010) 496 to 660, and Lapan et al. (655 adjacent LNCaP cells) [192, 247, 248, 255, 257, 386]. Additional fluorescence studies are discussed in the text, but are not included in this figure in the interest of readability, because they either report very wide ranges (for example Ogawa et al. 281 to 1097, Mandelkow et al. 400 to 1000, and Batnasan et al. 300 to 1000), or because their reported N lay substantially beyond the plotted x-range (for example Narotamo et al. ≥ 3500 nuclei per condition) [191, 224, 246, 252].

reported by Rana et al. (2021) and Gertych et al. (2013) fall near the middle of the plotted range [248, 249]. Other studies occupy a similar or moderately higher range: Ogawa et al. (2020) reported 281 to 1097 nuclei per condition in a nucleolar imaging assay, Gertych et al. (2010) reported 496 to 660 cells per condition, Mandelkow et al. (2017) reported 400 to 1000 nuclei per condition, and Batnasan et al. (2024) reported 300 to 1000 nuclei per condition in prostate cancer cell lines including LNCaP [191, 224, 247, 252]. Although the ranges reported in these studies were not all suitable for display in Figure 3.7, they are still informative, and serve to illustrate common fluorescence imaging practices.

At the upper end, substantially larger fluorescence datasets are somewhat less common, and usually associated with explicitly high-content/throughput workflows. This is also reflected in Figure 3.7: the 655 cell adjacent-LNCaP dataset from Lapan et al. (2009) lies close to the full extent of the LNCaP curve in Figure 3.7(a), whereas the 2773 cell non-adjacent-LNCaP subset from the same study and the ≥ 3500 nucleus cell-cycle analysis reported by Narotamo et al. (2021) exceed the plotted range for LNCaP and, in the latter case, also exceed the PNT2-C2 range [246, 386]. These are important precedents, but arise from workflows explicitly designed for scale, often making use of specialised imaging solutions designed specifically with these kinds of assays in mind, and should not be treated as the default for conventional scope-controlled fluorescence microscopy studies.

Relative to this fluorescence literature, the present non-permeabilised dataset ($N_{\max} = 657$ for LNCaP and $N_{\max} = 2378$ for PNT2-C2) therefore sits favourably within the published range. As seen directly in Figure 3.7(a), the LNCaP dataset reaches the upper end of many conventional fluorescence studies, though falls short of some of the sample sizes used in dedicated high-throughput studies. Likewise, Figure 3.7(b) shows that the PNT2-C2 dataset extends well beyond the sample sizes used in most targeted fluorescence studies, and approaches the lower end of several high-content analyses. This positions the present dataset well: it is large by the standards of many fluorescence microscopy studies, though remains below the most throughput-optimised systems.

Importantly, however, the stronger claim in this chapter does not arise from literature comparison alone. Rather, Figure 3.7 should be interpreted as contextual support for the more direct result of the convergence framework: under the specific non-permeabilised fluorescence workflow used here, the available nucleus counts were sufficient to estimate the selected DAPI-derived descriptors at the predefined precision target. Cross-study comparisons of sample-size are therefore informative, but not evidential in and of themselves, because preparation state, descriptor family, segmentation rules, imaging design, and underlying biological heterogeneity differ across studies.

Limitations

The convergence based analysis quantifies the precision of sample-mean estimates and provides a metric-specific, relative proxy for within-line variability. It does not, by itself, identify the mechanistic sources of dispersion. Plausible contributors include clonal diversity and evolving culture composition [385], heterogeneity in DNA content and chromatin state associated with cell-cycle distribution and replication stress [193, 194, 203, 206, 376–380], and nuclear architecture abnormalities linked to malignancy [9, 131, 381–384, 387, 388]. In LNCaP specifically, heterogeneity may also be shaped by androgen receptor pathway dynamics and subpopulations with differential androgen responsiveness [172, 370, 389–393]. These caveats qualify, but do not invalidate, the dispersion differences summarised in Table 3.1.

Also note that DAPI-derived measures are sensitive to fixation, dye concentration, and imaging parameters, and therefore require careful protocol standardisation if observed differences are to be interpreted as biological rather than technical [8, 12]. Finally, nuclei within a given experiment do not necessarily constitute independent biological replicates in the classical sense; accordingly, the present convergence results are best interpreted as within-sample population characterisation rather than inferential statistics over biological replicates (Table 3.1; [301]).

3.2 Morphological and intensity analysis

This section reports quantitative comparisons between LNCaP and PNT2-C2 nuclei in the non-permeabilised condition following confirmation from the convergence analysis that the sampled nucleus counts were sufficient for stable estimation of the reported metrics (Section 3.1.3 and Table 3.1). Unless otherwise stated, the uncertainties reported here refer to standard errors. The primary aim is to describe robust differences in nuclear morphology and DAPI-signal characteristics between PNT2-C2 and LNCaP, and to provide a limited biological interpretation in the context of the disease state.

3.2.1 Morphology

Morphology here refers to the DAPI-derived nucleus-level descriptors defined previously in the Methods: projected nuclear area (Equation (2.7)), nuclear aspect ratio (Equation (2.8)), and the coefficient of variation as the relative dispersion measure (Equation (2.27)). These descriptors are summarised in Figure 3.8, with the full area and aspect ratio distributions examined subsequently in Section 3.2.1. This separation distinguishes mean morphological shift (for example, larger nuclei) from morphological heterogeneity (for example, broader distributions), both of which are relevant to pleomorphism and nuclear atypia in prostate cancer [132, 135, 184, 188, 359].

Nuclear enlargement and reduced circularity in LNCaP

As summarised in Figure 3.8, LNCaP nuclei were larger on average than PNT2-C2 nuclei, with mean area increasing from $(136 \pm 1)\,\mu m^2$ in PNT2-C2 to $(195 \pm 3)\,\mu m^2$ in LNCaP. This corresponds to an absolute increase of approximately $59\,\mu m^2$ and a relative increase of approximately 45%. Mean nuclear aspect ratio likewise increased from 1.443 ± 0.006 in PNT2-C2 to 1.72 ± 0.02 in LNCaP, corresponding to an absolute increase of approximately 0.28 and a relative increase of approximately 20% (Figure 3.8).

Alongside the arithmetic mean, Figure 3.8 also reports the mode, defined here as the KDE peak position (Equation (2.34)). For both nuclear area and aspect ratio, the mean and mode-based comparisons were directionally concordant: LNCaP showed larger modal area and larger modal aspect ratio than PNT2-C2, with approximate increases of 30% and 10% respectively. However, in both cell lines the mean exceeded the mode, with a substantially larger mean-mode separation in LNCaP than in PNT2-C2. This is the pattern expected for positively skewed distributions in which a minority of very large or elongated nuclei displace the arithmetic mean towards larger values. The underlying distributional basis for this interpretation is shown explicitly in Figures 3.9, 3.10, 3.12 and 3.13.

Taken together, the morphology shifts summarised in Figure 3.8 and examined distributionally in Section 3.2.1 are consistent with malignancy-associated nuclear atypia, in which cancer cells frequently exhibit nucleomegaly and deviations from regular nuclear shape [131, 132, 312, 394]. In prostate

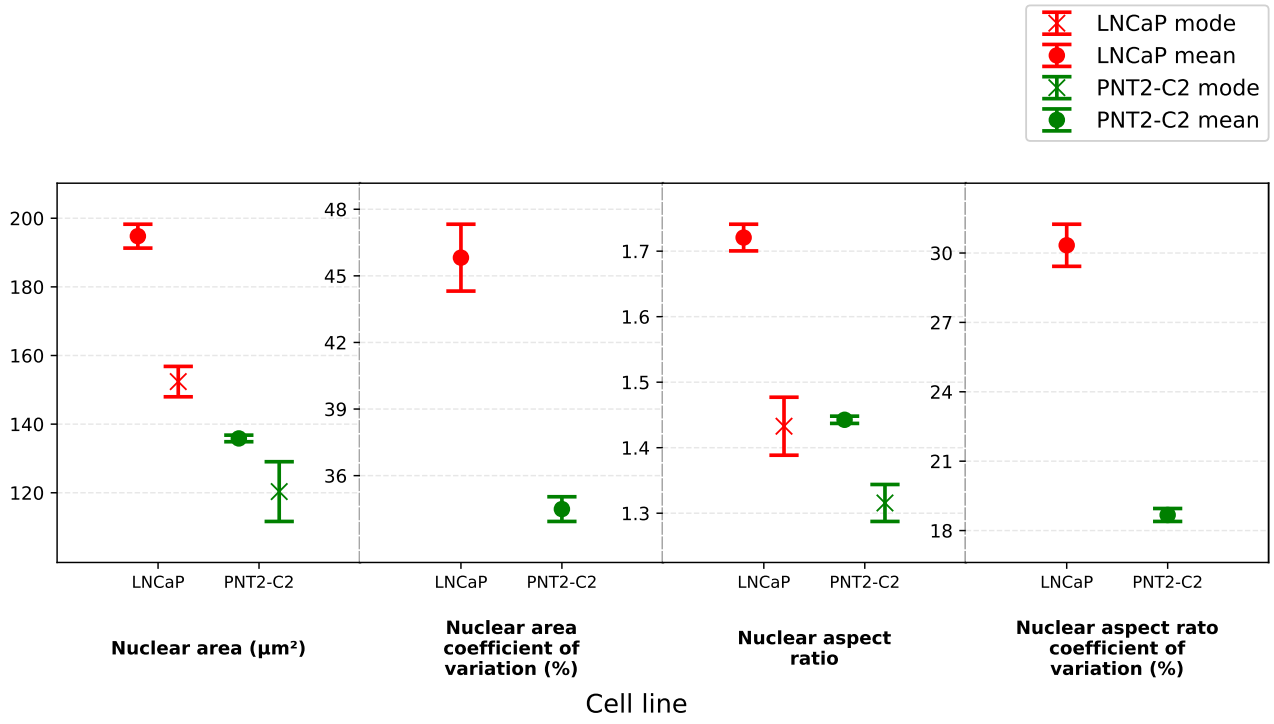


Figure 3.8: Plot showing the mean, mode and coefficient of variation (CV) of nuclear area and nuclear aspect ratio in non-permeabilised LNCaP and PNT2-C2 cells. Mode here refers to the peak position on the x-axis of a kernel density estimate for each metric (Equation (2.34)), consistent with the distributional analyses in Section 3.2.1. Error bars represent standard error calculated using the delta method where required for CV and mode estimates (Equations (2.28) and (2.29)). LNCaP nuclei had a mean area of $(195 \pm 3)\,\mu m^2$, modal area of $(152 \pm 4)\,\mu m^2$, area CV of $(46 \pm 2)\,\%$, mean aspect ratio of 1.72 ± 0.02, modal aspect ratio of 1.43 ± 0.04, and aspect ratio CV of $(30.3 \pm 0.9)\,\%$, while PNT2-C2 nuclei had a mean area of $(136 \pm 1)\,\mu m^2$, modal area of $(120 \pm 9)\,\mu m^2$, area CV of $(34.5 \pm 0.6)\,\%$, mean aspect ratio of 1.443 ± 0.006, modal aspect ratio of 1.32 ± 0.03, and aspect ratio CV of $(18.7 \pm 0.3)\,\%$.

cancer specifically, nuclear size and shape have long been linked to grading, prognosis, and malignant progression, motivating the expectation that a metastatic prostate cancer model such as LNCaP should show more abnormal nuclear morphology than a non-malignant epithelial comparator such as PNT2-C2 [133, 359, 360, 395, 396].

A mechanistically plausible contributor to the nuclear enlargement quantified in Figures 3.8 to 3.10 is altered genome content and instability in LNCaP. LNCaP was originally described as aneuploid, and subsequent work indicates complex and heterogeneous genomic states within the line [1, 2, 14, 165]. Because nuclear size is constrained by nucleocytoplasmic scaling and can vary with ploidy and chromatin state, such genomic alterations provide a plausible route to increased mean and modal nuclear area [397–400].

Rizzotto et al. (2024) identified nuclear size as a biologically informative feature with strong links to prostate-cancer grade, and reported that expression of specific proteins correlated strongly with nuclear size as well as with metastatic-associated behaviours including proliferation, migration, and adhesion [192]. They found that DHRS7 was downregulated in more aggressive/metastatic prostate-cancer contexts, and that experimental reduction of DHRS7 increased nuclear size whereas increased DHRS7 expression reduced it, with area changes of approximately 25% [192]. Although DHRS7 status was not measured in the present study, these results provide a concrete example of how molecular changes can generate the type of nuclear-size shift observed here in Figures 3.8 to 3.10. As such, altered gene-expression programmes remain a plausible route by which cancer cells may develop nuclear

pleomorphism.

Lamin A/C is also a major determinant of nuclear stiffness and shape stability under mechanical load; lamin A/C deficiency is associated with misshapen nuclei and reduced nuclear stiffness [330, 401]. Changes in lamin A/C expression are known to occur in prostate cancer and are strongly linked to metastatic potential [402, 403]. Modelling work has further indicated that bleb formation can emerge from non-uniform lamin meshwork mechanics and local instabilities [404]. These findings provide a mechanistic route from altered lamina composition to the pleomorphic nuclear contours implied by the enlarged area and increased aspect ratio observed here (Figures 3.8, 3.12 and 3.13). If abnormal nuclear shape reflects weakened envelope integrity or elevated mechanical stress, it may also increase the probability of envelope rupture or persistent bleb-associated damage, consistent with previous work showing that nuclear-envelope rupture can expose DNA to cytoplasmic factors and promote DNA damage [405].

However, 2D projected nuclear area is also influenced by nuclear flattening during adhesion and cell spreading. Accordingly, the area differences quantified in Figures 3.8 to 3.10 should not be interpreted as direct volume measurements without qualification, and may also reflect cell line-specific mechanics and substrate interactions rather than nuclear volume alone [367, 406].

Increased morphological heterogeneity in LNCaP: implications of CV

Beyond mean differences, LNCaP nuclei were also substantially more heterogeneous in both size and elongation, shown by the CV summaries in Figure 3.8. The nuclear area CV increased from $(34.5 \pm 0.6)\,\%$ in PNT2-C2 to $(46 \pm 2)\,\%$ in LNCaP, corresponding to a relative increase of approximately 30%. The nuclear aspect ratio CV likewise increased from $(18.7 \pm 0.3)\,\%$ in PNT2-C2 to $(30.3 \pm 0.9)\,\%$ in LNCaP, corresponding to a relative increase of approximately 60% (Figure 3.8).

This increase in CV is consistent with nuclear pleomorphism interpreted as a heterogeneity descriptor. In prostate morphometry, variance-based measures of nuclear irregularity and heterogeneity have previously been associated with progression risk and clinically relevant endpoints [135, 188, 381, 382], and previous work has explicitly linked the coefficient of variation of morphological quantities to heterogeneity [134, 407]. More broadly, nuclear-size variability has been linked to genomic-instability metrics across cancers, supporting the interpretation that the increased dispersion seen here in Figure 3.8 and in the slower convergence of Section 3.1 reflects genuine biological heterogeneity rather than measurement noise alone [115, 408].

In the specific context of LNCaP, elevated heterogeneity is plausible because bulk genomic data indicates that LNCaP comprises multiple subpopulations, which can broaden morphology distributions even under identical culture conditions [14]. In addition, nuclear size and shape are both cell-cycle and mechanics-dependent, so the broader distributions seen in Figures 3.8 to 3.10, 3.12 and 3.13 may reflect broader cell-cycle occupancy and/or a wider distribution of mechanical states, including adhesion, cytoskeletal tension, and chromatin compaction, all of which are commonly dysregulated in cancer [131, 377, 380, 397].

A key caveat is that PNT2-C2 is SV40-immortalised [3], and SV40-driven immortalisation perturbs p53/RB control and related cell-cycle checkpoints [169, 170]. Consequently, PNT2-C2 should not be treated as directly equivalent to primary healthy prostate epithelium. The morphology contrasts

quantified here against PNT2-C2 (Figure 3.8 and Section 3.2.1) are therefore best interpreted as differences relative to a non-malignant immortalised comparator, not as absolute differences relative to normal primary tissue. This caveat may mean that the malignant-versus-non-malignant contrast is, if anything, underestimated.

Distributional analysis using histograms

While Figure 3.8 summarises mean behaviour and relative dispersion, the nucleus-level histograms in Figures 3.9, 3.10, 3.12 and 3.13 provide the corresponding distributional structure. These plots are needed not only to visualise skewness, tail mass, and possible subpopulation structure, but also because the reported modal values are defined from the KDE peak rather than from histogram-bin counts (Equation (2.34)). This distributional perspective is motivated by prior morphometry work showing that population-level means and CVs can obscure rare but diagnostically meaningful subpopulations [174].

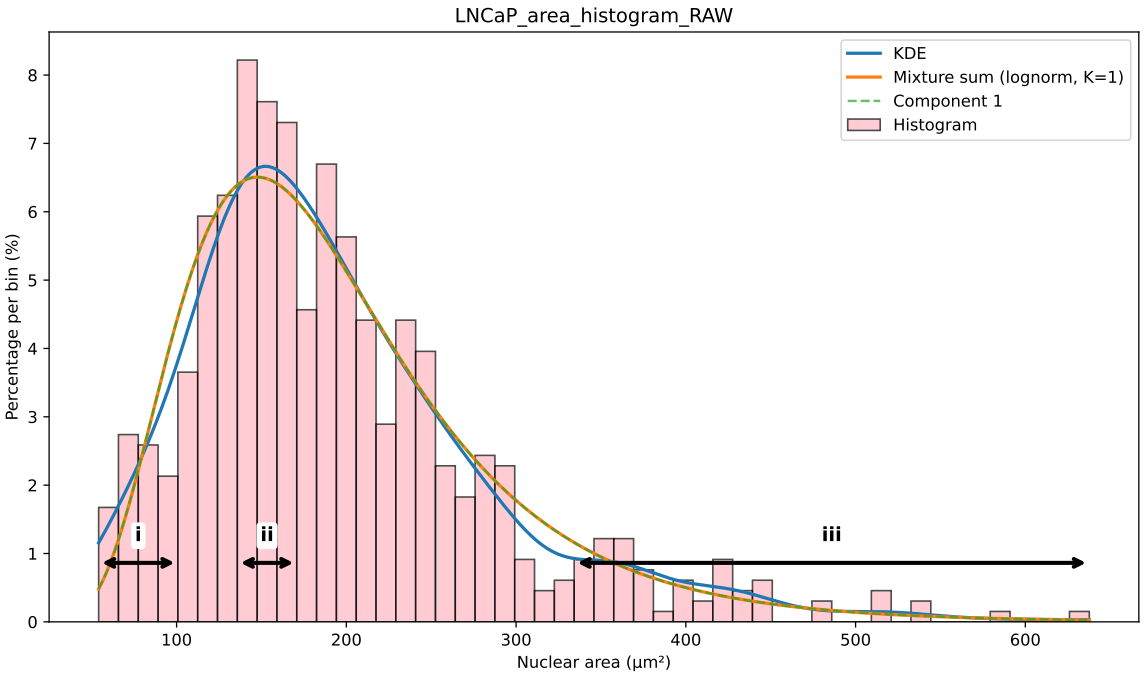
Because both area and aspect ratio are strictly non-negative, normal distributions should not be assumed. Instead, the observed distributions were evaluated using the KDE and ICL-based framework defined previously in Methods (Section 2.3.5 and Equation (2.34)) and visualised here in Figures 3.9, 3.10, 3.12 and 3.13. In this context, a unimodal but heavy-tailed distribution is consistent with a continuum of morphological states, whereas multimodality would suggest discrete subpopulations. Extreme tails may indicate rare abnormal nuclei that are biologically meaningful, segmentation artefacts, or a mixture of both.

The nuclear area histograms in Figures 3.9 and 3.10 move beyond the bulk summaries of Figure 3.8 by explicitly modelling the full nucleus-level area distribution within each cell line. Each segmented nucleus contributes one projected-area value, and the KDE-based distributional analysis therefore describes how nuclear area is distributed across the measured population rather than only how its mean and CV differ between cell lines.

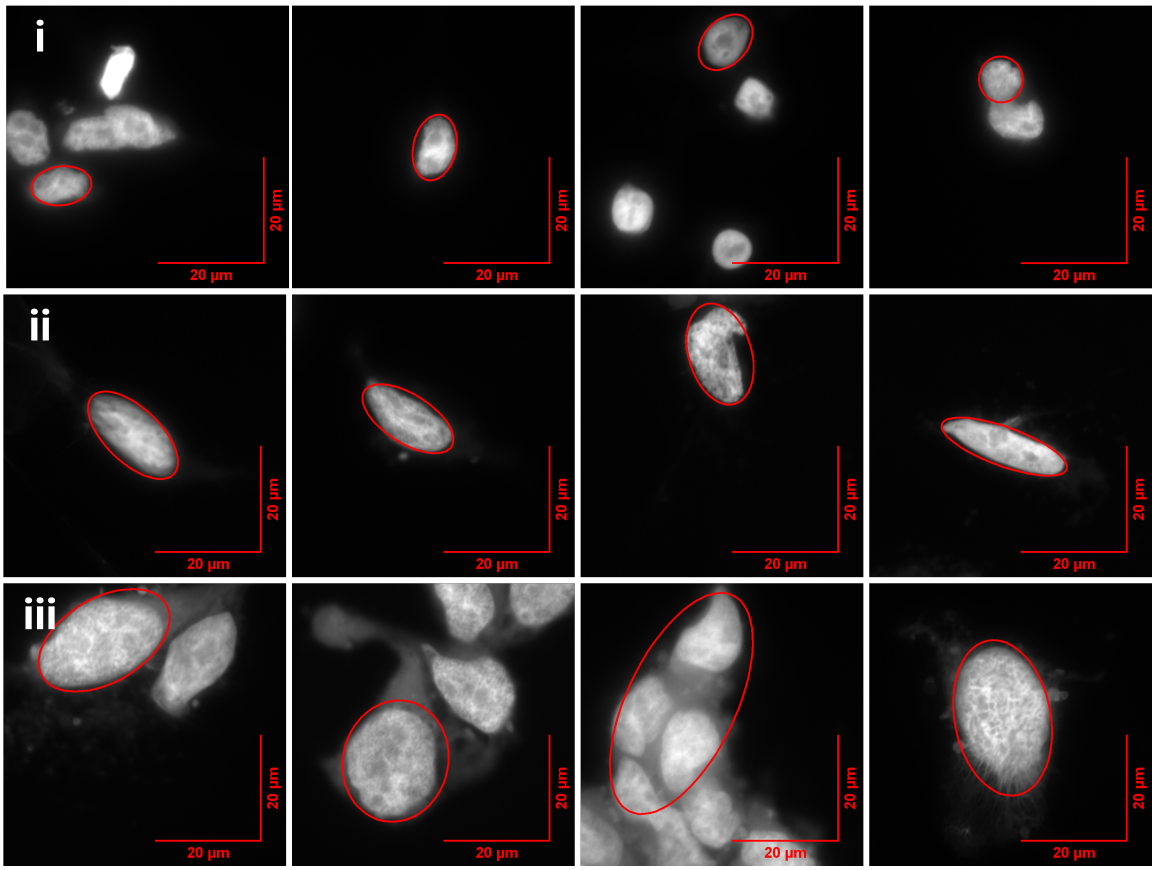
Using the modelling framework defined previously in Section 2.3.5, ICL-based model selection favoured a single log-normal component for nuclear area in both LNCaP and PNT2-C2 (Figures 3.9 and 3.10). The fitted-parameter values are reported in the corresponding figure captions. In the present section, the fitted distribution is used primarily as a compact parametric description of the overall histogram shape, while interpretation of central tendency and peak location remains anchored to the observed data.

Comparing the arithmetic mean with the fitted-distributional mean shows close agreement in both cell lines, supporting the use of the sample mean as the primary central-tendency descriptor. In LNCaP, the arithmetic mean was $(195 \pm 3)\,\mu m^2$ and the fitted-distribution mean was $(197 \pm 4)\,\mu m^2$; in PNT2-C2, the corresponding values were $(136 \pm 1)\,\mu m^2$ and $(136.2 \pm 0.9)\,\mu m^2$ (Figures 3.8 to 3.10). Because these estimates agree within uncertainty, all subsequent comparisons use the arithmetic mean.

For peak location, the mode can be reported either from the fitted log-normal curve or from the KDE. In the present area data, these two estimates were consistent within uncertainty in both cell lines, but the KDE mode was retained as the primary modal descriptor because it is defined directly from the empirical density estimate rather than from the fitted parametric approximation (Equation (2.34) and Figures 3.9(a) and 3.10(a)). Numerically, the KDE mode was $(152 \pm 4)\,\mu m^2$ in

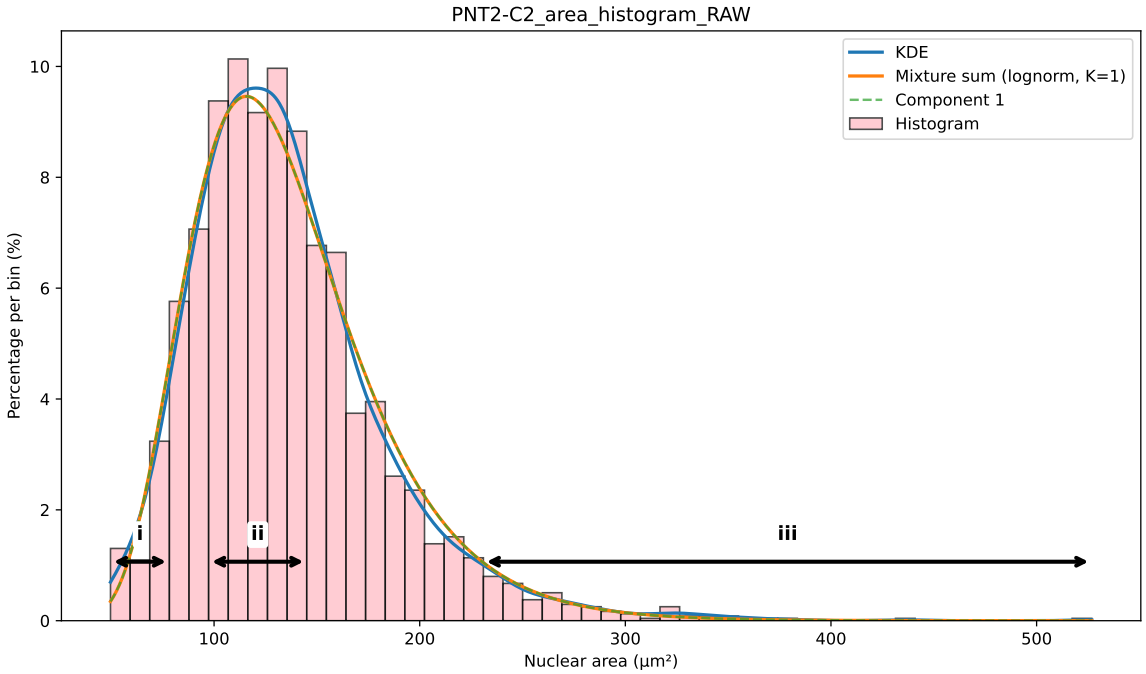


(a) Histogram showing the area in each LNCaP nucleus. A KDE is overlaid on top of the histogram, with fitting applied to the KDE to determine the overall distribution of the data [258]. The best model (chosen by ICL [259]) was a single lognormal distribution with s = 0.44 ± 0.01, loc = 0, and scale = 179 ± 3, with a peak at $(148 \pm 3)\,\mu m^2$. The KDE peak location was $(152 \pm 4)\,\mu m^2$.

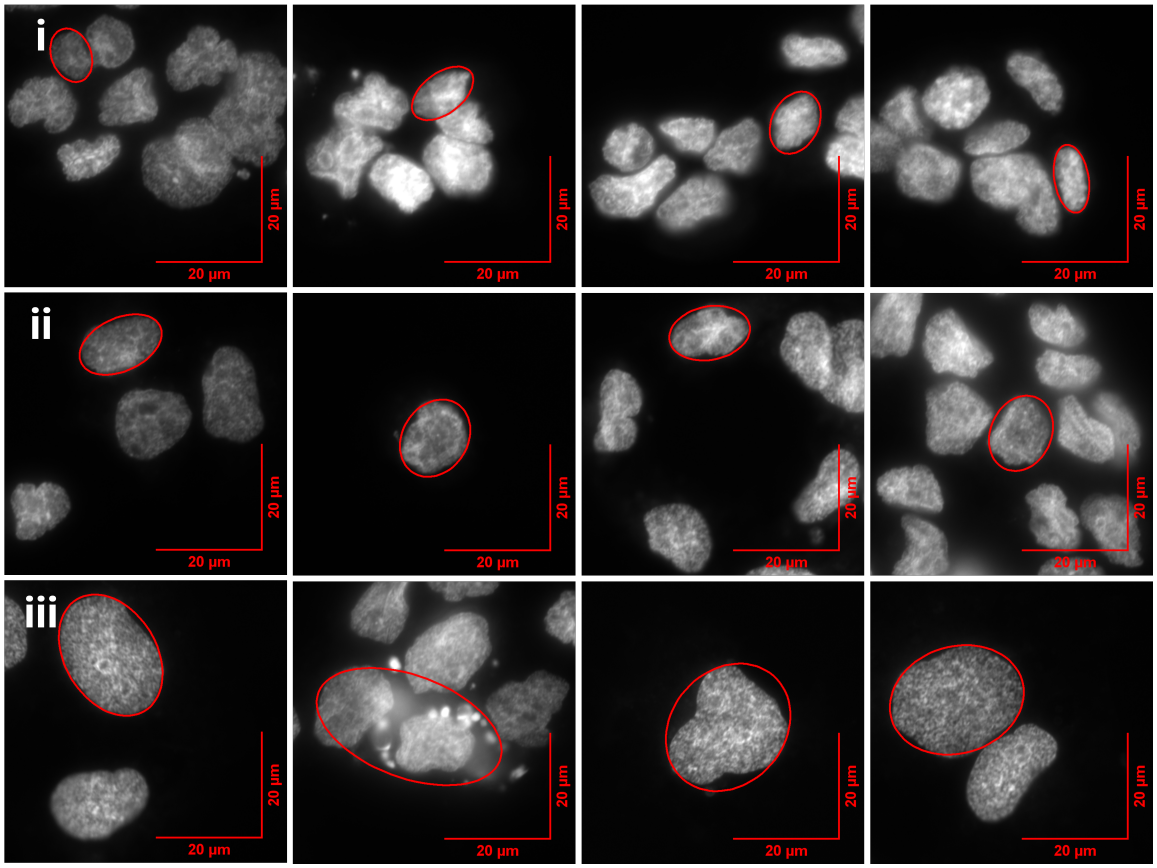


(b) Images of detected nuclei selected from the low (i), modal (ii), and high (iii) regions of the corresponding histogram. Red ellipses are drawn purely for illustrative purposes to show the approximate region which was detected as a nucleus by the segmentation code. Area was calculated using pixel counts within the nuclei mask. A 20 µm scale bar is shown in red at the bottom right corner of each image.

Figure 3.9: Histogram showing the nuclear area in non-permeabilised LNCaP cells ((a)), with some images of representative nuclei from the low, modal, and high regions of the histogram ((b)). The KDE is used as a basis for fitting, with the histogram being purely illustrative. The best fit was chosen using an Integrated Classification Likelihood (ICL) [259]. Uncertainties are calculated from the covariance matrix.



(a) Histogram showing the area in each PNT2-C2 nucleus. A KDE is overlaid on top of the histogram, with fitting applied to the KDE to determine the overall distribution of the data [258]. The best model (chosen by ICL [259]) was a single lognormal distribution with $s = 0.329 \pm 0.005$, loc $= 0$, and scale $= 129.0 \pm 0.9$, with a peak at $(115.7 \pm 0.9)\,\mu m^2$. The KDE peak location was $(120 \pm 9)\,\mu m^2$.



(b) Images of detected nuclei selected from the low (i), modal (ii), and high (iii) regions of the corresponding histogram. Red ellipses are drawn purely for illustrative purposes to show the approximate region which was detected as a nucleus by the segmentation code. Area was calculated using pixel counts within the nuclei mask. A 20 µm scale bar is shown in red at the bottom right corner of each image.

Figure 3.10: Histogram showing the nuclear area in non-permeabilised PNT2-C2 cells ((a)), with some images of representative nuclei from the low, modal, and high regions of the histogram ((b)). The KDE is used as a basis for fitting, with the histogram being purely illustrative. The best fit was chosen using an Integrated Classification Likelihood (ICL) [259]. Uncertainties are calculated from the covariance matrix.

LNCaP and $(120 \pm 9)\,\mu m^2$ in PNT2-C2, while the fitted modes were $(148 \pm 3)\,\mu m^2$ and $(115.7 \pm 0.9)\,\mu m^2$, respectively (Figures 3.9 and 3.10). The fitted modes are retained in the captions to document the model fully, but the quantity referred to as the mode in the main text is the KDE mode.

The typical nuclear area is materially larger in LNCaP, and this is visible directly in the central regions of Figures 3.9(a) and 3.10(a). The KDE mode increased from $(120 \pm 9)\,\mu m^2$ in PNT2-C2 to $(152 \pm 4)\,\mu m^2$ in LNCaP, corresponding to an approximate 25% positive shift in peak location. The arithmetic mean likewise increased from $(136 \pm 1)\,\mu m^2$ to $(195 \pm 3)\,\mu m^2$, corresponding to an increase of approximately 43% (Figures 3.8 to 3.10). The larger shift in the mean than in the mode indicates that LNCaP is not only shifted towards larger typical nuclei, but also exhibits a more influential positive tail, consistent with increased nucleus-level size heterogeneity in the malignant model [131, 381].

The shapes of the two area distributions also differ in a biologically informative way (Figures 3.9(a) and 3.10(a)). All model-selection criteria supported a single-component fit, providing negative evidence for clearly separable area-defined subpopulations at the present assay resolution. However, the LNCaP distribution exhibits a heavier positive tail. The fitted log-space dispersion was larger in LNCaP, the mean-mode separation was larger, and high-area tail measures were markedly enriched: approximately 10% of LNCaP nuclei exceeded $300\,\mu m^2$ vs. only $\sim 1\%$ of PNT2-C2 nuclei, and the 90th percentile was approximately $315\,\mu m^2$ in LNCaP vs. $195\,\mu m^2$ in PNT2-C2, consistent with increased pleomorphism/heterogeneity in the cancer-derived line (Figures 3.9 and 3.10) [132, 395]. These tail differences are largely invisible in mean ± SEM reporting alone, yet they provide direct evidence of increased nuclear pleomorphism.

A practical limitation is that some segmented nuclei correspond to algorithmic mergers of multiple real nuclei into a single mask, as illustrated in Figures 3.9(b) and 3.10(b). These failures appear more prevalent in LNCaP, plausibly because extranuclear DAPI staining more often obscures nuclear boundaries in that line (Figure 2.7). Such failures preferentially inflate the extreme right tail, but they cannot fully account for the central peak shift of approximately 25%, which arises from the high-density bulk of the distribution, rather than from rare extreme objects (Figures 3.9(a) and 3.10(a)). A conservative interpretation is therefore that the LNCaP high-area tail contains a mixture of genuinely large nuclei and merged segments. Because these classes overlap in size, a hard upper-area exclusion threshold would risk discarding biologically real large nuclei alongside artefacts.

Nuclear enlargement and increased morphological variability are well-established features of malignant transformation in prostate cancer, and quantitative nuclear morphometry has long been investigated as an objective indicator of grade and risk in tissue-based settings [132, 395]. The present analysis extends that logic to a controlled cell-line system by combining convergence-verified single-nucleus sampling with explicit distributional modelling and representative image examples anchored to low, modal, and high-area regions of the histograms (Section 3.1 and Figures 3.9 and 3.10). Under this framework, the non-permeabilised results support a specific statement: relative to PNT2-C2, LNCaP shows both a systematic shift to larger typical nuclear area and a heavier high-area tail while remaining unimodal under conservative mixture selection.

The nuclear area comparison above is intrinsically unbalanced in sample size ($N_{\text{PNT2-C2}} = 2378$ vs. $N_{\text{LNCaP}} = 657$). To isolate the effect of unequal N from genuine distributional differences, an additional control analysis was performed in which the PNT2-C2 dataset was randomly sub-sampled without replacement up to $N = 657$ nuclei, matching the available LNCaP count.

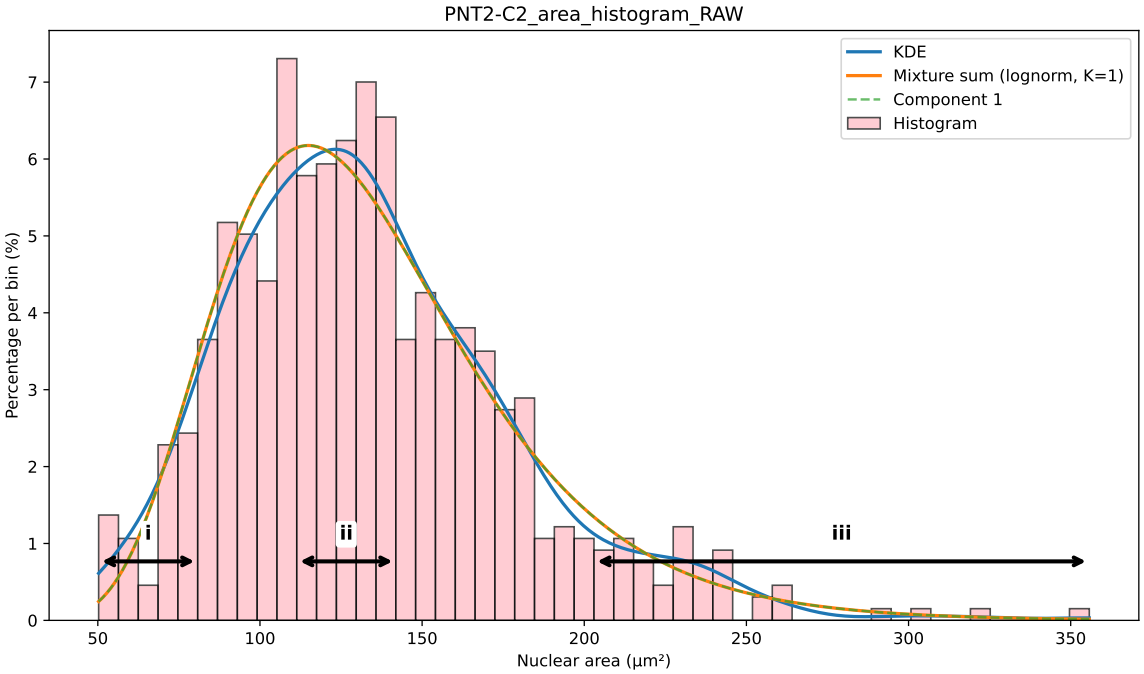


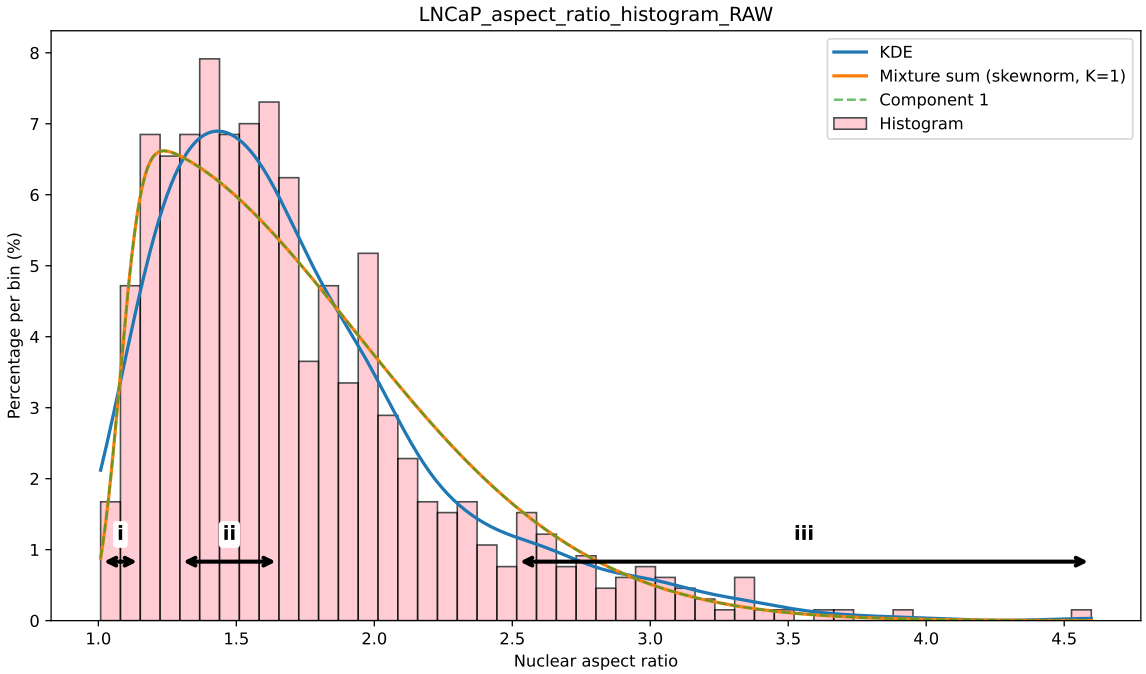
Figure 3.11: Histogram showing the nuclear area in each non-permeabilised PNT2-C2 nucleus using a random selection of nuclei to match the available number of nuclei in the LNCaP dataset. A KDE is overlaid on top of the histogram, with fitting applied to the KDE to determine the overall distribution of the data [258]. The best model (chosen by ICL [259]) was a single log-normal distribution with s = 0.326 ± 0.009, loc = 0, and scale = 128 ± 2, with a peak at $(115 \pm 2)\,\mu m^2$. The KDE peak location was $(123 \pm 4)\,\mu m^2$.

The matched-N PNT2-C2 histogram in Figure 3.11 retains the same unimodal, right-skewed form, and is again best described by a single log-normal density. The fitted modal location was $(115 \pm 2)\,\mu m^2$ and the KDE mode was $(123 \pm 4)\,\mu m^2$, both closely consistent with the full-PNT2-C2 analysis. The arithmetic mean also remained stable at $(134 \pm 2)\,\mu m^2$ vs. $(136 \pm 1)\,\mu m^2$ in the full dataset. More generally, the fitted width parameters changed by only approximately 1%, within fitting uncertainty. This demonstrates that the central tendency and overall spread of the PNT2-C2 area distribution are not materially altered by reducing N to match the LNCaP dataset (Figures 3.10 and 3.11).

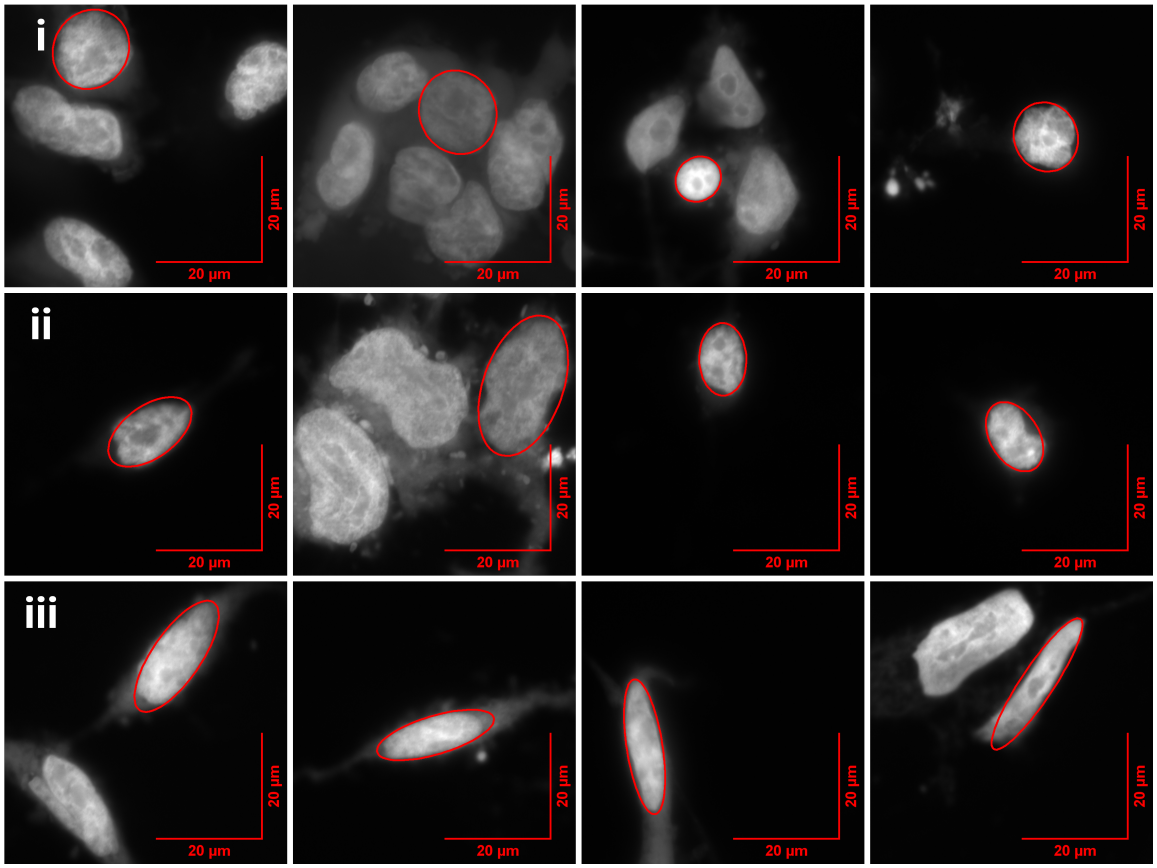
Tail descriptors support the same interpretation. In the matched-N subset, approximately 0.5 % of nuclei had area $\geq 300\,\mu m^2$, compared with $\sim 1\%$ in the full PNT2-C2 dataset (Figures 3.10 and 3.11). Given the rarity of such events, this modest difference is expected from sampling variation rather than from any substantive change in the underlying distribution. Combined with the independent convergence result for area (Section 3.1), this control supports the conclusion that the LNCaP-PNT2-C2 differences reported above are not an artefact of the smaller N available for LNCaP.

As such, subsequent histogram-based analyses retain all available nuclei for each cell line rather than enforcing matched-N by discarding valid PNT2-C2 data. This preserves sensitivity to rare-event tail behaviour and to any genuine subpopulation structure should it exist. Under that justification, an explicit LNCaP-versus-PNT2-C2$_{N=657}$ comparison would add little interpretative content beyond the already reported full-distribution comparison, while reducing precision.

Nuclear aspect ratio was defined previously in Methods as the ratio of the major to minor axes of the fitted ellipse for each segmented nuclear mask (Equation (2.8)). The corresponding histograms for LNCaP and PNT2-C2 are shown in Figures 3.12 and 3.13. In both cell lines, ICL model selection favoured a single skew-normal component, indicating that the observed shape variability is better

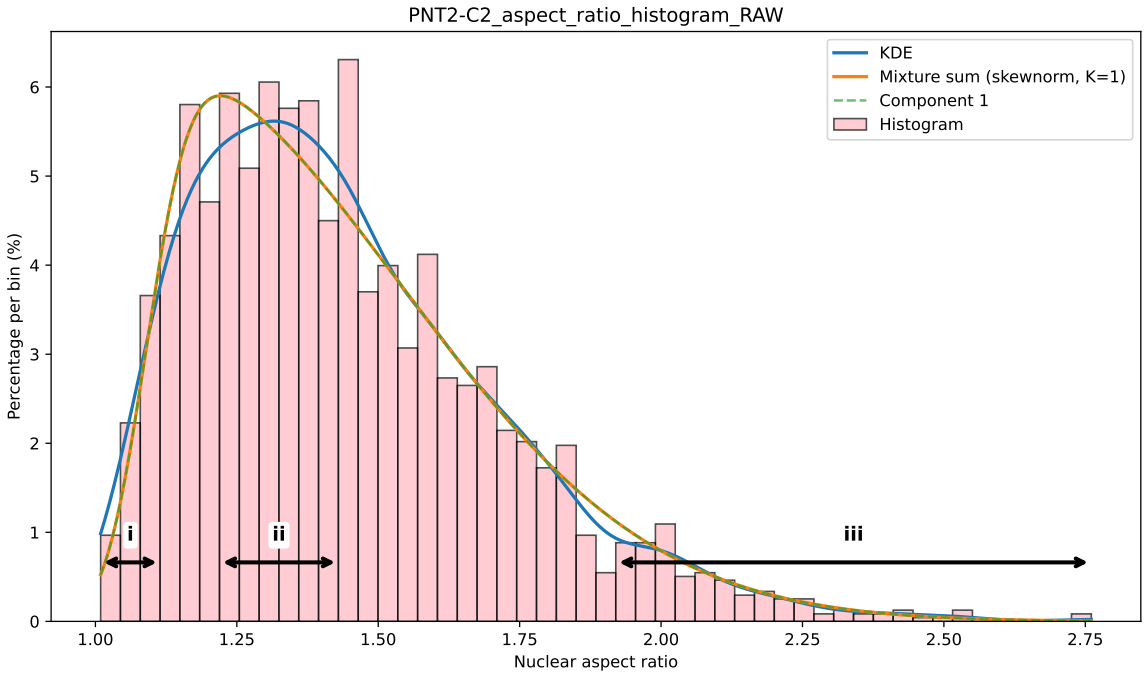


(a) Histogram showing the aspect ratio in each LNCaP nucleus. A KDE is overlaid on top of the histogram, with fitting applied to the KDE to determine the overall distribution of the data [258]. The best model (chosen by ICL [259]) was a single skew-normal distribution with a $= 14 \pm 4$, loc $= 1.08 \pm 0.02$, and scale $= 0.85 \pm 0.03$, with a peak at 1.24 ± 0.05. The KDE peak location was 1.43 ± 0.04.

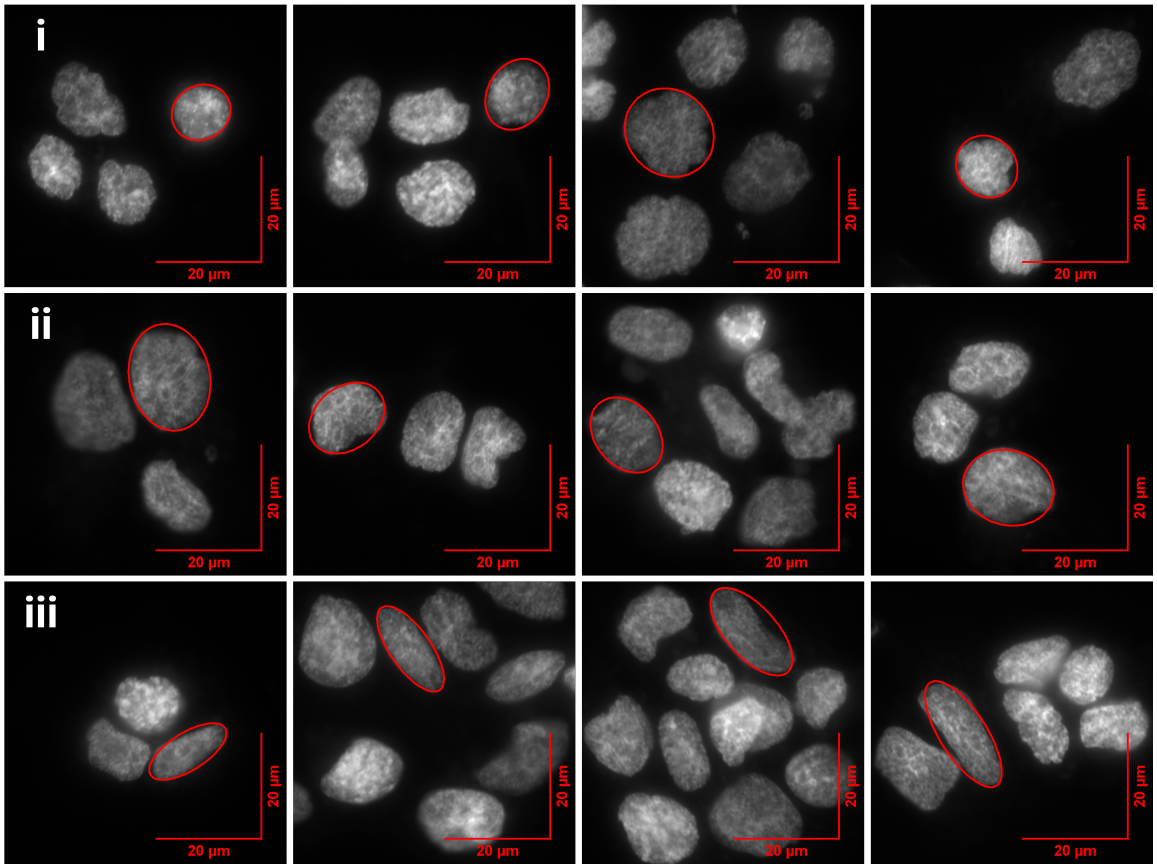


(b) Images of detected nuclei selected from the low (i), modal (ii), and high (iii) regions of the corresponding histogram. Red ellipses are drawn purely for illustrative purposes to show the approximate region which was detected as a nucleus by the segmentation code. Aspect ratio was calculated as the ratio of the major:minor axis of the fitted ellipse (which is distinct from the ellipses indicated here). A 20 µm scale bar is shown in red at the bottom right corner of each image.

Figure 3.12: Histogram showing the nuclear aspect ratio in non-permeabilised LNCaP cells ((a)), with some images of representative nuclei from the low, modal, and high regions of the histogram ((b)). The KDE is used as a basis for fitting, with the histogram being purely illustrative. The best fit was chosen using an Integrated Classification Likelihood (ICL) [259]. Uncertainties are calculated from the covariance matrix.



(a) Histogram showing the aspect ratio in each PNT2-C2 nucleus. A KDE is overlaid on top of the histogram, with fitting applied to the KDE to determine the overall distribution of the data [258]. The best model (chosen by ICL [259]) was a single skew-normal distribution with a $= 7.5 \pm 0.8$, loc $= 1.091 \pm 0.006$, and scale $= 0.447 \pm 0.008$, with a peak at 1.22 ± 0.01. The KDE peak location was 1.32 ± 0.03.



(b) Images of detected nuclei selected from the low (i), modal (ii), and high (iii) regions of the corresponding histogram. Red ellipses are drawn purely for illustrative purposes to show the approximate region which was detected as a nucleus by the segmentation code. Aspect ratio was calculated as the ratio of the major:minor axis of the fitted ellipse (which is distinct from the ellipses indicated here). A 20 µm scale bar is shown in red at the bottom right corner of each image.

Figure 3.13: Histogram showing the nuclear aspect ratio in non-permeabilised PNT2-C2 cells ((a)), with some images of representative nuclei from the low, modal, and high regions of the histogram ((b)). The KDE is used as a basis for fitting, with the histogram being purely illustrative. The best fit was chosen using an Integrated Classification Likelihood (ICL) [259]. Uncertainties are calculated from the covariance matrix.

described as a continuous spectrum than as discrete, clearly separable subpopulations. For PNT2-C2, AIC, BIC, and ICL all supported the same one-component skew-normal solution. In LNCaP, AIC favoured a three-component log-normal fit, but the resulting peaks were highly overlapping and did not match the observed KDE well. This supports use of ICL as the primary model-selection criterion here, because the aim is to identify clearly interpretable subpopulations rather than to add overlapping components merely to improve numerical approximation of the empirical curve (Figures 3.12(a) and 3.13(a)).

As for nuclear area, the mode for aspect ratio is taken as the KDE peak position, while the mean is computed directly from the numerical data (Equation (2.34) and Figures 3.12(a) and 3.13(a)). The modal and mean aspect ratios were 1.43 ± 0.04 and 1.72 ± 0.02 in LNCaP, vs. 1.32 ± 0.03 and 1.443 ± 0.006 in PNT2-C2. Thus, the modal aspect ratio was approximately 10% higher in LNCaP, indicating that the typical LNCaP nucleus is less circular. In both cell lines the mean exceeded the mode, consistent with right-skewness, but the mean-mode separation was substantially larger in LNCaP. This provides a quantitative indication that the high-aspect-ratio tail contributes more strongly in the malignant model (Figures 3.8, 3.12 and 3.13).

While both aspect-ratio distributions remain unimodal, the LNCaP distribution is visibly broader and contains a much larger high-aspect-ratio tail (Figures 3.12(a) and 3.13(a)). Approximately 25% of LNCaP nuclei had aspect ratio ≥ 2, compared with only $\sim 5\%$ of PNT2-C2 nuclei. The most irregular measured PNT2-C2 nucleus had aspect ratio ~ 2.75, whereas LNCaP included many nuclei beyond that threshold, consistent with the representative images in Figures 3.12(b) and 3.13(b). The 90th percentile likewise increased from 1.82 in PNT2-C2 to 2.47 in LNCaP. Accordingly, the difference between cell lines is not limited to a central shift in typical shape; it also includes substantial enrichment of rare, highly elongated nuclei in LNCaP.

As discussed previously (Section 3.2.1), nuclear-shape descriptors such as roundness factor and shape variance have long been associated with clinically aggressive prostate cancer and adverse outcomes, including metastasis [395, 409, 410]. Within that established context, the present aspect ratio results add two explicit pieces of distributional evidence. First, the modal LNCaP nucleus is less circular than the modal PNT2-C2 nucleus. Second, LNCaP contains a quantitatively enlarged high-aspect-ratio tail, implying a higher frequency of extreme nuclear elongation events (Figures 3.8, 3.12 and 3.13). The explicit comparison against a non-malignant comparator is important, because it is often absent from cell line studies.

Mechanistically, increased nuclear elongation and pleomorphism in prostate cancer models are consistent with altered nuclear-envelope organisation and nucleocytoskeletal coupling. Nuclear-shape stability has been linked to metastatic potential via DIAPH3-dependent cytoskeletal programmes [382], and prostate cancer cell systems are known to exhibit differential localisation and expression of nuclear-envelope components that correlate with abnormal nuclear phenotypes [330]. Aggressive prostate cancer nuclei have also been reported to be mechanically softer than non-malignant counterparts [380], providing a plausible physical route to increased deformability under adherent-culture conditions. Together, these mechanisms are consistent with the modal shift and enlarged tail quantified in Figures 3.12 and 3.13.

The most extreme aspect-ratio values correspond to nuclei that appear strongly elongated, as illustrated in Figures 3.12(b) and 3.13(b). Some of these may represent genuine biological states, for example

mechanically deformed nuclei in highly spread cells or nuclei with unstable envelope mechanics. However, sample preparation and image analysis can also contribute: cell shape is sensitive to adhesion geometry, actomyosin tension, and substrate stiffness, and chemical fixation can preserve what would otherwise be transient deformations [411–414]. Segmentation artefacts may also affect the measured aspect ratio. These caveats qualify the interpretation of the extreme tail, but they do not negate the broader distributional differences quantified in Figures 3.12 and 3.13.

While much work on prostate-cancer nuclear morphometry has focused on histological material and on scalar summary descriptors rather than explicitly modelled cell-line distributions [115, 132, 191, 409], the present non-permeabilised analysis combines convergence-verified sampling with KDE-based distribution modelling and objective ICL-guided component selection (Sections 3.1 and 3.2.1). Earlier in vitro work showed that nuclear-shape descriptors could discriminate between prostate-cancer cell lines such as PC3 and DU-145 while also noting marked cultured-cell polymorphism [415], but did not include a non-malignant comparator and did not evaluate the full distributional shape explicitly. In contrast, the present analysis provides a reproducible quantification of nuclear-area and nuclear-shape heterogeneity in LNCaP relative to PNT2-C2 through direct evaluation of modal shifts, tail probabilities, percentile differences, and model-based evidence for unimodality.

Nucleus-level histograms of nuclear area and aspect ratio were therefore used to complement the bulk summaries by interrogating skewness, tail mass, and the presence or absence of clearly separable subpopulations (Figures 3.9, 3.10, 3.12 and 3.13). In both metrics, density estimation was performed using a Gaussian KDE and then modelled parametrically with mixture complexity selected by ICL, as described previously in Section 2.3.5. This framework permits direct comparison of arithmetic means with KDE-defined modes, thereby quantifying how strongly tail behaviour displaces the mean away from the most probable nucleus.

Across both cell lines, nuclear area was best described by a single log-normal component and aspect ratio by a single skew-normal component, providing negative evidence for clearly separable subpopulations in these metrics at the present resolution (Figures 3.9, 3.10, 3.12 and 3.13). Nevertheless, in both descriptors LNCaP exhibited both a positive shift in peak location relative to PNT2-C2 and materially heavier positive tails, reflected by larger mean-mode separation and by increased high-end percentiles and tail probabilities. Collectively, these features constitute distribution-level evidence of increased pleomorphism and heterogeneity in the malignant model, while still favouring a largely unimodal interpretation under conservative mixture selection. Extreme tail events should be interpreted cautiously because they may contain both biologically meaningful nuclei and segmentation or preparation artefacts, particularly in the case of nuclear area.

The present morphology analysis therefore advances beyond conventional mean/CV reporting by explicitly modelling the full nucleus-level distributions using KDE-based density estimation and ICL-guided model selection (Section 3.2.1). This permits formal statements regarding unimodality versus multimodality, skewness, tail mass, and mean-mode separation. The finding that both nuclear area and aspect ratio remain unimodal under conservative model selection, yet exhibit heavier positive tails and larger mean-mode separation in LNCaP, provides a reproducible and parameterised analogue of the qualitative pathological description of malignant nuclei as larger and more pleomorphic (Figures 3.8 to 3.10, 3.12 and 3.13).

3.2.2 Intensity Analysis

In order to compare DNA-associated fluorescence and intranuclear organisation between cell lines, the DAPI-derived intensity descriptors defined previously in Methods were analysed for each segmented nucleus (Equations (2.9), (2.10) and (2.13) to (2.15)). The corresponding bulk summaries are shown in Figure 3.14, while the underlying density and texture-distribution analyses are developed subsequently in Section 3.2.2.

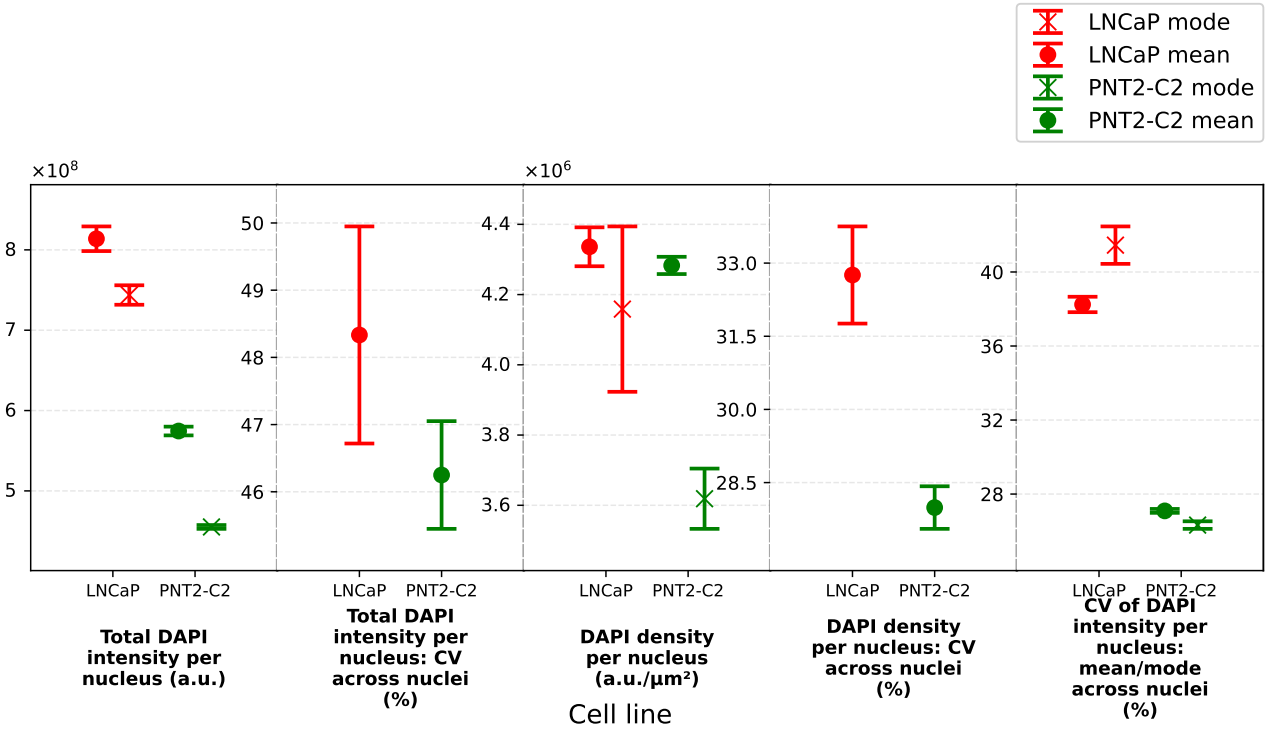


Figure 3.14: Plot showing key DAPI intensity summary statistics for non-permeabilised LNCaP and PNT2-C2 cells. Total DAPI intensity per nucleus was the sum of all pixel intensities in the nuclei mask for a given nucleus. DAPI density was the total intensity divided by nuclear area. CV of DAPI intensity per nucleus was the coefficient of variation of all pixel intensity values in the nuclei mask for a given nucleus.Mode here refers to the peak position of the KDE (Equation (2.34)), consistent with the distributional analyses in Section 3.2.2. Error bars represent standard error, calculated using the delta method where required for CV and mode estimates (Equations (2.28) and (2.29)).

As well as mean values, Figure 3.14 reports KDE modes, i.e. peak locations of the corresponding nucleus-level distributions (Equation (2.34)). Across the plotted intensity metrics, the mean and mode based comparisons were generally directionally concordant, indicating that the observed cell-line differences were not driven solely by a small population of extreme nuclei. Where the mean comparison was weak or ambiguous, however, the KDE mode typically still resolved a shift in the dominant population: for example, mean DAPI density overlapped within uncertainty, whereas the KDE mode remained higher in LNCaP than in PNT2-C2 by approximately 15% (Figures 3.14, 3.16 and 3.17). Conversely, for per-nucleus DAPI intensity heterogeneity, the modal separation in CV_{DAPI} was larger than the mean separation, consistent with non-Gaussian distributional structure (Figures 3.14, 3.18 and 3.19).

Total nuclear DAPI intensity I_{DAPI}, defined previously in Equation (2.9), was higher in LNCaP than in PNT2-C2 in the non-permeabilised baseline (Figure 3.14). At the mean level this corresponded to an increase of approximately 40%, while the KDE mode increased by approximately 60%, indicating that both the average and the typical nucleus were shifted upwards in the malignant model. This is

similar to the relative change in area discussed in Section 3.2.1 and shown in Figure 3.8, which may suggest that this is strongly linked to the nuclear area [416]. Because integrated DNA-dye intensity is widely used as an imaging proxy for DNA content under controlled acquisition settings, this result is consistent with the greater genomic complexity and aneuploid tendency of LNCaP as a malignant prostate cell line that is highly aneuploid (modal chromosome number ∼76 to 91, i.e. almost tetraploid) [2, 165, 361], whereas PNT2-C2 as a non-malignant prostate cell line is expected to be largely diploid. The higher total DAPI signal in LNCaP is therefore consistent with an increase in DNA content due to aneuploidy. Notably, aberrant DNA content (aneuploidy or whole-genome duplication) is a common hallmark of cancer cells, occurring in the majority of solid tumours [363], and in prostate cancer tends to correlate with disease aggressiveness and progression [362]. However, the interpretation remains conditional because DAPI intensity is influenced not only by DNA amount, but also by dye accessibility, chromatin state, fixation, and imaging parameters [8, 9].

Despite the difference in absolute DNA content, the variability in total intensity among cells was similar between the two cell lines. The CV of total nuclear DAPI intensity of both cell lines were within uncertainty of one another, indicating that the relative heterogeneity in DNA content per nucleus is comparable between both cell lines. In other words, while the distribution of DNA content in LNCaP is shifted to higher values, its spread (as a fraction of the mean) is not significantly different than that of PNT2-C2. As such, while both populations exhibit a degree of nucleus-level DNA content variability (potentially due to varying cell cycle stages or subclonal differences), the greater mean value observed for LNCaP is mainly a global shift.

To verify if the observed change in total DAPI intensity is solely a result of increasing nuclear size and also assess any differences in overall chromatin compaction, the nuclear DNA density (total DAPI intensity divided by nuclear area) between the cell lines was compared. While mean DAPI density was slightly higher in LNCaP than in PNT2-C2, these values were still within uncertainty of each other, and so no statistically clear difference in average chromatin density can be determined. Overall then, although LNCaP nuclei contain more DNA in absolute terms, they are also larger, and the net density of DNA is not dramatically changed; based on this analysis, the chromatin in LNCaP nuclei cannot be stated to be uniformly more condensed overall. The additional DNA is distributed over a larger nuclear volume such that the average concentration remains similar to that in PNT2-C2 nuclei. This is consistent with previous findings that cancer cells often maintain a balance between nuclear size and DNA content, preventing excessive bulk density even as ploidy increases [399]. The CV of DAPI density, however, did show significant differences between LNCaP and PNT2-C2 nuclei. An increase of $\sim 20\%$ was observed from PNT2-C2 to LNCaP. This may suggest that despite the nuclear DNA density being similar over both populations, LNCaP nuclei show greater variability, with more cells having very high (or low) DNA density. Asides from the expected changes in chromatin organisation in cancer cells, this may be explained as a result of the LNCaP being more likely to be found undergoing apoptosis or mitosis, as the regulation of the cell cycle is known to be impaired in cancer cells, and LNCaP cells may have been more susceptible to damage during sample preparation [163, 207, 367, 376, 377].

Additionally, the CV of intensity in each individual nucleus was evaluated, and the mean reported in Figure 3.14. An approximately 40% increase in mean CV was seen between PNT2-C2 ($(27.1 \pm 0.1)\,\%$) and LNCaP ($(38.2 \pm 0.4)\,\%$). The coefficient of variation of nuclear DAPI staining is a quantitative measure of intranuclear DNA density heterogeneity, [358] with a higher CV indicating greater variability

in local DNA concentration within each nucleus. The trends observed for CV here imply that LNCaP cells contain both intensely stained (densely packed) and weakly stained (decondensed) DNA regions. As such, the elevated CV in LNCaP suggests more irregular chromatin organisation relative to the more uniform DNA distribution expected in normal prostate cells.

DAPI intensity reflects the mix of heterochromatin and euchromatin: condensed heterochromatin (DAPI-bright) and open euchromatin (DAPI-dim) [417]. A high intranuclear CV can arise if large heterochromatic foci exist with decondensed domains. Prostate cancer cells often alter histone modifications that maintain heterochromatin; for example, trimethylation of H3K9 (a mark of constitutive heterochromatin) is linked to aggressive prostate cancer and therapy resistance [366, 418]. In experimental systems, relaxing chromatin homogenises DNA distribution, substantially lowering DAPI CV [419]. Thus, the persistence of higher heterogeneity in LNCaP implies the opposite: dysregulated epigenetic controls that preserve distinct compact and open regions. Aberrant patterns of histone methylation or acetylation in LNCaP may result in "patchy" compaction [418], yielding the larger DAPI intensity fluctuations observed. More simply, epigenetic reprogramming in cancer may increase spatial chromatin variance.

In prostate cancer models, reorganisation of A and B type lamins and the emergence of lamin B-deficient microdomains (LDMDs), often coinciding with nuclear folding, provide a direct mechanism for redistributing heterochromatin and generating low and high density chromatin regions [420]. Consistently, prostate cancer cell systems show distorted nuclear morphology with altered lamin expression [330], and metastatic prostate nuclei with low lamin levels are softer with reduced chromatin condensation [380]. These observations plausibly explain the higher mean DAPI-CV per nucleus in LNCaP vs. PNT2-C2 despite the relatively stable mean DAPI density as reflecting weakened chromatin tethering and more spatially uneven chromatin packing.

Additionally, increased prevalence of nucleoli may contribute to the increase in DAPI CV in LNCaP cells. Any increase in the spatial coexistence of low and high intensity regions will increase the variance in intensity (and typically CV) without requiring a change in the nucleus-wide mean intensity. Nucleoli are known to be DNA-poor, and are therefore able to be identified in fluorescence based analyses using negative markers i.e. regions from which DAPI signal is excluded [364]. An increase in nucleolar size and/or number would therefore be expected to introduce larger low-intensity regions within each nucleus, increasing intranuclear intensity dispersion and providing a direct route to an increase in DAPI-CV in LNCaP nuclei.

This interpretation is mechanistically consistent with prostate neoplasia, where nucleolar enlargement and increased nucleolar number are early and common nuclear alterations, and can be driven by oncogenic programmes that upregulate ribosome biogenesis. In particular, activation of the MYC oncogene has been shown to drive nucleolar structural changes in prostatic neoplasia and to maintain nucleolar programmes in prostate cancer cell models, including dependence on nucleolar proteins such as fibrillarin [17]. At the same time, clinical prostate cancer commonly exhibits increased rRNA abundance, which is also consistent with elevated nucleolar activity and altered nucleolar organisation [222]. In recent studies, prostate cancer transcription factor programmes have been shown to directly engage rDNA repeats and regulate rRNA transcription, further supporting a transcriptional mechanism for enhanced nucleolar output in prostate cancer lineages [421].

Genome instability and pleomorphism can further accentuate local variability in DNA density. Tumorous

nuclei are reported to exhibit irregular subnuclear compartmentalisation consistent with a "patchy" DAPI landscape [419]. Moreover, recent super-resolution imaging work has linked transcription-coupled genome folding/supercoiling to large-scale chromatin reconfiguration [422], supporting the interpretation that elevated CV in DAPI intensity reflects stronger partitioning into domains of differing compaction and access. It is important to note however that the coefficient of variation is not a direct readout of compaction alone; it is also sensitive to staining/imaging/segmentation, and as such should be framed as a robust indicator of cancer-associated nuclear architectural disruption mechanistically consistent with lamina changes and altered chromatin mechanics [330, 380].

Bright/dark ratio box plots

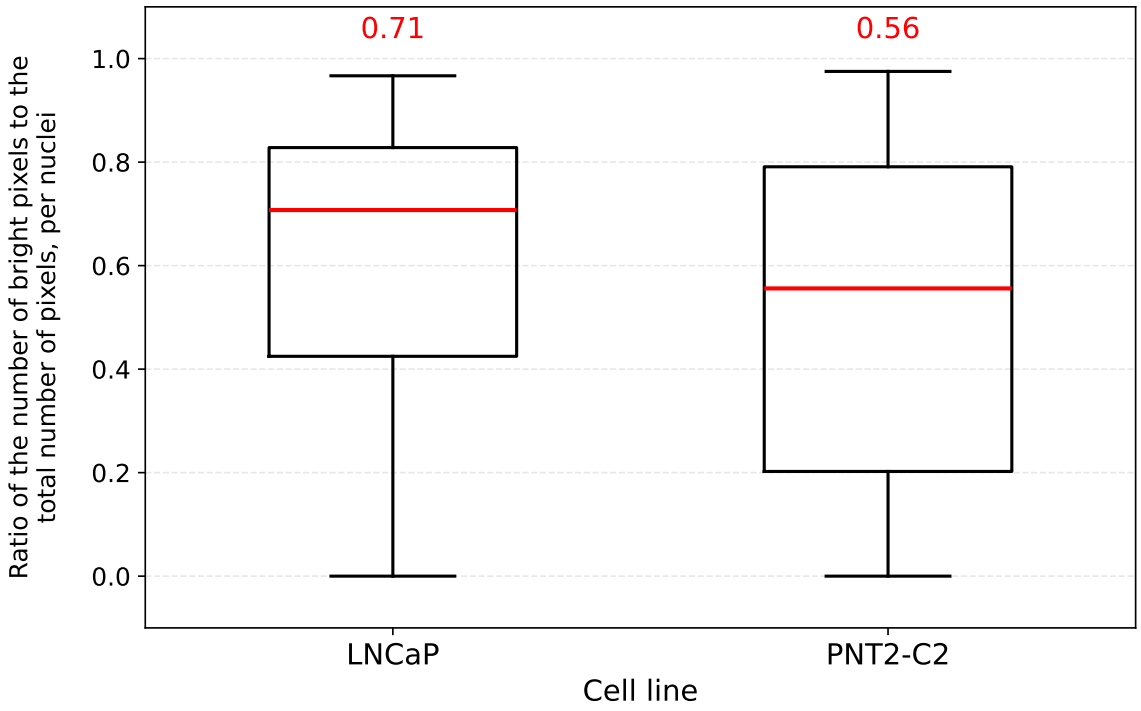
To support interpretation of the within-nucleus DAPI coefficient of variation, two threshold-based metrics were computed for each segmented nucleus and summarised as box plots: the bright-pixel fraction and dark-pixel fraction. Operationally, for a given nucleus with N_{pix} pixels and DAPI intensity I per pixel, these are defined as

$$R_{\text{bright}} = \frac{N(I > I_{\text{bright}})}{N_{\text{pix}}}, \qquad R_{\text{dark}} = \frac{N(I < I_{\text{dark}})}{N_{\text{pix}}},$$

where I_{bright} and I_{dark} are fixed intensity thresholds applied across both cell lines of 22000 and 10000 respectively, and were chosen to closely reflect the average of the second and first Otsu thresholds across all nuclei during segmentation. These descriptors do not attempt to reconstruct chromatin state directly; rather, they quantify how much of the nucleus resides in the high and low intensity tails of the DAPI distribution, and are therefore tentatively interpreted as proxies for intranuclear intensity heterogeneity[200, 419, 423].

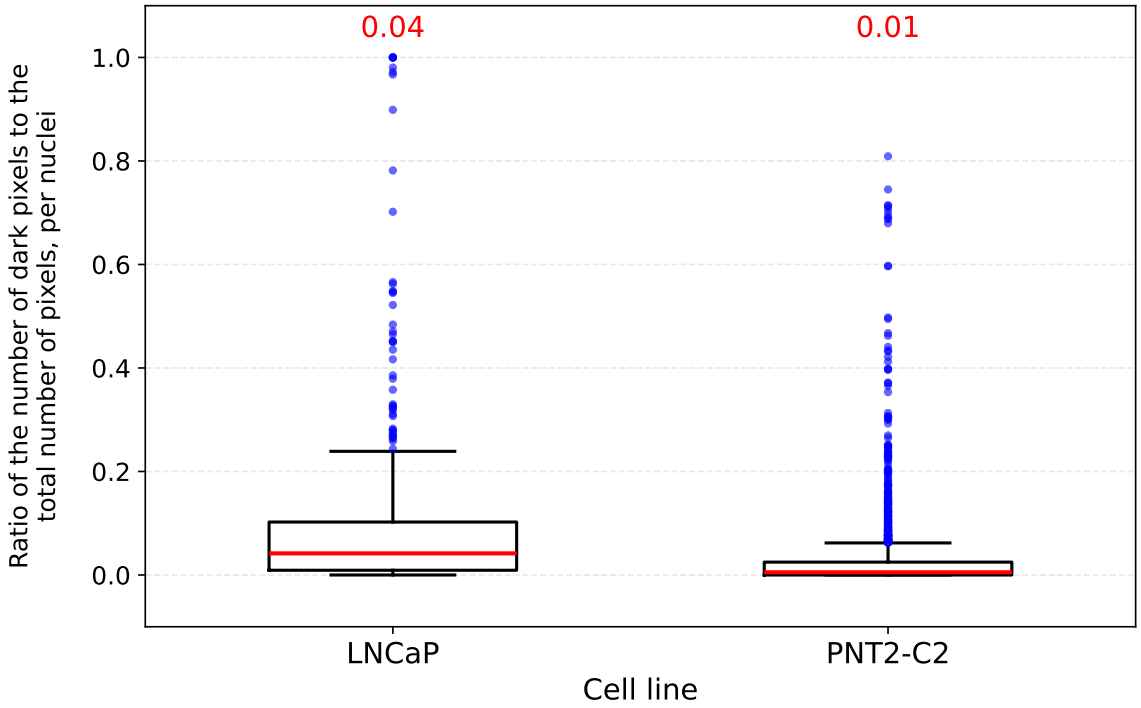
As detailed in Figure 3.15, LNCaP exhibited an increased proportion in both bright and dark tails relative to PNT2-C2. For R_{bright}, the median (IQR) was 70.7% (42.5% – 82.8%) in LNCaP vs. 55.6% (20.2% – 79.1%) in PNT2-C2. For R_{dark}, the median (IQR) was 4.2% (0.9% – 10.2%) in LNCaP vs. 0.5% (0.00% – 2.5%) in PNT2-C2. The qualitative pattern of larger R_{bright} and R_{dark} is consistent with a more heavy-tailed, spatially compartmentalised DAPI intensity landscape in LNCaP, which matches the independently observed increase in within-nucleus DAPI-CV [419]. Mechanistically, enrichment of high-intensity pixels is compatible with increased local DNA-dye signal associated with condensed chromatin subdomains and heterochromatin aggregation, noting that DAPI preferentially binds the DNA minor groove, and that DAPI can sensitively report chromatin compaction state under controlled staining/imaging conditions [191, 200, 205, 358, 423]. Consistent with this general link between DAPI texture and heterochromatin organisation, DAPI-based texture and heterochromatin aggregation changes have been explicitly quantified in prostate cancer cell models (e.g. PC3 morphology/state transitions) using 3D nuclear texture features [249].

Conversely, the increased R_{dark} in LNCaP supports a greater prevalence or area fraction of DAPI-poor intranuclear regions. This strongly suggests that a possible contributor may be the nucleolus: nucleoli are characteristically DNA-poor, and can often appear as low-intensity "holes" within the DAPI channel, enabling nucleolar segmentation strategies that explicitly exploit DAPI exclusion as mentioned previously [139, 364]. In prostate cancer, nucleolar enlargement and prominence are long-established histopathologic correlates of malignancy and risk, with nucleolar size carrying predictive value in prostate diagnostic contexts [72, 178]. At the level of prostate cancer cell biology, nucleolar hypertrophy is



LNCaP: Q1=0.42, Q3=0.83, IQR=0.40, L.Out=0, H.Out=0
PNT2-C2: Q1=0.20, Q3=0.79, IQR=0.59, L.Out=0, H.Out=0

(a) Boxplot showing the ratio of the number of bright pixels to the total number of pixels in each nucleus in non-permeabilised LNCaP (left) and PNT2-C2 (right) cells. The median is reported in red above each plot, with the quartiles, IQR, and number of outliers at each end reported underneath. Bright pixels are defined as pixels with an intensity greater than 22000 a.u.



LNCaP: Q1=0.01, Q3=0.10, IQR=0.09, L.Out=0, H.Out=45
PNT2-C2: Q1=0.00, Q3=0.02, IQR=0.02, L.Out=0, H.Out=239

(b) Boxplot showing the ratio of the number of dark pixels to the total number of pixels in each nucleus in non-permeabilised LNCaP (left) and PNT2-C2 (right) cells. The median is reported in red above each plot, with the quartiles, IQR, and number of outliers at each end reported underneath. Dark pixels are defined as pixels with an intensity lower than 10000 a.u.

Figure 3.15: Box plots showing bright and dark ratios (R_{bright} and R_{dark}) of non-permeabilised LNCaP (left) and PNT2-C2 (right) cells. Outliers are overlaid as blue dots.

mechanistically consistent with increased ribosome biogenesis demand in cancer [136] and with androgen receptor (AR)-centred programmes that elevate rRNA transcription in androgen-responsive prostate cancer cells, including LNCaP, via angiogenin-mediated regulation of rDNA transcription [424]. Kabler et al. (1996) also found that nucleolar activity is tightly coupled to androgen signalling; androgens increase rRNA synthesis in LNCaP and modulate the abundance of polymerase I associated transcription factors (such as UBF), providing a clear and explicit route by which the LNCaP-specific regulatory state could promote nucleolar hypertrophy. The joint increase of R_{bright} and R_{dark} provides a simple, threshold-based corroboration of increased intranuclear intensity heterogeneity (and thus DAPI-CV) in LNCaP, plausibly arising from co-existing condensed chromatin subdomains and expanded DAPI-poor compartments in the form of nucleoli, both of which are compatible with prostate cancer-driven nuclear changes.

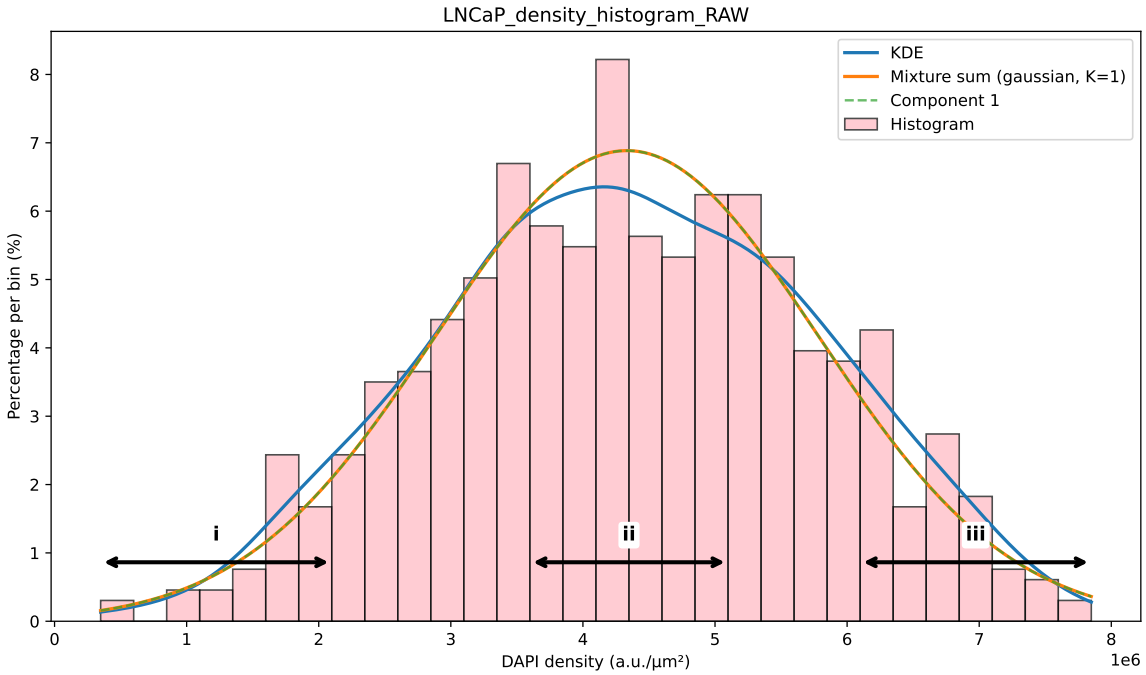
The threshold-based bright and dark ratios provide an orthogonal, tail-probability confirmation of the distribution-level findings derived from CV and histogram modelling. By quantifying the allocation of nuclear area to intensity extremes under fixed thresholds, this analysis reinforces the conclusion that LNCaP nuclei occupy a broader intensity state-space. The convergence of CV-based, KDE-based, and threshold-based descriptors strengthens the robustness of the heterogeneity inference. Nevertheless, these ratios remain strictly method-dependent; because I_{bright} and I_{dark} are fixed in absolute intensity space, R_{bright} and R_{dark} can easily be shifted by staining kinetics, imaging parameters/exposure time, background subtraction, and segmentation quality, and thus to avoid overstating the interpretations, their primary value here is as convergent evidence that LNCaP nuclei allocate more area to the intensity extremes of the DAPI signal distribution, consistent with the expected altered chromatin organisation in the disease state.

Analysis of histograms

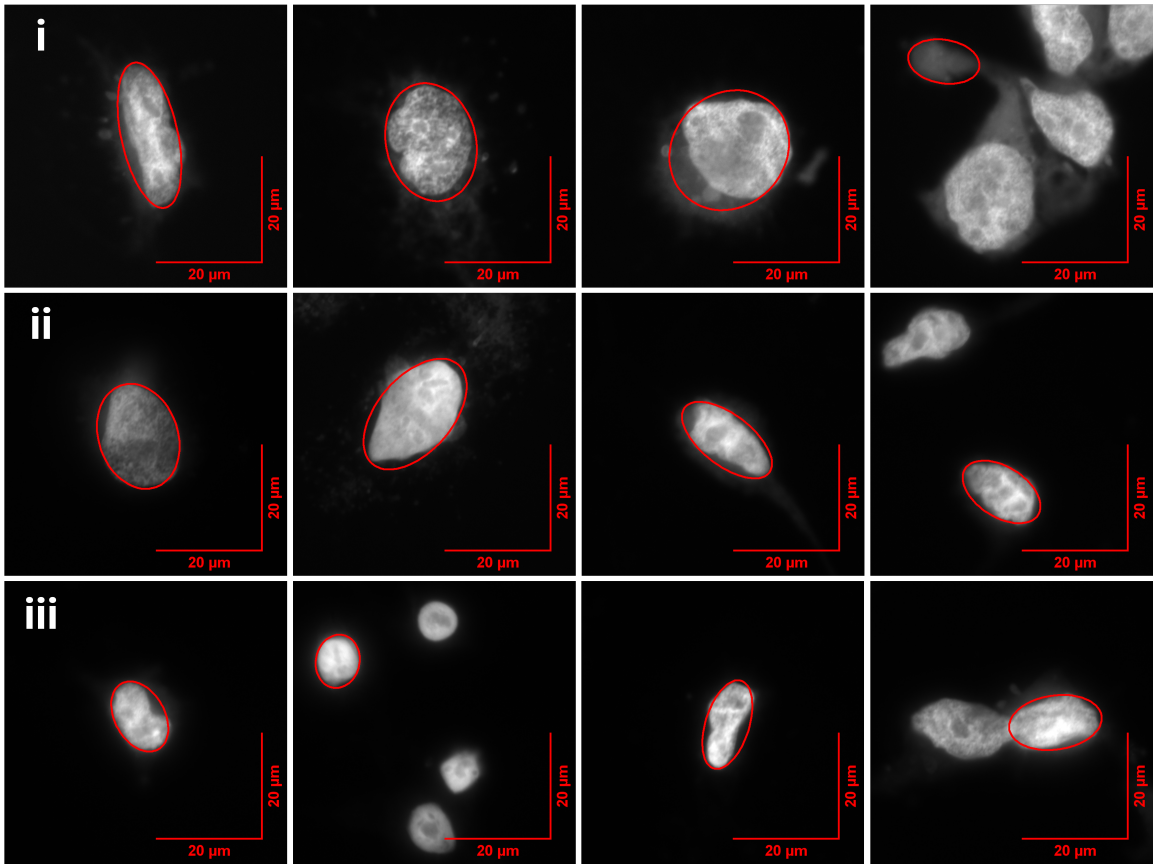
As mentioned previously (Section 3.2.1), sole reliance on summary statistics such as the mean can obscure important details [174]. As such, histograms were used to further investigate some of the quantities discussed above. Under fixed staining and imaging conditions, integrated DNA-dye fluorescence is widely used as an imaging proxy for DNA content and for cell-cycle staging at single-cell resolution [9, 193, 204]. However, total DAPI intensity is not a pure DNA-mass readout: it can also vary with nuclear size, dye-binding accessibility and base-composition preferences, fixation/permeabilisation chemistry, and optical factors [8, 9]. As such, DNA-content interpretation is plausible, but not uniquely identifiable without calibration (such as DNA standards).

The DAPI density histograms presented in Figures 3.16 and 3.17 quantify the distribution of per-nucleus DAPI signal normalised by projected nuclear area, defined in Equation (2.10). This normalisation aims to decouple intensity from the nuclear enlargement reported in Section 3.2.1, and therefore provides a more direct readout of DNA content per unit area. Parametric modelling selected a single Gaussian description for LNCaP, and a single Gamma description for PNT2-C2, indicating that the inherent shape of the DAPI density distribution differs between the disease model and non-malignant comparator.

Importantly, the histogram analysis adds new information beyond the earlier mean-based summary; although the arithmetic means were similar with overlapping uncertainties ($(4.34 \pm 0.06) \times 10^6$ a.u. µm^{-2} for LNCaP and $(4.28 \pm 0.02) \times 10^6$ a.u. µm^{-2} for PNT2-C2), the modal value (taken as the KDE peak as in

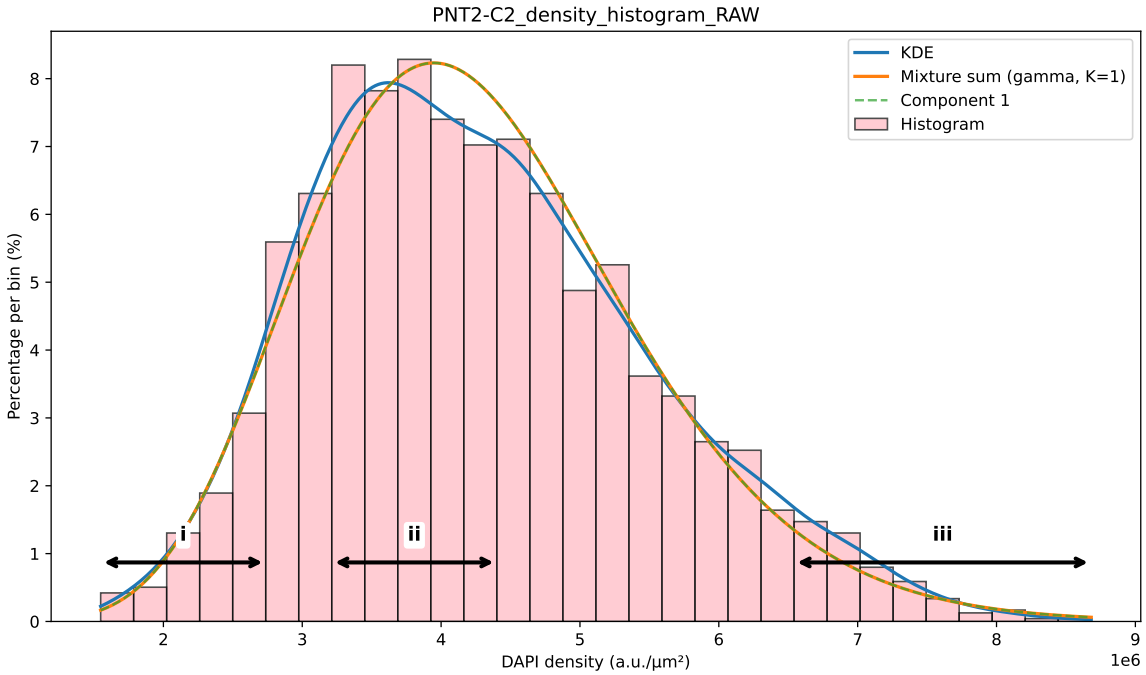


(a) Histogram showing the DAPI density in each LNCaP nucleus. A KDE is overlaid on top of the histogram, with fitting applied to the KDE to determine the overall distribution of the data [258]. The best model (chosen by ICL [259]) was a single Gaussian distribution with $\mu = (4.33 \pm 0.01) \times 10^6$ and $\sigma = (1.45 \pm 0.01) \times 10^6$, with a peak at $(4.33 \pm 0.01) \times 10^6$ a.u. µm^{-2}. The KDE peak location was $(4.2 \pm 0.2) \times 10^6$ a.u. µm^{-2}.

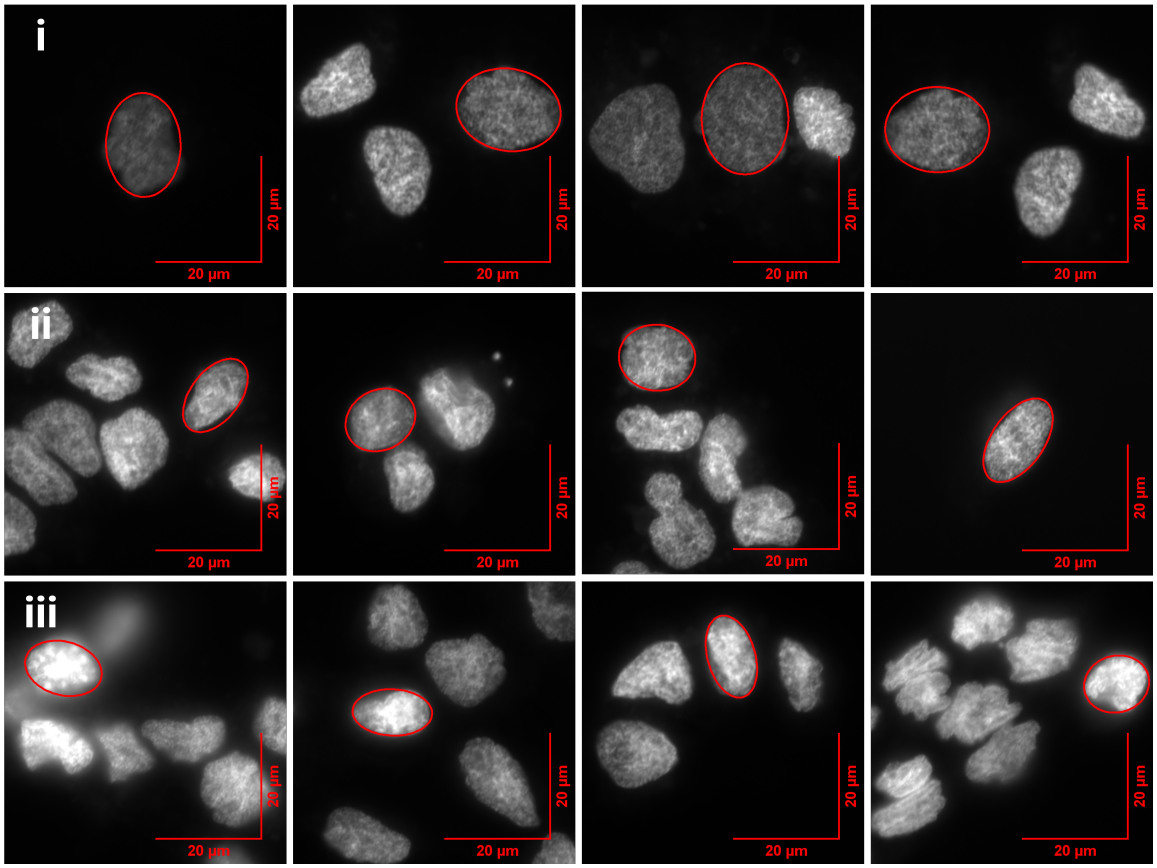


(b) Images of detected nuclei selected from the low (i), modal (ii), and high (iii) regions of the corresponding histogram. Red ellipses are drawn purely for illustrative purposes to show the approximate region which was detected as a nucleus by the segmentation code. DAPI density was calculated as the sum of the intensity of all pixels in the nuclei mask divided by the nuclear area. A 20 µm scale bar is shown in red at the bottom right corner of each image.

Figure 3.16: Histogram showing the DAPI density in non-permeabilised LNCaP cells ((a)), with some images of representative nuclei from the low, modal, and high regions of the histogram ((b)). The KDE is used as a basis for fitting, with the histogram being purely illustrative. The best fit was chosen using an Integrated Classification Likelihood (ICL) [259]. Uncertainties are calculated from the covariance matrix.



(a) Histogram showing the DAPI density in each PNT2-C2 nucleus. A KDE is overlaid on top of the histogram, with fitting applied to the KDE to determine the overall distribution of the data [258]. The best model (chosen by ICL [259]) was a single Gamma distribution with a $= 12.90 \pm 0.08$, loc $= 0$, and scale $= (3.317 \pm 0.007) \times 10^5$, with a peak at $(3.95 \pm 0.02) \times 10^6$ a.u. μm^{-2}. The KDE peak location was $(3.62 \pm 0.09) \times 10^6$ a.u. μm^{-2}.



(b) Images of detected nuclei selected from the low (i), modal (ii), and high (iii) regions of the corresponding histogram. Red ellipses are drawn purely for illustrative purposes to show the approximate region which was detected as a nucleus by the segmentation code. DAPI density was calculated as the sum of the intensity of all pixels in the nuclei mask divided by the nuclear area. A 20 µm scale bar is shown in red at the bottom right corner of each image.

Figure 3.17: Histogram showing the DAPI density in non-permeabilised PNT2-C2 cells ((a)), with some images of representative nuclei from the low, modal, and high regions of the histogram ((b)). The KDE is used as a basis for fitting, with the histogram being purely illustrative. The best fit was chosen using an Integrated Classification Likelihood (ICL) [259]. Uncertainties are calculated from the covariance matrix.

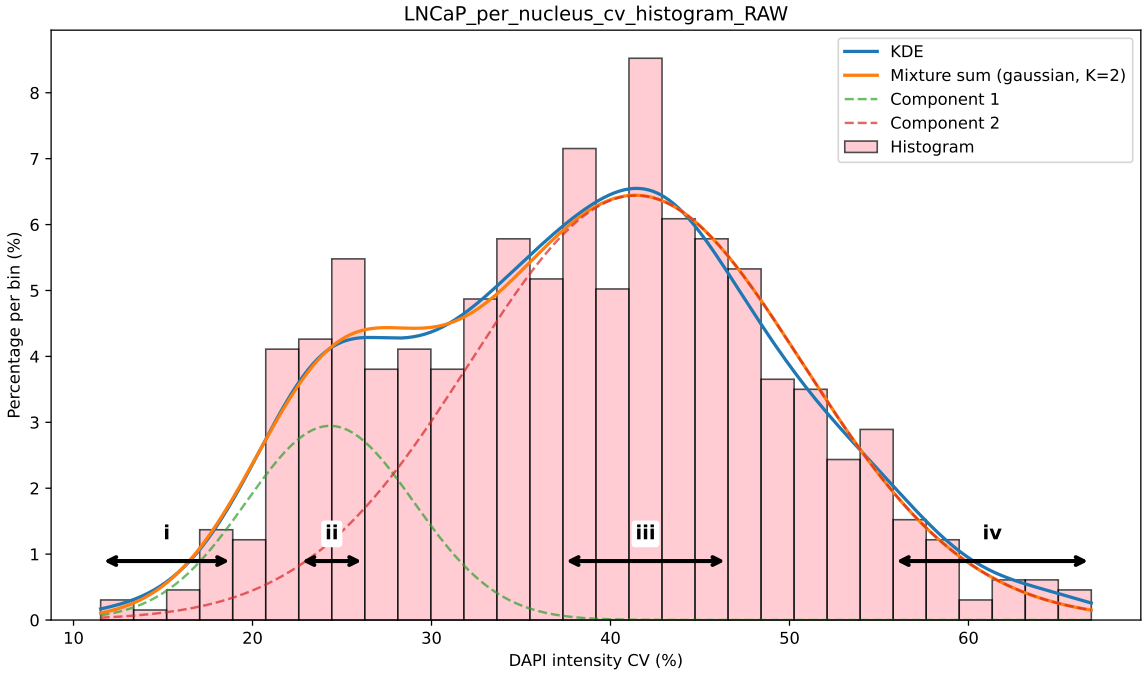
the morphology histograms) was higher notably in LNCaP than in PNT2-C2 ($(4.2 \pm 0.2) \times 10^6$ a.u. µm^{-2} vs. $(3.62 \pm 0.09) \times 10^6$ a.u. µm^{-2}), corresponding to an approximately 15% increase in the peak location from PNT2-C2 to LNCaP. The reason this peak shift is largely invisible to the mean is distributional; in PNT2-C2 the mean exceeds the mode by $\sim 20\%$ (right-skew, consistent with the Gamma fit), whereas in LNCaP the mean exceeds the mode by only $\sim 5\%$ (near-symmetric, consistent with the Gaussian fit). Additionally, the distributions are relatively "wide" in both cases, and so the standard deviation (thus SEM) is higher as a result; the combined effect of both of these factors results in the overlap of mean values for density. When comparing the percentage of data in the tails, LNCaP still shows consistently higher values compared to PNT2-C2. For example, $\sim 15\%$ of LNCaP nuclei had a density $\geq 6 \times 10^6$ a.u. µm^{-2} compared to $\sim 10\%$ of PNT2-C2 nuclei, and the 90th percentile in LNCaP fell at 6.26×10^6 a.u. µm^{-2}, while in PNT2-C2 it was 5.99×10^6 a.u. µm^{-2}. Even despite the greater positive skew in PNT2-C2, the modal shift in LNCaP is great enough that a greater proportion of the data remains in the high portion of the distribution when compared to PNT2-C2.

This work provides an explicit distribution-level result; while PNT2-C2 cells have a lower modal density, they have a greater proportion of high-density (relative to the modal density) nuclei which work to increase the mean DAPI density, and LNCaP nuclei show a higher typical density, with a greater distribution of values and less of a positive skew. This may be due to an increase in the overall variability of DAPI density increasing the presence of both low and high density nuclei, thus "evening out" the observed distribution. This distinction is not observable from reporting the mean alone, and illustrates the value of histogram-based modelling when comparing prostate cancer cell-line states.

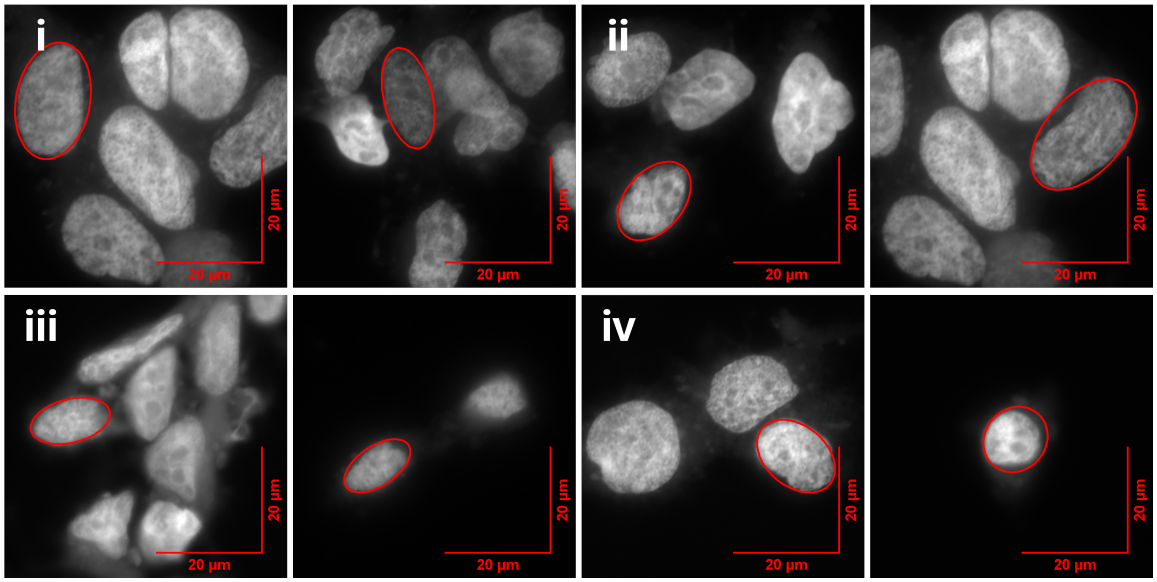
It is important to mention that although DAPI density is less directly inflated by poor segmentation than area, extreme values can still arise, especially as a result of the extranuclear staining in LNCaP nuclei. As seen in Figure 3.16(b), in some rare cases, regions of extranuclear staining were actually detected as individual nuclei. These "nuclei" show much lower relative staining intensity, which may contribute to the reduction in skew observed in the LNCaP DAPI density distribution. Additionally, the high density nuclei appear to be the result of increased chromatin condensation and DNA compaction, which could possibly be from cells undergoing mitosis or apoptosis, both of which can manifest as a condensation of DNA and shrinking of nuclei, effectively resulting in much higher detected DAPI density [191, 193]. Serum starvation was carried out to minimise the variation in cell cycles, and so the population of cells in active mitosis was small, and likely did not constitute a significant proportion of the total measured nuclei.

Biologically, DAPI density is commonly interpreted as a imaging proxy for DNA-associated signal per unit nuclear area and, more qualitatively, for chromatin packaging state (with increased local compaction producing locally brighter staining), but remains sensitive to sample preparation, image acquisition parameters, and data-processing [8]. Within these constraints, the observed increase in modal DAPI density in LNCaP is consistent with cancer-associated alterations to chromatin organisation and nuclear texture reported in prostate cancer biomarker studies, though most prior work quantifies texture/intensity features on tissue rather than modelling the full nucleus-level density distribution in cell-line systems [132, 425].

Histograms of the per-nucleus DAPI intensity coefficient of variation (CV) are shown in Figures 3.18 and 3.19. Each segmented nucleus contributed a single descriptor of intranuclear intensity heterogene-

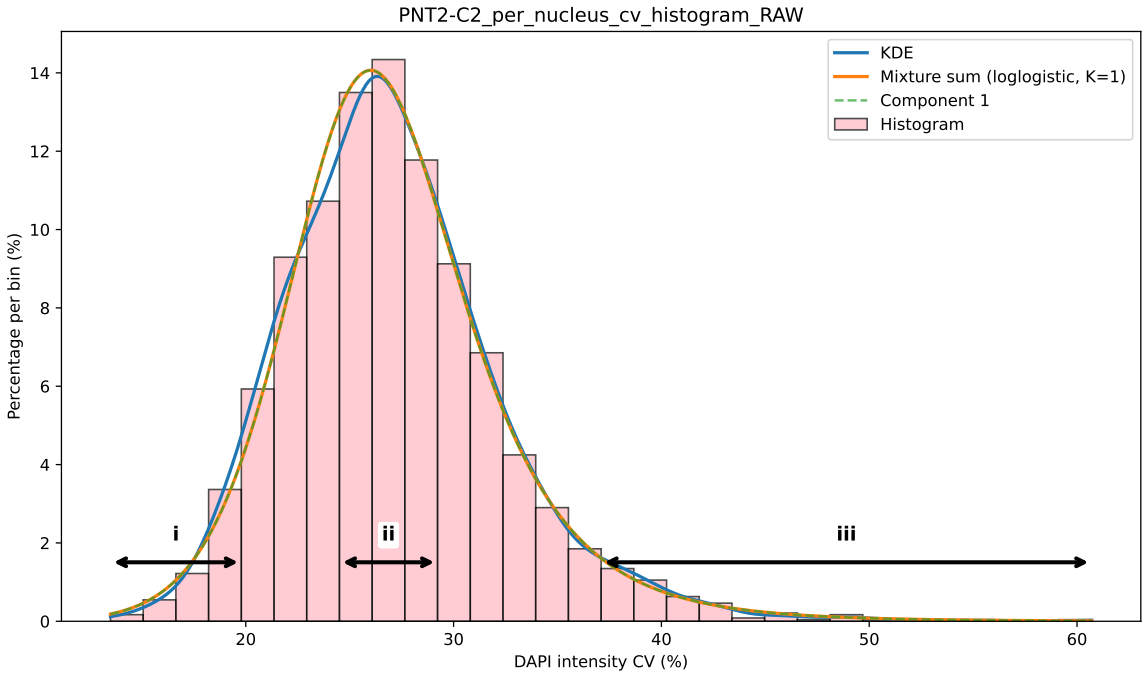


(a) Histogram showing the CV of DAPI intensity in each LNCaP nucleus. A KDE is overlaid on top of the histogram, with fitting applied to the KDE to determine the overall distribution of the data [258]. The best model (chosen by ICL [259]) was a Gaussian distribution with two components, the first had $\mu = 24.3 \pm 0.4$ and $\sigma = 4.7 \pm 0.3$, with a peak at $(24.3 \pm 0.4)\,\%$, the second had $\mu = 41.4 \pm 0.4$ and $\sigma = 9.3 \pm 0.3$, with a peak at $(41.4 \pm 0.4)\,\%$. The KDE peak location was $(41 \pm 1)\,\%$.

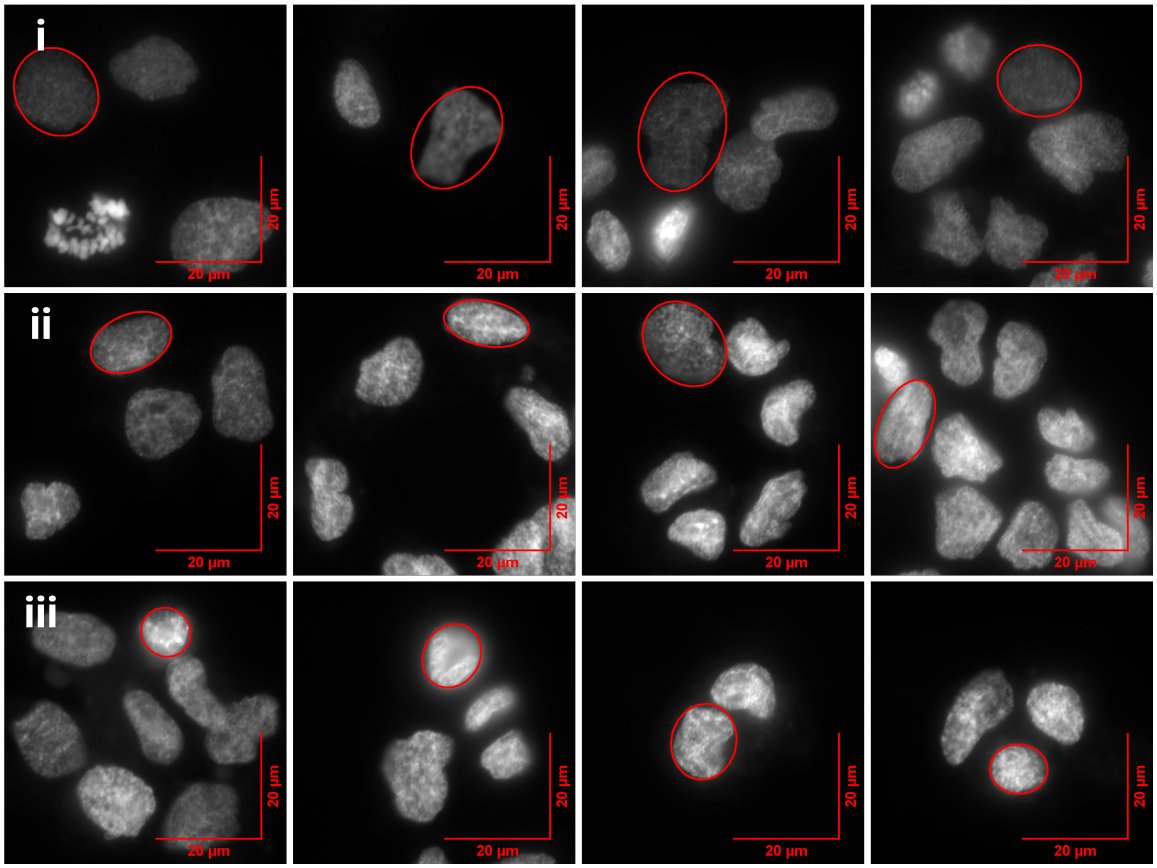


(b) Images of detected nuclei selected from the low (i), subpopulation (ii), modal (iii), and high (iv) regions of the corresponding histogram. Red ellipses are drawn purely for illustrative purposes to show the approximate region which was detected as a nucleus by the segmentation code. CV was calculated as 100 times the standard deviation of all pixel intensities in the nuclei mask divided by the mean. A 20 µm scale bar is shown in red at the bottom right corner of each image.

Figure 3.18: Histogram showing the CV of DAPI intensity in non-permeabilised LNCaP cells ((a)), with some images of representative nuclei from the low, subpopulation, modal, and high regions of the histogram ((b)). The KDE is used as a basis for fitting, with the histogram being purely illustrative. The best fit was chosen using an Integrated Classification Likelihood (ICL) [259]. Uncertainties are calculated from the covariance matrix.



(a) Histogram showing the CV of DAPI intensity in each PNT2-C2 nucleus. A KDE is overlaid on top of the histogram, with fitting applied to the KDE to determine the overall distribution of the data [258]. The best model (chosen by ICL [259]) was a single log-logistic distribution with $c = 9.4 \pm 0.2$, $\text{loc} = 0$, and $\text{scale} = 26.6 \pm 0.1$, with a peak at $(26.0 \pm 0.1)\,\%$. The KDE peak location was $(26.3 \pm 0.2)\,\%$.



(b) Images of detected nuclei selected from the low (i), modal (ii), and high (iii) regions of the corresponding histogram. Red ellipses are drawn purely for illustrative purposes to show the approximate region which was detected as a nucleus by the segmentation code. CV was calculated as 100 times the standard deviation of all pixel intensities in the nuclei mask divided by the mean. A 20 µm scale bar is shown in red at the bottom right corner of each image.

Figure 3.19: Histogram showing the CV of DAPI intensity in non-permeabilised PNT2-C2 cells ((a)), with some images of representative nuclei from the low, modal, and high regions of the histogram ((b)). The KDE is used as a basis for fitting, with the histogram being purely illustrative. The best fit was chosen using an Integrated Classification Likelihood (ICL) [259]. Uncertainties are calculated from the covariance matrix.

ity,

$$\mathrm{CV}_{\mathrm{DAPI}} = \frac{\sigma(I)}{\mu(I)} \times 100,$$

where I denotes the set of pixel intensities within the nuclear mask, and $\mu(I)$ and $\sigma(I)$ are the mean and standard deviation over those pixels. Higher $\mathrm{CV}_{\mathrm{DAPI}}$ indicates greater spatial contrast between bright and dim regions within a nucleus, whereas lower values indicate a more uniform nuclear intensity field. This type of CV-based quantification has previously been used as a practical readout of chromatin compaction/heterogeneity, although the mapping from intensity statistics to specific chromatin states is not necessarily direct [358].

A key distribution-level difference is that LNCaP was best described as a two-component Gaussian mixture, whereas PNT2-C2 was best described by a single log-logistic distribution. In LNCaP, the fitted component peaks were at $(24.3 \pm 0.4)\,\%$ and $(41.4 \pm 0.4)\,\%$, indicating the presence of a lower-CV subpopulation within the dominant higher-CV population. Importantly, in this case the KDE mode reports the location of the primary density maximum (i.e. the dominant population peak), in this case $(41 \pm 1)\,\%$ and therefore does not characterise the subpopulation peak. The subpopulation location is instead summarised by the first fitted component peak reported above and in Figure 3.18(a). In contrast, PNT2-C2 exhibited a single peak at $(26.3 \pm 0.2)\,\%$, with the log-logistic form capturing a small positive tail without any clearly separable second component.

Comparisons of mean and mode further clarify how the distributional shapes differ. The mean CV was $(38.2 \pm 0.4)\,\%$ for LNCaP, and $(27.1 \pm 0.1)\,\%$ for PNT2-C2, i.e. a relative increase of $\sim 40\%$ in the malignant model. Using the KDE peak as the primary mode descriptor for the dominant population, the modal separation was larger still ($\sim 60\%$: $(41 \pm 1)\,\%$ in LNCaP vs. $(26.3 \pm 0.2)\,\%$ in PNT2-C2), reflecting that the main LNCaP distribution had a substantially higher intranuclear heterogeneity. Notably, in PNT2-C2 the mean slightly exceeded the mode (by $\sim 3\%$ relative to the mode), consistent with a unimodal right-skewed distribution. In LNCaP, the opposite is observed (mean $\sim 8\%$ below the primary population mode), which is expected when a significant lower-CV subpopulation is present. As such, in LNCaP the mean is not representative of any single nuclear texture state, and the mixture structure constitutes materially new information beyond bulk reporting. Importantly, even the low-CV subpopulation in LNCaP is centred at a higher peak value than the PNT2-C2 mode value (though they only differ by about 3% relative to the PNT2-C2 value).

The tail behaviour here highlights that the disease-associated difference is not restricted to a small number of extreme nuclei. Approximately 45% of LNCaP nuclei had $\mathrm{CV}_{\mathrm{DAPI}} \geq 40\,\%$, compared with $\sim 2\%$ in PNT2-C2 (nearly a $\sim 25\times$ increase), and the 90th percentile was 52.4 % in LNCaP, and 33.6 % in PNT2-C2, demonstrating a broad, population-level shift in the frequency of highly heterogeneous nuclei rather than an effect driven solely by rare outliers. Qualitatively, the representative nuclei anchored to each region (Figures 3.18(b) and 3.19(b)) support this interpretation: low-CV examples can be seen to exhibit comparatively smooth (and typically lower) nuclear intensity fields, while high-CV examples show stronger intranuclear contrast with notable bright and dark spots, giving the nuclei a much more noticeable "texture", consistent with more pronounced spatial heterogeneity in DNA-associated signal.

Biologically, increased heterogeneity of DNA staining patterns is consistent with the long-standing observation that prostate cancer progression is accompanied not only by nuclear enlargement and

pleomorphism, but also by altered chromatin organisation that is measurable by quantitative texture descriptors in DNA-stained nuclei [187, 217, 426, 427]. The present work extends this literature to a controlled cell line setting by measuring the CV of DAPI staining at the single-nucleus level with high N, and explicitly modelling the full distribution rather than collapsing to a single mean/CV summary, thus demonstrating that the malignant model is best described by a mixed distribution that includes a distinct lower-heterogeneity subpopulation in addition to a dominant higher-heterogeneity population. Mechanistically, multiple routes could result in this structure without requiring a fully separable genetic subclone: for example, cell-cycle compartmentation, variable chromatin condensation/dispersion states resulting from mitosis, apoptosis, or stress responses, and epigenetically driven cellular heterogeneity are all plausible contributors to intranuclear texture variation in adherent cultures [191, 193, 207, 428]. Consistent with a physical interpretation, prostate cancer cell systems have also been reported to exhibit altered nuclear mechanical state and chromatin condensation signatures relative to non-malignant counterparts, providing a feasible link between disease state, nuclear deformability, and DNA intensity organisation [380].

Finally, several technical limitations must be considered when interpreting these results. Very large or irregular masks (such as those resulting from merged nuclei) can introduce additional heterogeneity, thereby inflating the observed CV independently of true chromatin organisation. The dominant peak shift (and the need for a two-component model in LNCaP under ICL), however, indicate that this is not driven solely by a small population of highly irregular nuclei/artefacts. The malignant model exhibits a genuine, population-level enrichment of nuclei with higher intranuclear DAPI heterogeneity, together with evidence for an additional lower-heterogeneity subpopulation; information that is invisible when considering bulk descriptors alone.

3.3 Summary

This chapter evaluated the non-permeabilised fixed-cell condition as the minimally perturbed baseline for the study as a whole. The central question here was whether the malignant LNCaP model and the non-malignant comparator PNT2-C2 differed under this baseline in nuclear central tendency, nucleus-level heterogeneity, and full distributional structure. A second question was whether the available nucleus counts were sufficient to support those comparisons robustly (Sections 3.1.3 and 3.2). Convergence analysis showed that the principal descriptors followed the expected $\mathrm{SEM}(N) \propto N^{-1/2}$ behaviour. LNCaP required larger effective sample sizes than PNT2-C2 to satisfy the $\%\mathrm{SEM} < 2.5\%$ convergence criterion for all three selected convergence probes: $N_{\mathrm{conv}} = 365$ vs. 225 nuclei for area ($\sim 60\%$ higher), 170 vs. 70 for aspect ratio ($\sim 140\%$ higher), and 190 vs. 145 for DAPI density ($\sim 30\%$ higher) (Table 3.1). The fitted convergence-rate proxy γ (Equation (2.38)) was likewise lower in LNCaP across all three metrics, by approximately 50% for area and aspect ratio and $\sim 15\%$ for DAPI density, consistent with greater nucleus-level heterogeneity under matched preparation (Table 3.1). As such, this chapter established both sample-size sufficiency and a quantified heterogeneity contrast before biological interpretation was attempted.

Taken together, the non-permeabilised baseline showed that LNCaP nuclei were larger, more elongated, more heterogeneous, and more texturally variable than PNT2-C2 nuclei, with the strongest contrast observed for intranuclear DAPI texture (Figures 3.8 and 3.14 and Table 3.1).

Morphologically, LNCaP nuclei were consistently larger and more elongated than PNT2-C2 nuclei

(Figure 3.8 and Section 3.2.1). Mean nuclear area increased from $(136 \pm 1)\,\mu m^2$ in PNT2-C2 to $(195 \pm 3)\,\mu m^2$ in LNCaP (approximately 45%), while mean aspect ratio increased from 1.443 ± 0.006 to 1.72 ± 0.02 (approximately 20%; Figure 3.8). KDE modes showed the same directional changes, but with smaller shifts: modal area increased by $\sim 30\%$, and modal aspect ratio by $\sim 10\%$, indicating that the dominant population also differed in the malignant model, rather than the mean being displaced only by a small number of extreme nuclei (Figure 3.8 and Section 3.2.1).

Relative dispersion was also increased, with the area CV rising from $(34.5 \pm 0.6)\,\%$ in PNT2-C2 to $(46 \pm 2)\,\%$ in LNCaP ($\sim 30\%$ relative increase) and the aspect ratio CV from $(18.7 \pm 0.3)\,\%$ to $(30.3 \pm 0.9)\,\%$ ($\sim 60\%$ relative increase), consistent with greater pleomorphism in the malignant model, and supported by the greater heterogeneity in LNCaP area and aspect ratio assessed by convergence testing (Figure 3.8 and Table 3.1) [131, 132, 360].

The histogram-based analyses showed that these morphology contrasts were not reducible to mean shifts alone. For both area and aspect ratio, conservative ICL-based model selection (Section 2.3.5) favoured a single-component description in both cell lines, while the LNCaP distributions remained broader and more positively tailed (Section 3.2.1). For area, the mean-mode separation was approximately 30% in LNCaP, compared with $\sim 15\%$ in PNT2-C2 (Figures 3.9 and 3.10). Considering tail measures from the same figures explicitly, around 10% of LNCaP nuclei exceeded $300\,\mu m^2$ vs. only $\sim 1\%$ of PNT2-C2 nuclei, and the 90th percentile was $\sim 315\,\mu m^2$ in LNCaP vs. $\sim 195\,\mu m^2$ in PNT2-C2 (Figures 3.9 and 3.10). For aspect ratio, approximately 25% of LNCaP nuclei had aspect ratio ≥ 2 vs. only $\sim 5\%$ in PNT2-C2, and the 90th percentile increased from ~ 1.82 to ~ 2.47 (Figures 3.12 and 3.13).

A matched-N control for PNT2-C2 nuclear area, where distributional analysis was repeated using a random subset of PNT2-C2 nuclei up to $N_{\max}$ for LNCaP, such that both cell lines were assessed using a common number of nuclei, retained the same unimodal and right-skewed structure as when all available nuclei were used Figure 3.11. This suggests that the principal morphological differences observed were not simply artefacts of unequal sample size, and taken together, these results provide a distribution-level characterisation of malignant nuclear atypia in a controlled prostate cell line system, extending the largely tissue-dominated nuclear morphometry literature into an explicitly converged, nucleus-resolved fluorescence setting [132, 395].

DAPI-derived intensity analysis added two further key findings (Figure 3.14 and Section 3.2.2). First, total nuclear DAPI intensity was higher in LNCaP by approximately 40% at the mean level, and $\sim 60\%$ at the KDE mode level, consistent with the known genomic complexity and aneuploid tendency of this model (Figure 3.14) [2, 361]. Conversely, the CV of total nuclear DAPI intensity across nuclei remained within uncertainty of that of PNT2-C2, indicating that the absolute DNA-associated signal was shifted upwards without a strong increase in relative dispersion (Figure 3.14).

Second, once nuclear area was accounted for, the DAPI-density comparison became more nuanced. Mean DAPI density between cell lines overlapped within uncertainty, but the KDE mode remained approximately 15% higher in LNCaP, and the fitted distributional forms differed between cell lines (Gaussian in LNCaP vs. Gamma in PNT2-C2), indicating that both the typical density and the overall shape of the density distribution differed, even when mean separation was limited (Figures 3.14, 3.16 and 3.17). This is an important result: overlap in the did not imply distributional equivalence. Rather, PNT2-C2 retained a more positively skewed density distribution, whereas LNCaP showed a broader distribution with a higher typical density state, consistent with altered chromatin organisation under

fixed staining and imaging conditions [8, 9, 358].

The clearest cell line separation was observed for intranuclear texture. Mean CV_{DAPI} increased from $(27.1 \pm 0.1)\,\%$ in PNT2-C2 to $(38.2 \pm 0.4)\,\%$ in LNCaP ($\sim 40\%$ relative increase), while the KDE mode increased from $(26.3 \pm 0.2)\,\%$ to $(41 \pm 1)\,\%$, ($\sim 60\%$ relative increase; Figures 3.14, 3.18 and 3.19). Crucially, the PNT2-C2 CV_{DAPI} distribution was adequately unimodal, whereas LNCaP was best described by a two-component Gaussian mixture, indicating that the malignant model differed not only in the magnitude of intranuclear heterogeneity, but also in distributional complexity, with evidence for coexisting low and high CV_{DAPI} components (Figures 3.18 and 3.19).

This was not a rare-outlier effect: approximately 45% of LNCaP nuclei had $CV_{DAPI} \geq 40\%$, compared with only $\sim 2\%$ of PNT2-C2 nuclei, and the 90th percentile was $\sim 52.4\%$ vs. $\sim 33.6\%$ respectively (Figures 3.18 and 3.19). Threshold-based tail descriptors supported the same interpretation: the median bright-pixel fraction increased from 55.6% in PNT2-C2 to 70.7% in LNCaP, while the median dark-pixel fraction increased from 0.5% to 4.2%, indicating that LNCaP nuclei occupied larger fractions of both high and low intensity extremes (Figure 3.15). In combination, these results indicate that the malignant model exhibited a substantially broader range of intranuclear DAPI intensity states, consistent with literature linking prostate cancer progression to altered chromatin texture and nuclear architectural irregularity, while noting that much of the published evidence is tissue-based, and therefore not directly equivalent to the present single-nucleus cell line analysis [72, 132, 187, 426].

However, this non-permeabilised baseline does not by itself establish mechanistic routes to the observed changes. Heavy tails and apparent secondary structure may arise from a combination of effects, including segmentation edge cases (as shown in Figure 2.7), and DAPI staining alone cannot identify whether specific subnuclear compartments are responsible for the observed texture contrast. As such, this chapter should be read as a quantitatively resolved baseline rather than a complete mechanistic account. Its principal value for Chapter 4 is that it defines the baseline nuclear state prior to detergent treatment, and thus frames the next question directly: whether these robust DAPI-derived differences remain interpretable following permeabilisation, and whether they are consistent with directly measured nucleolar reorganisation using fibrillarin-based descriptors [17, 72, 136, 424].

Chapter 4

Results: Analysis of permeabilised cells

This chapter extends the non-permeabilised baseline (Chapter 3) to a permeabilised workflow in order to quantify how detergent permeabilisation itself perturbs DAPI-derived nuclear descriptors, and enable subsequent direct, protein-specific quantification of nucleolar content using immunofluorescence. Throughout, "permeabilisation" refers to detergent-mediated disruption of membrane lipid organisation by exposure to Triton X-100 to permit antibody access to intranuclear targets (namely, the nucleolus), a step which while not required for DAPI staining of the nucleus, is required for nucleolar immunolabelling [11, 193, 202]. This distinction is of critical importance: detergent exposure is not necessarily a neutral sample preparation step, even after fixation, and may result in differential extraction/redistribution of cellular components and altered retention of macromolecules [16, 295, 296]. As such, the permeabilised dataset is treated as a controlled quantification of the perturbation caused by permeabilisation, rather than as a direct continuation of the non-permeabilised condition.

The chapter is further split to separate preparation-induced effects in the DAPI channel from marker-associated effects and new metrics following the addition of anti-fibrillarin staining, while keeping the biological question in view. Nuclear size and shape remain biologically relevant descriptors: in prostate malignancy, nuclear atypia/pleomorphism (altered nuclear size, elongation, and chromatin organisation) are long-standing diagnostic features, and are mechanistically linked to genome instability and altered nuclear architecture [131, 132, 184]. The interpretation of fluorescence-derived nuclear morphometry, however, is only meaningful if sample preparation effects are either minimised or explicitly quantified and controlled for. Accordingly, a DAPI-only permeabilised condition is analysed first (Section 4.1) to provide a direct estimate of detergent-induced shifts in nuclear morphology and DAPI intensity structure, against the non-permeabilised baseline. The full DAPI and anti-fibrillarin dataset (Section 4.2) is then analysed, in which the principal objective is to quantify the nucleolar state directly and comparatively between the malignant LNCaP model and the non-malignant PNT2-C2 comparator.

Briefly, nuclei were segmented using DAPI staining, and a set of size, shape, and intensity descriptors were extracted from each nucleus, consistent with the non-permeabilised analysis. Specifically, projected nuclear area (2D area of the nuclear mask), aspect ratio of the best-fitting ellipse, total DAPI intensity per nucleus I_{DAPI}, and DAPI density Equation (2.10) were considered. Then, for nucleolar analysis, fibrillarin was used as the nucleolar marker, as it is enriched within the dense fibrillar component

of nucleoli, and has been used routinely used to identify nucleoli in fluorescence microscopy [17, 219, 417, 429]. In this work, nucleoli were defined as discrete intranuclear regions, positive for anti-fibrillarin immunofluorescence within the DAPI-defined nuclear mask, enabling both morphological and intensity-based nucleolar measures to be extracted. Nucleolar morphology was summarised by absolute nucleolar area and the nucleolar:nuclear area ratio defined in Equations (2.20) and (2.22) (Figure 4.18). Complementary intensity metrics were then introduced, with nucleolar intensity I_{nucleoli} (background-suppressed fibrillarin signal) and the nucleoli:DAPI ratio $R_{N:D}$ defined in Equations (2.21) and (2.24) providing per-nucleus measures of absolute nucleolar signal magnitude and nucleolar enrichment relative to nuclear DNA-associated fluorescence (Figure 4.31). This framing aligns with the broader cancer cell literature, in which nucleolar activation and ribosomal biogenesis are implicated in malignant growth programmes, while also retaining a microscopy-grounded, single-cell quantitative definition [63, 136, 212, 214].

As in Chapter 3, interpretation was conditioned on first demonstrating statistical sufficiency of the available nucleus counts for each metric, since permeabilisation can alter both effective heterogeneity and measurement noise [16]. Convergence testing therefore remained an explicit first step in both the permeabilised DAPI-only and DAPI+FIB analyses, yielding metric-specific convergence points N_{conv} under a fixed criterion (%SEM $< 2.5\%$). This is method-aware, and determines which comparisons are supported by stable estimation.

To summarise, the permeabilised results show that detergent exposure can shift nuclear descriptors by amounts comparable to (and in some instances larger than) the baseline LNCaP-PNT2-C2 differences, and therefore must be treated as an experimental factor, rather than as a simple neutral preparatory step. In the DAPI-only permeabilised condition, bulk morphology indicated a clear reversal in nuclear area ordering relative to the non-permeabilised baseline; PNT2-C2 nuclei were larger than LNCaP, both in mean and mode (mean $(211 \pm 2)\,\mu\text{m}^2$ vs. $(179 \pm 2)\,\mu\text{m}^2$, mode $(185 \pm 6)\,\mu\text{m}^2$ vs. $(152 \pm 3)\,\mu\text{m}^2$), as seen in Figure 4.2, while the ordering for nuclear shape descriptors remained intact (with LNCaP displaying a higher aspect ratio and aspect ratio CV, reflecting shape dispersion). Intensity analysis further indicated that permeabilisation alters DAPI intensity structure in a cell line dependent manner (evidenced through LNCaP mean I_{DAPI} decreasing by $\sim 25\%$, with comparatively little change in PNT2-C2 mean I_{DAPI}, but a substantial reduction in mean DAPI density, likely driven by the increase in nuclear area for PNT2-C2; Figure 4.7), reinforcing that DAPI derived intensity measures should not be assumed to be preparation-invariant [295, 296]. Under the full DAPI+FIB workflow, nucleolar morphology showed a modest elevation of nucleolar occupancy in LNCaP (mean $R_{N:NA}$ 0.093 ± 0.001 vs. 0.0869 ± 0.0009), while absolute nucleolar area remained essentially unchanged within uncertainty between cell lines Figure 4.18. In contrast, nucleolar intensity measures showed a substantially elevated nucleolar signal state in LNCaP (mean I_{nucleoli} of $(6.4 \pm 0.2) \times 10^7$ a.u. vs. $(5.07 \pm 0.08) \times 10^7$ a.u., and mean $R_{N:D}$ of 0.087 ± 0.002 vs. 0.0644 ± 0.0009; Figure 4.31), consistent with the well-established association between prostate malignancy and nucleolar activation [17, 63, 140, 222]. Importantly, the analysis was structured such to distinguish a global increase in nuclear fibrillarin signal from a compartment-specific redistribution towards nucleolar assigned pixels, since the latter constitutes the more specific interpretation when background nuclear staining is non-negligible (Figure 4.31).

Taken together, the chapter provides a baseline for DAPI-derived nuclear morphometry under permeabilised conditions, and a protein specific, distribution-level quantification of nucleolar state using fibrillarin, expressed in terms of both morphological ($R_{N:NA}$) and intensity related measures

(I_{nucleoli}, $R_{N:D}$). This was intended to support the mechanistic interpretations from DAPI+FIB data by clarifying which contrasts between LNCaP and PNT2-C2 persist across preparation states, and which invert (and therefore plausibly reflect preparation sensitivity). The final nucleoli-specific analysis was then carried out using a specific nucleolar protein marker rather than inferred indirectly from DAPI texture alone, enabling full characterisation of the malignant and non-malignant models of LNCaP and PNT2-C2 [136, 139, 430].

4.1 DAPI-only analysis

Permeabilised-cell results are presented as a controlled, stepwise perturbation series: the non-permeabilised DAPI baseline (Chapter 3), chemical permeabilisation with DAPI staining only (this section), and chemical permeabilisation with combined DAPI and anti-fibrillarin staining (Section 4.2). The purpose of the intermediate DAPI-only condition is to isolate preparation-induced effects (i.e. the imact of permeabilisation) from marker-induced effects when interpreting the subsequent nucleolar labelling data.

Here, "permeabilisation" denotes disruption of membrane lipid organisation using a detergent (Triton X-100) to allow access to intranuclear antibody targets. This step is required for antibody-based immunolabelling, but is not typically required for DAPI nuclear counterstaining [193, 202]. As such, repeating the Chapter 3 analysis under permeabilised, DAPI-only conditions provides a direct estimate of how the permeabilisation step itself shifts DAPI-derived morphometric and intensity statistics. Mechanistically, detergent exposure can induce partial solubilisation of membrane lipids and related loss or redistribution of cellular material, which may propagate into measured nuclear intensity/density distributions, even when imaging parameters remain unchanged [12, 16]. In prostate cancer cell-line studies, permeabilisation workflows have previously been used to control antibody access and distinguish extracellular from intracellular epitopes, including in LNCaP-derived contexts [431, 432]. However, in prostate cell line studies, it is uncommon to find an explicit, like-for-like comparison in which the same metrics, measured using the same acquisition/analysis pipeline, across the same cell lines are evaluated before and after detergent permeabilisation, with permeabilisation treated as the only experimental change. As such, the present DAPI-only permeabilised dataset is reported as a necessary control condition that quantifies the shift in each DAPI-based distribution, solely as a result of sample preparation, thereby enabling the subsequent dual-stain (DAPI and anti-fibrillarin) results to be interpreted as the incremental effect of additional labelling beyond the permeabilisation step itself, rather than as a redundant repetition of Chapter 3.

4.1.1 Convergence testing

As in the non-permeabilised baseline (Section 3.1), all subsequent interpretation is conditioned on first demonstrating that the available sample size is sufficient for stable estimation of the reported summary statistics and fitted distribution parameters. Progressive subsampling was used to quantify the decay of SEM as a function of nucleus count N, with repeated random subsampling used to minimise resampling error [334–336]. SEM decay was then modelled using the framework defined in Equation (2.37), and N_{conv} was defined using the same percentage standard error criterion as in the non-permeabilised analysis Equation (2.39) [301].

The analysis is repeated here because detergent permeabilisation can introduce preparation-dependent

variability (e.g. heterogeneous lipid solubilisation and differential loss/redistribution of cellular material), which may alter both the effective heterogeneity of measured nuclear descriptors and the measurement noise floor [12, 16, 295]. As such, the permeabilised DAPI-only convergence behaviour is reported for the same metrics as previously in each cell line to support a direct interpretation of any differences observed between non-permeabilised and permeabilised conditions. The full convergence pipeline detailed in Section 2.3.6 and first applied in Section 3.1 was applied here, however for brevity, only key plots and summary data are shown, with additional data provided in Section B.2.

Key quantitative outcomes

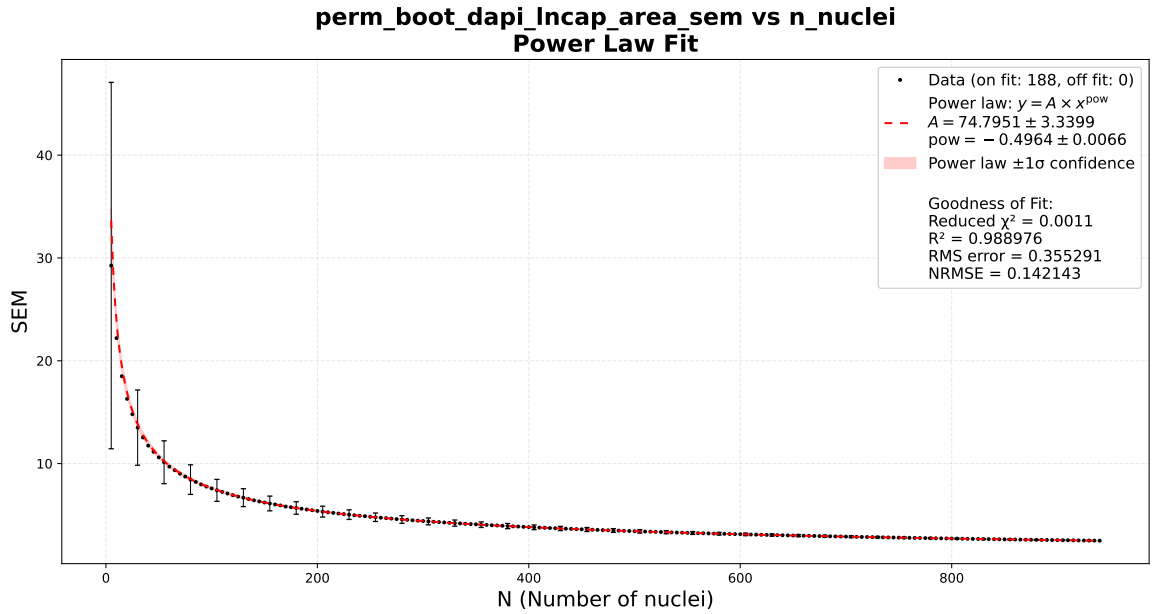
Across all three metrics (nuclear area, nuclear aspect ratio, and DAPI density), the fitted exponents remained close to $pow = -0.5$ (Table 4.1), supporting the continued application of the $\mathrm{SEM}(N) \propto N^{-1/2}$ framework for the permeabilised dataset [346, 347, 350]. Absolute mean variation from the ideal SEM scaling was larger than in the non-permeabilised set, however the largest absolute deviation from -0.5 occurred for LNCaP aspect ratio ($pow = -0.494 \pm 0.008$), corresponding to a $\sim 1.1\%$ shift in the point estimate; however, this remains consistent with -0.5 within uncertainty, and therefore is not indicative of any anomalous SEM scaling.

Under the $\%\mathrm{SEM} < 2.5\%$ convergence threshold, the required nucleus counts were $N_{\mathrm{conv}} = 340$ (area), 195 (aspect ratio), and 140 (DAPI density) for LNCaP, and $N_{\mathrm{conv}} = 310$, 100, and 300 for PNT2-C2 respectively. All convergence points remain comfortably below the available nucleus counts, confirming that the permeabilised DAPI-only dataset is statistically sufficient for stable descriptor estimation under the specified criterion. Expressed as a sampling burden ($N_{\mathrm{conv}}/N_{\mathrm{max}}$), convergence required approximately 15 % to 35 % of available nuclei in LNCaP, and 7 % to 20 % in PNT2-C2. Relative to the non-permeabilised baseline (Table 3.1), this represents a net reduction in sampling burden for LNCaP (previously 25 % to 55 %), but a net increase for PNT2-C2 (previously 3 % to 10 %), indicating that permeabilisation alters the relative statistical "difficulty" of achieving a fixed precision target in a cell-line-dependent manner. These convergence results provide a quantitative baseline under a fixed analysis pipeline and precision target. Post-fixation permeabilisation measurably changes the sampling burden required for stable DAPI-derived descriptor estimation in a cell-line-dependent manner.

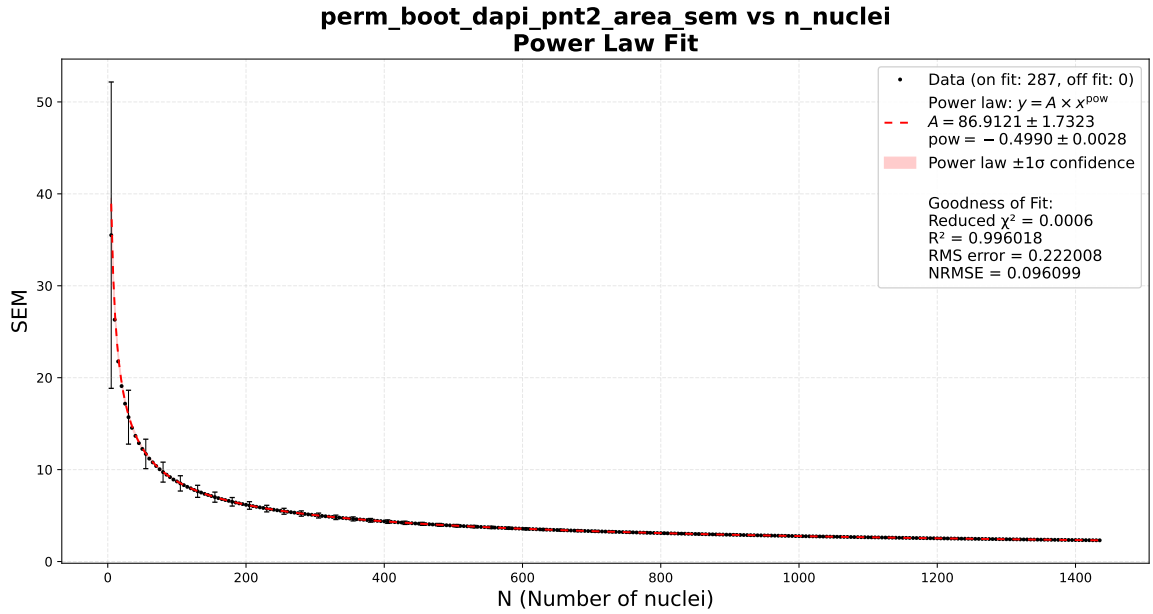
Between-cell-line differences post-permeabilisation

Because the exponent remained near $-1/2$, differences in convergence behaviour are dominated by amplitude differences rather than departures from the expected SEM scaling. Operationally, for $pow \approx -1/2$, $\mathrm{SEM}(N) \approx \sigma/\sqrt{N}$, where σ is the nucleus-level dispersion. In this regime, the fitted amplitude A (more simply interpretable as $\gamma = \frac{1}{A}$) provides a direct, metric-specific proxy for absolute SEM scale, and therefore heterogeneity [346, 347].

Post-permeabilisation, the area convergence curves (Figure 4.1) show that the separation between absolute SEM in LNCaP and PNT2-C2 is markedly reduced compared to the non-permeabilised case (Figure 3.2). Quantitatively, N_{conv} differs by only $\sim 10\%$ (340 vs 310 nuclei) in the permeabilised state, while the non-permeabilised baseline differed by $\sim 60\%$ (365 vs 225 nuclei). Furthermore, the convergence rates suggest a modest reversal in the direction of absolute SEM amplitude for nuclear area: γ is higher in LNCaP than in PNT2-C2, implying that at matched N, the absolute SEM of nuclear area is lower in LNCaP even though the relative-convergence threshold N_{conv} remains slightly higher. Despite the apparent disagreement of these measures, this is not contradictory: N_{conv} is defined



(a) The average SEM of nuclear area in permeabilised LNCaP cells stained with only DAPI, calculated from 10000 repeated random subsamples plotted against N. A power law fit is applied, with fit parameters of $A = 75 \pm 3$, and $pow = -0.496 \pm 0.007$. Goodness of fit metrics were NRMSE $= 0.142$, $\chi^2_\nu = 0.001\,15$, and $R^2 = 0.989$.



(b) The average SEM of nuclear area in permeabilised PNT2-C2 cells stained with only DAPI, calculated from 10000 repeated random subsamples plotted against N. A power law fit is applied, with fit parameters of $A = 87 \pm 2$, and $pow = -0.499 \pm 0.003$. Goodness of fit metrics were NRMSE $= 0.0961$, $\chi^2_\nu = 5.90 \times 10^{-4}$, and $R^2 = 0.996$.

Figure 4.1: Graphs showing the standard error of the mean (SEM) convergence of LNCaP (a) and PNT2-C2 (b) nuclear area post-permeabilisation and staining with only DAPI as a function of N, the number of nuclei included in the calculation. Error bars are drawn every five points, and indicate the resampling standard deviation of each datapoint. NRMSE refers to normalised root mean square error (normalised to the terminal SEM), χ^2_ν refers to reduced chi squared, and R^2 refers to the coefficient of determination. Fits use weighted least squares regression with inverse variance weighting applied, and uncertainties were calculated using covariance aware error propagation (Equation (2.40)).

Cell line	Metric	γ	*pow*	NRMSE	χ^2_ν	N_{conv}
LNCaP ($N_{\max} = 944$)	Area	0.0134 ± 0.0006	-0.496 ± 0.007	0.142	0.001 15	340
	Aspect ratio	2.0 ± 0.1	-0.494 ± 0.008	0.127	0.0011	195
	DAPI density	$(1.05 \pm 0.02) \times 10^{-6}$	-0.499 ± 0.003	0.0586	4.97×10^{-4}	140
PNT2-C2 ($N_{\max} = 1438$)	Area	0.0115 ± 0.0002	-0.499 ± 0.003	0.0961	5.90×10^{-4}	310
	Aspect ratio	3.3 ± 0.1	-0.497 ± 0.005	0.138	9.46×10^{-4}	100
	DAPI density	$(9.1 \pm 0.2) \times 10^{-7}$	-0.499 ± 0.003	0.0898	5.49×10^{-4}	300

Table 4.1: Summary of convergence analysis for permeabilised LNCaP and PNT2-C2 cells stained with only DAPI. Power law fit parameters and goodness of fit statistics refer to weighted fits performed to the mean SEM calculated through repeated random subsampling as a function of N, where γ is the convergence rate proxy defined in Equation (2.38), and *pow* is the fitted exponent from Equation (2.37). The convergence point, N_{conv}, is defined using the percentage standard error criterion from Equation (2.39), and is reported as a discrete value without uncertainty. NRMSE refers to normalised root mean square error, normalised to terminal SEM, and χ^2_ν refers to reduced chi squared. $N_{\max}$ is the total number of nuclei available in each cell line.

in terms of %SEM $= {}^{\text{SEM}}/_{\bar{x}}$, and therefore depends on both dispersion and the mean descriptor value, while γ reflects the absolute SEM scale. As such, the convergence results for nuclear area indicate that permeabilisation acts to reduce the strong cell-line heterogeneity contrast which is observed sans-permeabilisation, and that after disrupting the nuclear membrane (by detergent treatment), nuclear area variability is no longer accurately described by a simple cancer > non-malignant ordering under this preparation.

For aspect ratio, in contrast, the expected ordering is preserved: LNCaP remains less rapidly convergent than PNT2-C2 ($N_{\text{conv}} = 195$ vs. 100; $\gamma = 2.0 \pm 0.1$ vs. 3.3 ± 0.1), indicating that nuclear-shape variability remains elevated in the malignant model even after detergent treatment. The degree of separation between cell lines, however, is reduced relative to the non-permeabilised case (N_{conv} of 170 vs. 70 nuclei, with convergence rates of 1.94 ± 0.08 and 3.72 ± 0.04). This supports the interpretation that while permeabilisation acts to homogenise the variability in nuclear area between cell lines, nuclear shape is much more resistant to these changes, and the relative difference in the heterogeneity in this metric is mostly preserved after permeabilisation.

The most substantial post-permeabilisation change emerges in DAPI density. In the non-permeabilised baseline, LNCaP required more nuclei to reach %SEM $< 2.5\%$ for DAPI density (190 vs 145), consistent with greater intensity-linked heterogeneity in the cancer model. This was also reflected in lower convergence rates in LNCaP ($(7.1 \pm 0.2) \times 10^{-7}$ vs. $(8.36 \pm 0.07) \times 10^{-7}$). After permeabilisation, this relationship was reversed; PNT2-C2 required more than twice the nuclei of LNCaP to reach the same relative precision (300 vs 140), and displayed slower convergence rates ($(9.1 \pm 0.2) \times 10^{-7}$ vs $(1.05 \pm 0.02) \times 10^{-6}$), indicative of greater absolute SEM than the malignant model (Table 4.1). This inversion constitutes a key result: for an intensity normalised descriptor, post-fix permeabilisation is sufficient to reverse the apparent stability ordering between cell lines under a fixed relative-precision criterion. This therefore indicates that stability in intensity-per-area is substantially more challenging

to achieve in the non-malignant comparator after permeabilisation, a trend which was mirrored in total DAPI intensity (Tables B.1 and B.2), though the magnitude of the change in total intensity was much less than the change in density. This suggests that this change is driven by permeabilisation induced changes to the DAPI density (DAPI density = $I_{\mathrm{DAPI}}/\mathrm{area}$), rather than by changes in total DAPI intensity alone. Detergent permeabilisation is known to introduce variable extraction/redistribution of fixed material in cells, and modulate DAPI uptake, thereby altering the effective nuclear boundary and/or coupling between total DAPI intensity and nuclear area, which would preferentially destabilise the observed DAPI density more strongly than area or intensity alone [16, 206, 433–435]. This is a critical outcome; subsequent comparisons (especially any interpretation of intranuclear texture or density heterogeneity) must account for the fact that permeabilisation can play a large role in the observed convergence behaviour for DAPI density in a cell line-specific manner.

Effect of permeabilisation relative to the non-permeabilised baseline

Directly comparing the convergence metrics before and after permeabilisation (Tables 3.1 and 4.1) isolates the effect of detergent permeabilisation on measurement stability, under an otherwise matched analysis pipeline and convergence criterion. Through this comparison, several trends are quantitatively clear.

First, in LNCaP, permeabilisation produced a net improvement in the rate of convergence for the two descriptors most directly related to nuclear size; N_{conv} decreased from 365 to 340 for nuclear area ($\sim 10\%$ reduction), and from 190 to 140 for DAPI density ($\sim 25\%$ reduction). Consistently, γ increased by $\sim 20\%$ for area and $\sim 50\%$ for DAPI density, indicating a smaller SEM amplitude and therefore reduced absolute dispersion post-permeabilisation. Aspect ratio behaved differently, however, showing an increase in N_{conv} from 170 to 195 ($\sim 15\%$), with an approximate 5% increase in γ, suggesting that detergent treatment slightly increases the number of nuclei required to stabilise nuclear shape estimates in LNCaP, while decreasing the number required for metrics dependent on nuclear area. More simply, permeabilisation appears to reduce the variability of nuclear size and DAPI density, while increasing the variability of nuclear shape.

Second, in PNT2-C2, permeabilisation increased the sampling requirement for all three descriptors. The increases are moderate for purely morphological measures, but notably larger for DAPI density. N_{conv} increased from 225 to 310 for area ($\sim 40\%$ increase), from 70 to 100 for aspect ratio ($\sim 40\%$ increase), and from 145 to 300 for DAPI density ($\sim 105\%$ increase). The changes in convergence rates, however, were not consistent across metrics; γ dropped by approximately 50% for area and $\sim 10\%$ for aspect ratio, but in the case of DAPI density, γ increased slightly by around $\sim 10\%$, despite the doubling of N_{conv}. This again highlights that N_{conv} is controlled by relative precision ($\mathrm{SEM}/\bar{x}$) rather than absolute SEM alone. Notably, the reversal in the magnitude of N_{conv} between cell lines observed for DAPI density is directionally concordant with the corresponding changes in nuclear area convergence (with N_{conv} reducing for LNCaP nuclear area after permeabilisation and increasing for PNT2-C2) and total DAPI intensity (converging at similar N_{conv} sans-permeabilisation and increasing by a greater amount in PNT2-C2 post-permeabilisation) (Tables B.1 and B.2). The approximate doubling of N_{conv} for DAPI density implies a substantial increase in the effective coefficient of variation (CV) of this descriptor post-permeabilisation: under $\mathrm{SEM}(N) \propto N^{-1/2}$, $N_{\mathrm{conv}} \propto \mathrm{CV}^2$, so this increase corresponds to an inferred $\sim 45\%$ increase in effective CV for DAPI density in PNT2-C2, even if the absolute dispersion (reflected by γ) changes only modestly.

Third, considered as a heterogeneity contrast between cell lines, permeabilisation systematically reduces the strong LNCaP > PNT2-C2 ordering observed pre-permeabilisation for morphological descriptors, and fully inverts the ordering for intensity based descriptors. More explicitly, the pre-permeabilisation DAPI-density requirement $N_{\text{conv,LNCaP}}/N_{\text{conv,PNT2-C2}} \approx 190/145$ becomes 140/300 post-permeabilisation, indicating a switch from LNCaP having a $\sim 15\%$ higher effective CV to having a $\sim 30\%$ lower effective CV. This is strong evidence that permeabilisation materially changes the heterogeneity landscape between cell lines for intensity-based metrics, and therefore that "permeabilised" and "non-permeabilised" DAPI-derived heterogeneity should not be treated as directly interchangeable readouts of underlying nuclear biology.

Interpretation of permeabilisation effects

The above results are consistent with three mechanistic interpretations, each aligned with established preparation artefacts associated with detergent permeabilisation.

First, detergent permeabilisation can induce heterogeneous loss or redistribution of cellular material even after aldehyde fixation. Previous work on breast cancer cells has shown that Triton X-100 treatment following PFA fixation can produce substantial decreases in cellular mass density and increases in membrane permeability, with significant cell-to-cell variability in the magnitude of mass loss [16]. Additionally, Melan and Sluder (1992) used mammalian cell cultures to demonstrate that detergent exposure post-fixation can redistribute and differentially extract soluble proteins, leading to preparation-dependent changes that are not present in the living state [295]. In the context of this study, such effects provide a direct route by which permeabilisation could broaden the distributions of nuclear area estimates and DAPI density (through cell-to-cell differences in nuclear permeability, dye access, and/or partial loss of nuclear material), thereby shifting convergence rates and N_{conv}.

Second, the direction and magnitude of the convergence shifts are cell-line-dependent, consistent with differential susceptibility to preparation-induced selection (i.e. weakly attached cells may be more likely to be lost through permeabilisation and washing steps) and/or detergent interaction. LNCaP is well documented to be weakly adherent and prone to growth in loosely attached clusters, and plating/coating conditions can measurably alter LNCaP morphology, including nuclear area [163, 367]. As such, an important caveat is that the permeabilisation and wash sequence may preferentially detach or exclude a subset of cells/nuclei, which could result in a reduction in the measured heterogeneity in the retained population without necessarily reflecting a true reduction in intrinsic biological heterogeneity. This provides a plausible explanation for why LNCaP shows faster apparent convergence for area and DAPI density post-permeabilisation, despite being the more heterogeneous line in the non-permeabilised baseline. Conversely, PNT2-derived cultures are typically more uniformly adherent and epithelial in morphology [151, 436], which may reduce detachment-mediated selection while leaving permeabilisation as the dominant new perturbation; consistent with the marked increase in N_{conv} for PNT2-C2 DAPI density.

Third, membrane-lipid composition provides an additional, biologically grounded axis that could modulate detergent effects. Triton X-100 works by solubilising membrane lipids, and the extent of the resultant structural disruption is therefore expected to depend on lipid and cholesterol composition [16, 437]. Prostate cancer cell lines including LNCaP have been reported to exhibit increased levels of markers associated with cholesterol-rich membrane regions relative to non-malignant comparators,

consistent with dysregulated membrane cholesterol organisation in the disease state [329]. While this does not by itself predict the convergence changes observed here, it does motivate the interpretation that detergent–membrane interactions (and their heterogeneity across cells) may differ systematically between malignant and non-malignant prostate models, and may contribute to the observed inversion in DAPI-density convergence behaviour.

The direct impact of permeabilisation does not appear to have been explicitly reported in prostate cancer cell lines, and so the information detailed above constitutes valuable new knowledge. Considering the above points collectively, permeabilisation clearly cannot be treated as a neutral preparatory step with respect to DAPI intensity derived convergence. Instead, it measurably reshapes descriptor stability in a cell line specific manner, with the largest impact on intensity-linked metrics in PNT2-C2. Practically, this implies that subsequent quantitative comparisons of permeabilised datasets should maintain explicit convergence validation for each intensity-linked descriptor and stain condition, and interpret any observed changes in heterogeneity as potentially reflecting variability introduced through sample preparation in addition to biology. The present DAPI-only permeabilised convergence analysis thus provides a critical control baseline for the subsequent fibrillarin based analyses that follow, by quantifying the magnitude and direction of the preparation-induced perturbation prior to any marker-specific interpretation. Accordingly, all downstream analyses of permeabilised samples are interpreted with this quantified preparation dependence in mind, and are accompanied by descriptor-specific convergence validation where intensity normalisation or texture-derived measures are used.

4.1.2 Morphological and intensity analysis

4.1.2.1 Morphology

As in the non-permeabilised baseline (Section 3.2.1), nuclear morphology is assessed using the same DAPI-derived descriptors defined previously: projected nuclear area, nuclear aspect ratio, and the coefficient of variation (CV) for across-nucleus dispersion reporting (Equations (2.7), (2.8) and (2.27)). The key aim here is to quantify how detergent permeabilisation perturbs these morphological descriptors, and critically, whether that perturbation is cell-line dependent.

This is important as nuclear size and shape are widely used as markers of malignancy, and are explicitly leveraged in prostate cancer nuclear morphometry [131, 217, 366]. However, fixation/permeabilisation are not morphologically neutral sample-preparation steps, as discussed in Section 4.1.1. Triton X-100 solubilises membrane lipids to increase intracellular accessibility, and can induce substantial loss of cellular mass density even after PFA fixation [16]. Permeabilisation has already been shown to cause significant changes in the rate of convergence (and thus variance) of several measures, including nuclear area and aspect ratio, and so assessing its effect on these metrics directly (rather than just the standard error convergence) is a logical and necessary next step.

Bulk statistics (mean, mode, and CV) for permeabilised nuclei are summarised in Figure 4.2. For context, the corresponding non-permeabilised values are shown in Figure 3.8. The primary result is that detergent permeabilisation induces a pronounced shift in nuclear area between LNCaP and PNT2-C2, sufficiently large to reverse the ordering observed in non-permeabilised cells.

Post-permeabilisation, PNT2-C2 nuclei were approximately 20% larger than LNCaP nuclei when considering the mean values. This is also reflected by the KDE mode, which also increased by

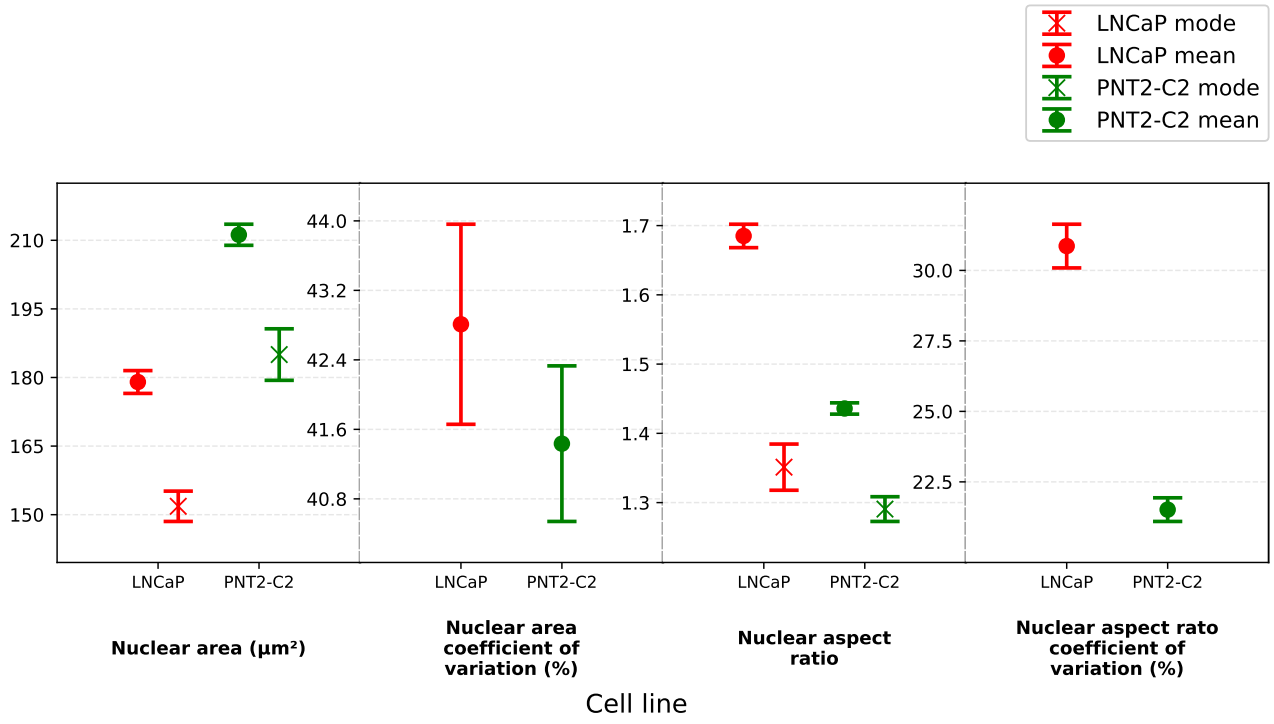


Figure 4.2: Plot showing the mean, mode and coefficient of variation (CV) of nuclear area and nuclear aspect ratio in permeabilised cells stained with only DAPI. Mode here refers to the peak position of the KDE (Equation (2.34)), consistent with the distributional analyses in Section 4.1.2.1. Error bars represent standard error (calculated using the delta-method where required for CV and mode estimates; Equations (2.28) and (2.29)). LNCaP nuclei had a mean area of $(179 \pm 2)\,\mu m^2$, modal area of $(152 \pm 3)\,\mu m^2$, area CV of $(43 \pm 1)\,\%$, mean aspect ratio of 1.69 ± 0.02, modal aspect ratio of 1.35 ± 0.03, and aspect ratio CV of $(30.9 \pm 0.8)\,\%$, while PNT2-C2 nuclei had a mean area of $(211 \pm 2)\,\mu m^2$, modal area of $(185 \pm 6)\,\mu m^2$, area CV of $(41.4 \pm 0.9)\,\%$, mean aspect ratio of 1.436 ± 0.008, modal aspect ratio of 1.29 ± 0.02, and aspect ratio CV of $(21.5 \pm 0.4)\,\%$.

$\sim 20\%$ between LNCaP and PNT2-C2. The area CV, however, was commensurate within uncertainty between the two cell lines ($(43 \pm 1)\,\%$ in LNCaP vs. $(41.4 \pm 0.9)\,\%$ in PNT2-C2). This suggests that permeabilisation drives both cell lines towards comparable relative area dispersion, despite a clear shift in absolute scale, a trend which is supported by the findings from convergence testing (Table 4.1). Nuclei in the malignant cell line remained more elongated after permeabilisation; LNCaP had a greater mean and modal aspect ratio than PNT2-C2 (1.69 ± 0.02 and 1.35 ± 0.03 in LNCaP vs. 1.436 ± 0.008 and 1.29 ± 0.02 in PNT2-C2), with the mean value being approximately 20% higher in LNCaP, and the mode value $\sim 5\%$ higher. The greater difference in means compared to modes suggests that LNCaP has a greater fraction of nuclei with large aspect ratios, with more details and a full analysis of the distribution in Section 4.1.2.1. Similarly, shape heterogeneity remains higher in LNCaP (aspect ratio CV $\sim 30\%$ vs. $\sim 20\%$), however the difference between cell lines is materially smaller than in the non-permeabilised baseline (Figure 3.8), indicating that permeabilisation again acts to erode differences in dispersion between cell lines (with this also being in agreement with conclusions from convergence testing).

Direct comparison of the bulk morphology summaries in (Figures 3.8 and 4.2) shows that detergent permeabilisation introduces a strong, cell line-dependent change in nuclear area. In LNCaP, the mean nuclear area decreased from $(195 \pm 3)\,\mu m^2$ before permeabilisation to $(179 \pm 2)\,\mu m^2$ after, a reduction of $\sim 10\%$. In PNT2-C2, however, the mean nuclear area increased from $(136 \pm 1)\,\mu m^2$ to $(211 \pm 2)\,\mu m^2$ (an increase of approximately 55%) under the same conditions. This constitutes a reversal of the between cell line ordering; non-permeabilised LNCaP nuclei were larger than PNT2-C2 by approximately 45%,

whereas following permeabilisation, PNT2-C2 nuclei were larger than LNCaP by $\sim 20\%$. This is a clear, and significant change induced as a result of permeabilisation, demonstrating that this pre-analytical step can dominate the apparent biological separation in nuclear size.

The heterogeneity in this measure is also differentially perturbed. LNCaP area CV decreases slightly (however remains within uncertainty), while PNT2-C2 area CV increases substantially from $(34.5\pm0.6)\,\%$ to $(41.4 \pm 0.9)\,\%$, a relative change of $\sim 20\%$. Simply, permeabilisation appears to not only increase the absolute nuclear area, but also the relative dispersion in the non-malignant model, while decreasing the nuclear area in the malignant model, but leaving the dispersion relatively unchanged. It must be said that while the mean nuclear area did change materially in LNCaP after permeabilisation, the magnitude of this change is relatively small, and so may simply be the result of natural variations between experiments.

Considering aspect ratio, permeabilisation can be seen to have a smaller effect on mean elongation than on size. LNCaP mean aspect ratio decreased slightly following permeabilisation from 1.72 ± 0.02 to 1.443 ± 0.006 (a decrease of $\sim 15\%$), while PNT2-C2 did not change outside of uncertainty. In contrast, PNT2-C2 aspect ratio CV increased from $(18.7 \pm 0.3)\,\%$ to $(21.5 \pm 0.4)\,\%$ (an approximate 15% increase), while LNCaP aspect ratio CV remained unchanged within uncertainty. Taken as a whole, the observed changes in bulk statistics are compatible with a predominantly isotropic size change (i.e. uniform in all directions), in conjunction with broadly preserved shape trends, however distributional analysis is required to determine whether these bulk shifts result from a global rescaling, or tail-specific changes.

Mechanistically, the differing change in area between cell lines is consistent with permeabilisation altering the physical state of fixed cells in a line-specific manner. Detergent permeabilisation removes membrane lipids to enable molecular access [437], and can drive leakage of intracellular macromolecules [16]. Additionally, once the plasma membrane barrier is disrupted, permeabilised cells can show enhanced volume responses to osmotic conditions [438], implying altered nucleo-cytoplasmic coupling. Since nuclear size and shape reflect chromatin compaction, lamina composition, and cytoskeletal stresses, differences in nuclear mechanical properties across prostate-derived lines (such as differences in stiffness/lamin-dependent mechanics reported across prostate models) provide a plausible basis for why PNT2-C2 and LNCaP may respond differently [131, 380].

Another possible cause is the preparation-induced selection mentioned in Section 4.1.1. Potentially, both cell lines are affected equally by permeabilisation, and a similar nuclear enlargement would be observed for both; however, the larger nuclei in the more weakly adherent LNCaP line may have been disproportionately lost during sample preparation, such that the observed mean nuclear area appeared smaller as a result [163, 367]. More explicitly, this selection could be amplified if LNCaP nuclei are mechanically less robust as a result of nuclear-envelope alterations associated with malignancy. Across prostate cancer cell systems, deformed nuclei and reduced/mislocalised lamins (including lamin A/C) increase in prominence with disease progression, and reduced lamin A/C has been linked to lower nuclear stiffness and increased nuclear-envelope failure in previous studies[197, 330, 380]. In LNCaP specifically, nuclear blebbing has been characterised as an abnormal morphological state relevant to prostate cancer [439]. Given that detergent permeabilisation solubilises lipid membranes in a concentration and time dependent manner, detergent exposure may preferentially compromise the most deformed/blebbed nuclei, thereby increasing detachment during subsequent washes and/or exclusion during image-analysis

quality control [440–442]. Nonetheless, because nuclei were not tracked longitudinally through the permeabilisation step (i.e. the exact same nuclei were not imaged before and after permeabilisation), these data quantify the overall effect of permeabilisation on the measured morphological quantities, but do not isolate an exact biophysical pathway by which they may occur.

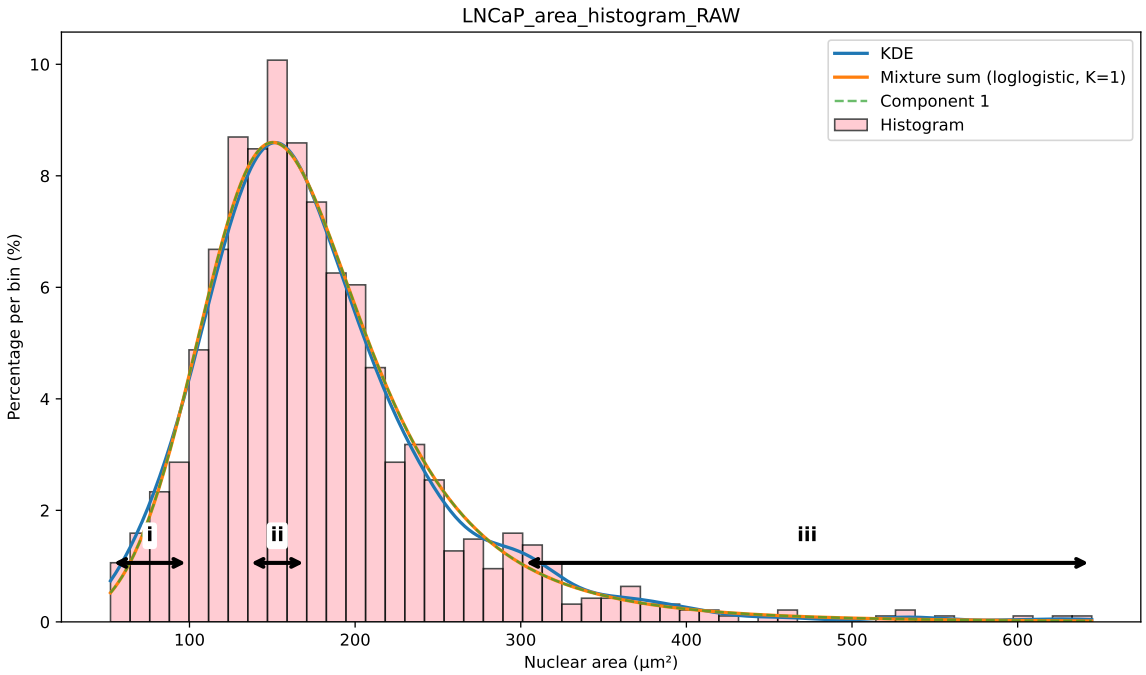
Distributional analysis using histograms

As previously Section 3.2.1, analysis of histograms/kernel density estimates was carried out to expose skewness and tail behaviour which is not clearly visible in bulk statistics, as well as contextualise the modal values given in Figure 4.2. The central question here is whether the large increase in area observed in PNT2-C2 nuclei arises from a uniform distributional shift, or from preferential expansion of a subpopulation (such as an enhanced high area tail), and whether any analogous tail changes occur in LNCaP.

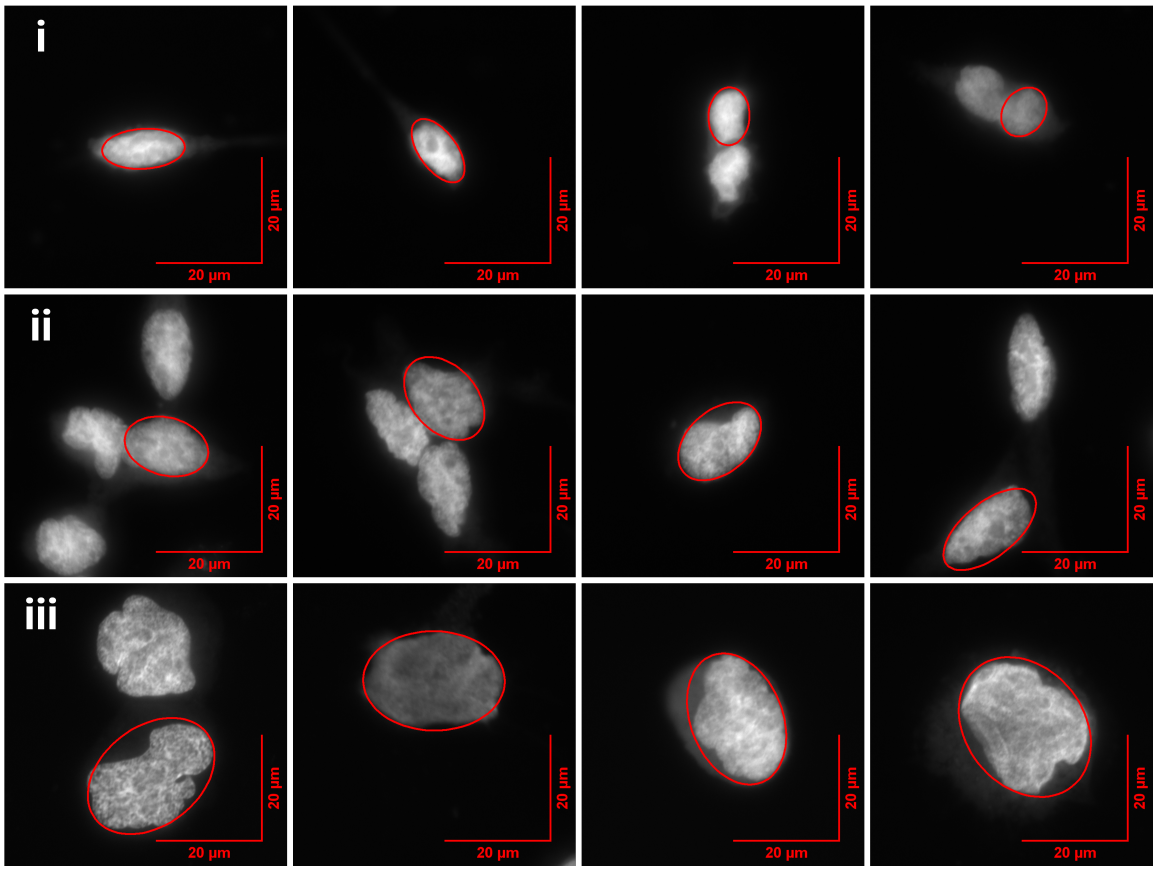
The nuclear area distribution across both cell lines was best described by a single log-logistic model (Figures 4.3 and 4.4), rather than the log-normal model used in the non-permeabilised baseline (Figures 3.9 and 3.10). Practically, this indicates that a heavier-tailed parametric form provides a better approximation after permeabilisation, consistent with increased probability mass at large areas. However, the biological interpretation depends on how that tail mass differs between cell lines, and relative to the non-permeabilised state.

In LNCaP (Figure 4.3), the KDE mode lay at $(152 \pm 3)\,\mu m^2$ and the distribution was right-skewed with a significant tail extending beyond $600\,\mu m^2$. Consistent with this skew, the mean nuclear area ($(179 \pm 2)\,\mu m^2$) exceeded the KDE mode by approximately 20%, quantifying the leverage of rare high-area nuclei on the mean. Considering the tail explicitly, $\sim 6.5\%$ of nuclei had an area exceeding $300\,\mu m^2$, and the 90th percentile was 272 μm^2. Compared with the non-permeabilised baseline (where the KDE mode was $(152 \pm 4)\,\mu m^2$, the mean was $(195 \pm 3)\,\mu m^2$ (corresponding to a mean-mode separation of $\sim$ 30%), $\sim$ 10% of LNCaP nuclei exceeded $300\,\mu m^2$, and the 90th percentile was $\sim 315\,\mu m^2$; Section 3.2.1), the LNCaP distribution therefore exhibits a marked reduction in mean-mode separation and extreme tail behaviour following permeabilisation. The consistency of the KDE mode before and after permeabilisation confirms that the reduction in mean area is driven by a reduced number of high area nuclei, rather than a shift in the size of the typical LNCaP nucleus. Representative nuclei shown in Figure 4.3(b) demonstrate that the high area region contains genuinely large nuclei, as well as occasional morphologies consistent with partial mergers/over-segmentation, and as such the tail should be interpreted as a combination of biological extremes and an upper bound on segmentation failure rate. Importantly, the tail contraction relative to the non-permeabilised baseline provides distribution-level evidence to suggest that permeabilisation does not simply "inflate" all nuclei in LNCaP. Instead, it preferentially reduces the prevalence of the most enlarged nuclei (or otherwise acts to reduce their measured projected area), thus shifting the mean towards the mode, without affecting the mode itself.

In permeabilised PNT2-C2 (Figure 4.4), the distribution was also right-skewed, but was globally shifted to larger areas than LNCaP, with KDE mode $(185 \pm 6)\,\mu m^2$, and a greater proportion of data residing in the high-area tail. The mean nuclear area ($(211 \pm 2)\,\mu m^2$) exceeded the mode by approximately 15%, again demonstrating that tail structure materially inflates the mean above the modal nuclear

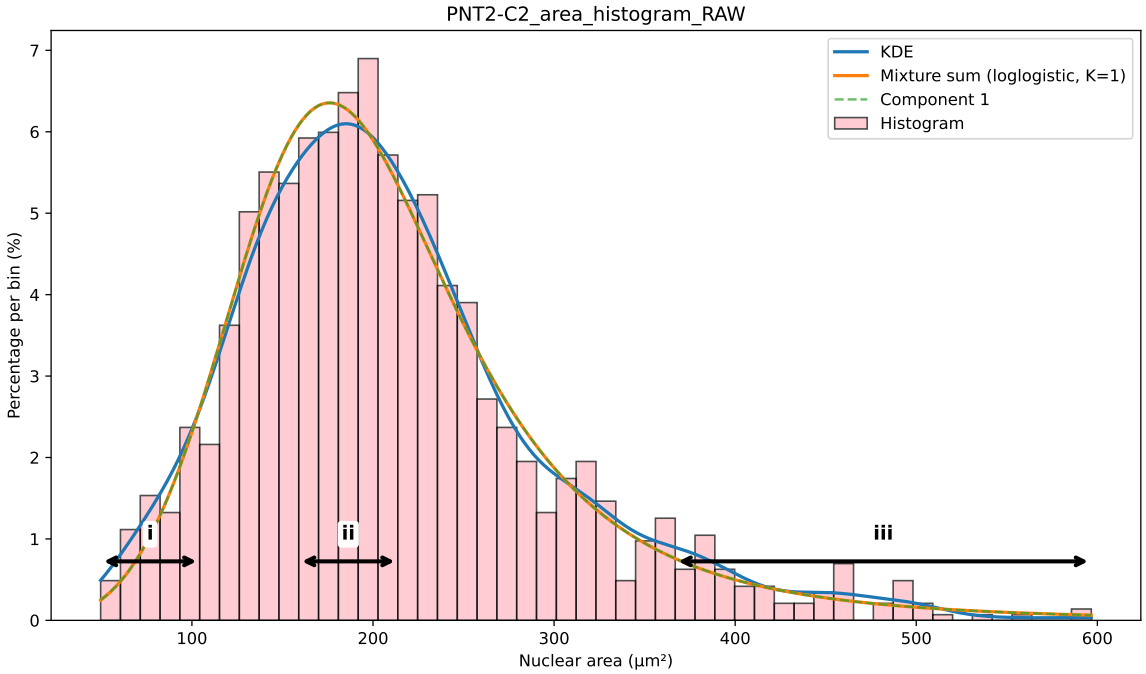


(a) Histogram showing the area in each LNCaP nucleus. A KDE is overlaid on top of the histogram, with fitting applied to the KDE to determine the overall distribution of the data [258]. The best model (chosen by ICL [259]) was a single log-logistic distribution with $c = 4.6 \pm 0.1$, loc $= 0$, and scale $= 166 \pm 2$, with a peak at $(150 \pm 2)\,\mu m^2$. The KDE peak location was $(152 \pm 3)\,\mu m^2$.

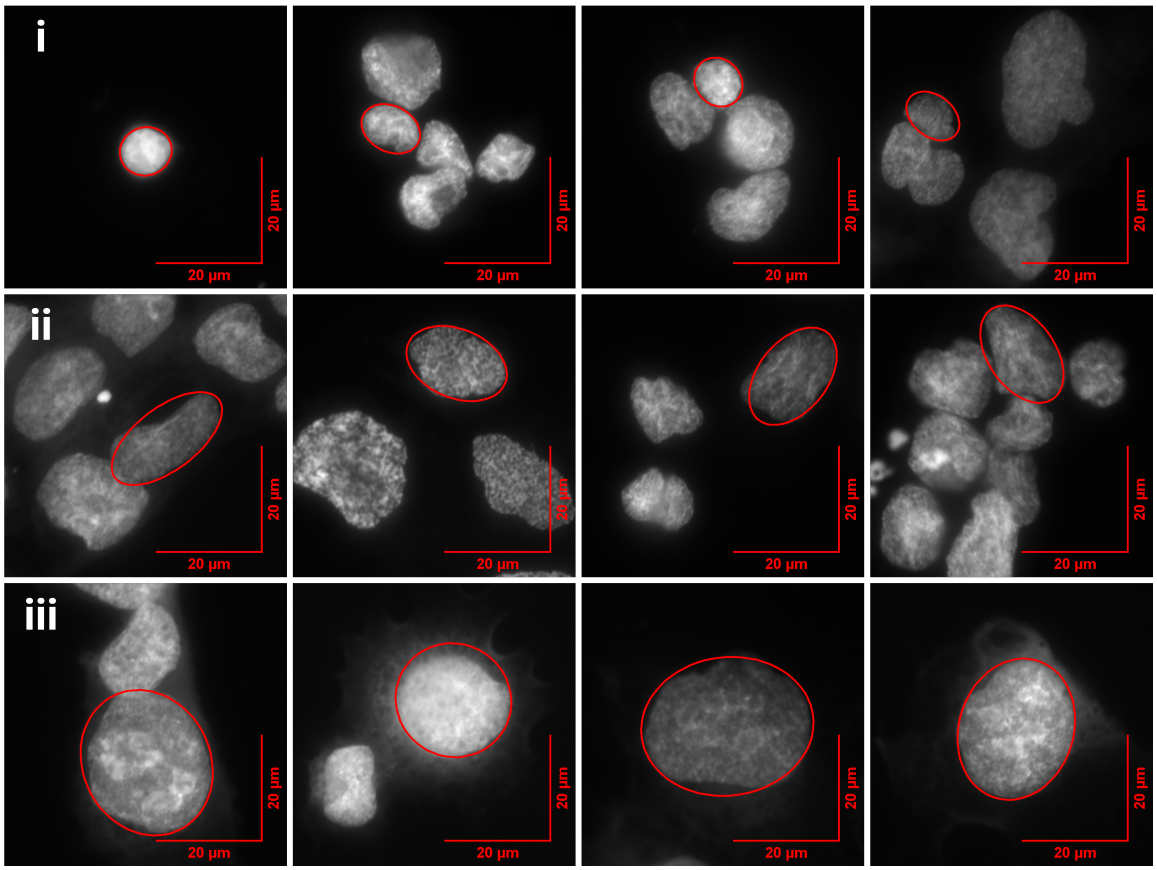


(b) Images of detected nuclei selected from the low (i), modal (ii), and high (iii) regions of the corresponding histogram. Red ellipses are drawn purely for illustrative purposes to show the approximate region which was detected as a nucleus by the segmentation code. Area was calculated using pixel counts within the nuclei mask. A 20 µm scale bar is shown in red at the bottom right corner of each image.

Figure 4.3: Histogram showing the nuclear area in permeabilised LNCaP cells stained with only DAPI ((a)), with some images of representative nuclei from the low, modal, and high regions of the histogram ((b)). The KDE is used as a basis for fitting, with the histogram being purely illustrative. The best fit was chosen using an Integrated Classification Likelihood (ICL) [259]. Uncertainties are calculated from the covariance matrix.



(a) Histogram showing the area in each PNT2-C2 nucleus. A KDE is overlaid on top of the histogram, with fitting applied to the KDE to determine the overall distribution of the data [258]. The best model (chosen by ICL [259]) was a single log-logistic distribution with c = 4.3 ± 0.1, loc = 0, and scale = 196 ± 2, with a peak at (176 ± 2) µm^2. The KDE peak location was (185 ± 6) µm^2.



(b) Images of detected nuclei selected from the low (i), modal (ii), and high (iii) regions of the corresponding histogram. Red ellipses are drawn purely for illustrative purposes to show the approximate region which was detected as a nucleus by the segmentation code. Area was calculated using pixel counts within the nuclei mask. A 20 µm scale bar is shown in red at the bottom right corner of each image.

Figure 4.4: Histogram showing the nuclear area in permeabilised PNT2-C2 cells stained with only DAPI ((a)), with some images of representative nuclei from the low, modal, and high regions of the histogram ((b)). The KDE is used as a basis for fitting, with the histogram being purely illustrative. The best fit was chosen using an Integrated Classification Likelihood (ICL) [259]. Uncertainties are calculated from the covariance matrix.

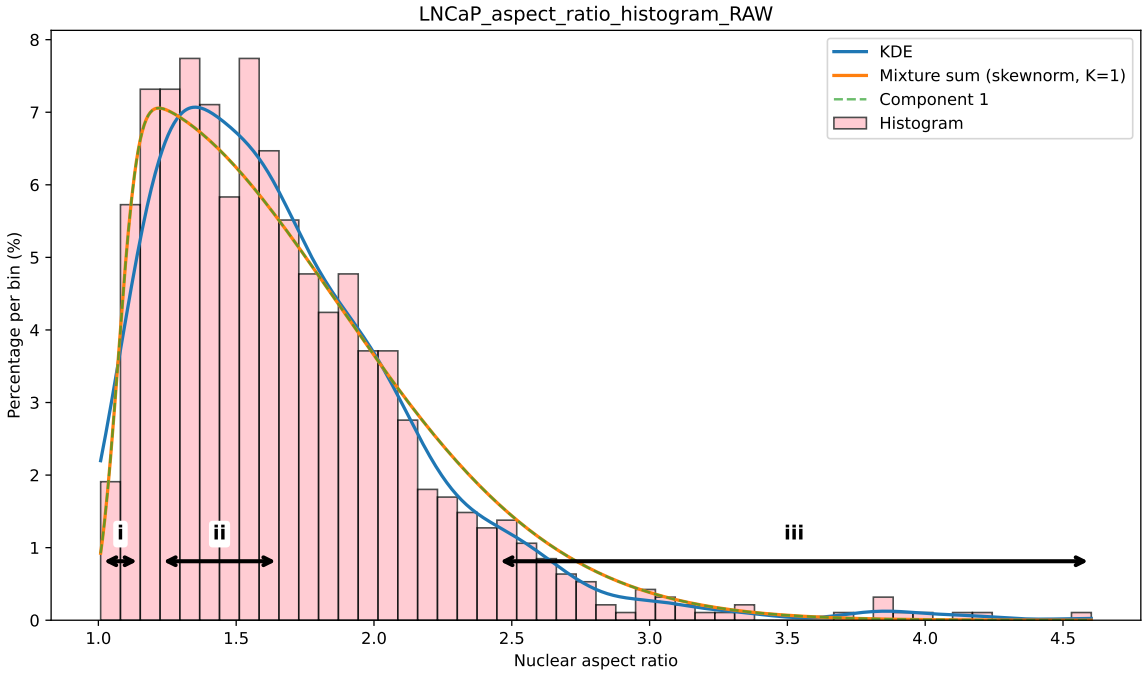
size. Quantitatively, $\sim$ 15% of nuclei had an area exceeding 300 µm^2, and the 90th percentile was $\sim$330 µm^2. These values contrast starkly with the non-permeabilised baseline, where the KDE mode was (120 ± 9) µm^2, mean area was (136 ± 1) µm^2 (giving a mean-mode separation of $\sim$ 15%), only $\sim$ 1% of PNT2-C2 nuclei had an area greater than 300 µm^2, and the 90th percentile was $\sim$195 µm^2 (Section 3.2.1). Therefore, the large increase in mean area in PNT2-C2 after permeabilisation is not solely the result of a shift near the mode; the modal change is accompanied by a significant increase in "tailedness". The increase in mean-mode separation is consistent with both the large increase in tail mass, and the increase in area dispersion (with area CV rising from (34.5 ± 0.6) % to (41.4 ± 0.9) %). Notably, although the absolute mean-mode separation in nuclear area is similar between PNT2-C2 and LNCaP ($\sim$25 µm^2), the much larger fraction of high area nuclei in PNT2-C2 indicates that the post-permeabilisation ordering is dominated by differences in upper-tail mass rather than a shift of the central bulk.

Biologically, PNT2-C2 is derived from normal prostate epithelium and immortalised by SV40 transfection, and is often used as a comparatively non-malignant reference cell line in prostate studies [3, 167, 369]. The observation that permeabilisation induces both an increase in modal area, as well as a strong "giant nucleus" tail in this cell line is therefore unlikely to represent a genuinely disease-associated morphometric feature. Rather, it provides direct quantitative evidence that detergent permeabilisation can impose a morphology distribution that mimics (and exceeds) established markers of nucleomegaly used in cancer morphometry [131, 217, 359, 394]. This behaviour is mechanistically consistent with detergent-induced perturbation of membrane integrity and material content, combined with altered susceptibility to osmotic effects and reduced mechanical coupling following permeabilisation [16, 438]. Since nuclear size is regulated by chromatin state, lamina composition, and cytoskeletal stresses, any intrinsic differences in nuclear mechanics or chromatin organisation between "normal" and malignant prostate models may also plausibly result in differing morphological responses to permeabilisation [131, 184, 397, 400]. Additionally, the representative nuclei in Figure 4.4(b) indicate that the extreme high-area region contains very large individual nuclei (rather than exclusively mergers/segmentation artefacts), supporting the interpretation that the tail at least partially reflects a real increase in the prevalence of measured high area nuclei in the permeabilised state.

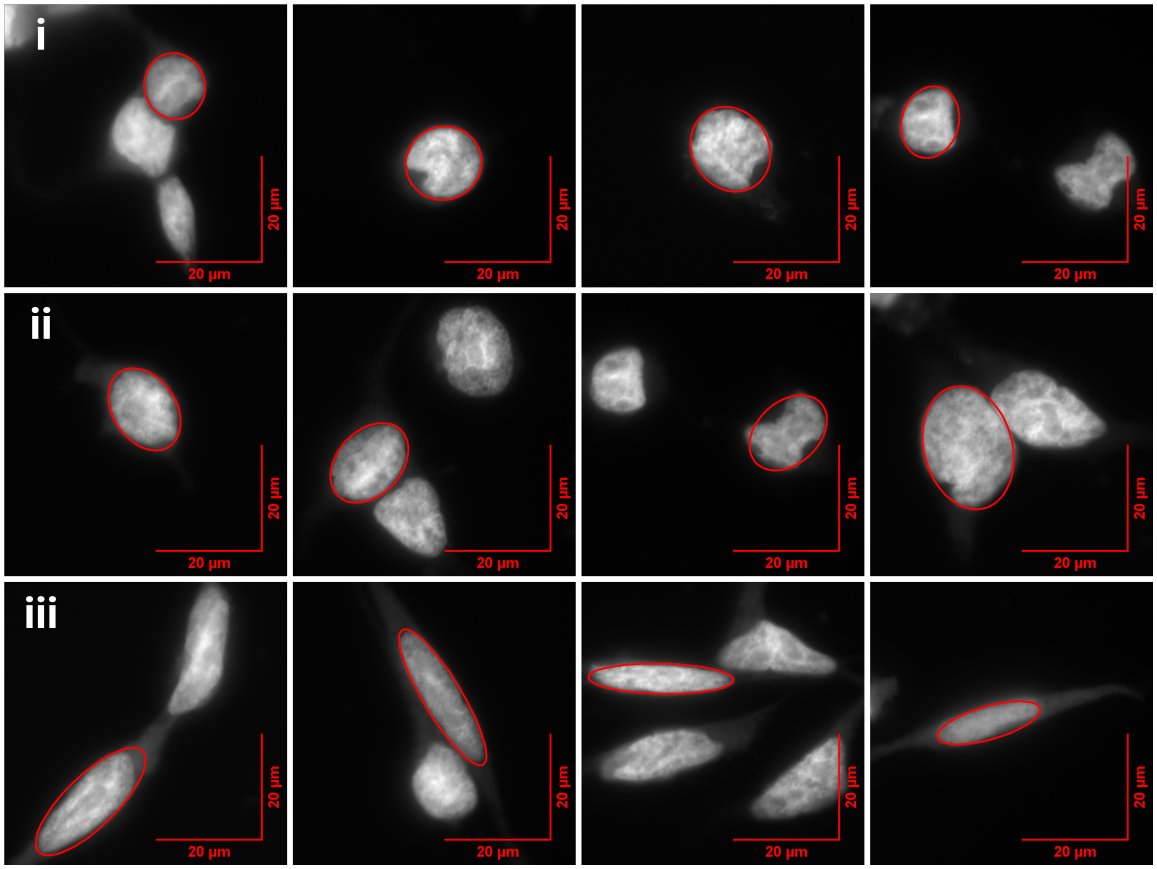
A crucial finding provided by this analysis is that detergent permeabilisation can have an effect strong enough to invert the apparent nuclear size relationship between LNCaP and PNT2-C2 cells; without a baseline comparison, it is possible that one could incorrectly conclude that PNT2-C2 has intrinsically larger nuclei than LNCaP based on the results of this experiment, whereas the paired analysis reveals that this ordering is driven by the effects of permeabilisation, and was not observed prior to detergent exposure.

Moving on to nuclear aspect ratio, this measure remained positively skewed across both cell lines, with skew-normal distributions best describing the spread of data, as in the non-permeabilised baseline.

In permeabilised LNCaP (Figure 4.5), the distribution showed a strong positive skew, with modal aspect ratio being 1.35 ± 0.03, and a had long tail containing aspect ratio values in excess of 4. The mean aspect ratio of 1.69 ± 0.02 surpassed the mode by approximately 25%, demonstrating the substantial influence of the tail in this metric. Using a practical tail threshold, $\sim$ 20% of nuclei had aspect ratio > 2, and the 90th percentile was 2.47. Relative to the non-permeabilised baseline (Figure 3.12), the 90th percentile decreased from 2.47 to 2.32 (a reduction of $\sim$ 5%), and the percentage of nuclei with



(a) Histogram showing the aspect ratio in each LNCaP nucleus. A KDE is overlaid on top of the histogram, with fitting applied to the KDE to determine the overall distribution of the data [258]. The best model (chosen by ICL [259]) was a single skew-normal distribution with a $= 14 \pm 3$, loc $= 1.07 \pm 0.01$, and scale $= 0.80 \pm 0.02$, with a peak at 1.22 ± 0.04. The KDE peak location was 1.35 ± 0.03.



(b) Images of detected nuclei selected from the low (i), modal (ii), and high (iii) regions of the corresponding histogram. Red ellipses are drawn purely for illustrative purposes to show the approximate region which was detected as a nucleus by the segmentation code. Aspect ratio was calculated as the ratio between major and minor ellipse axes. A 20 µm scale bar is shown in red at the bottom right corner of each image.

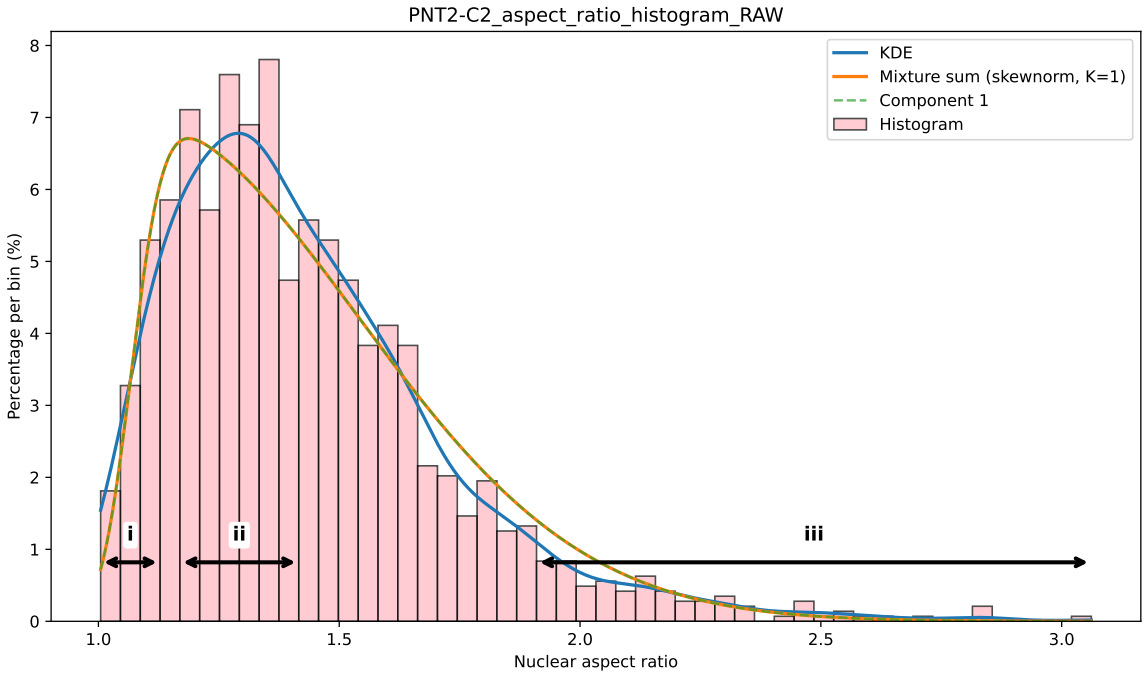
Figure 4.5: Histogram showing the nuclear aspect ratio in permeabilised LNCaP cells stained with only DAPI ((a)), with some images of representative nuclei from the low, modal, and high regions of the histogram ((b)). The KDE is used as a basis for fitting, with the histogram being purely illustrative. The best fit was chosen using an Integrated Classification Likelihood (ICL) [259]. Uncertainties are calculated from the covariance matrix.

an aspect ratio greater than 2 reduced from $\sim 25\%$ to $\sim 20\%$, indicating that permeabilisation only very modestly suppresses the most elongated nuclei.

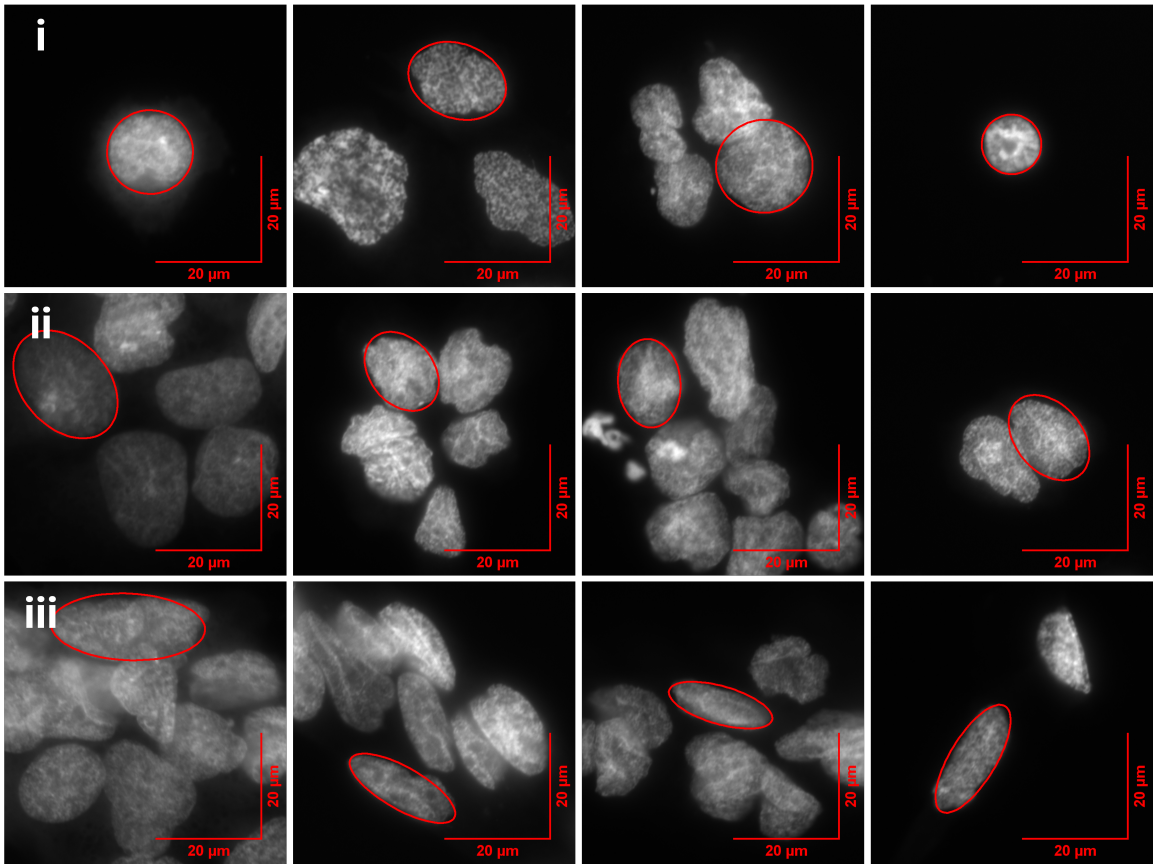
Notably, both the modal and mean aspect ratio decreased following permeabilisation. Specifically, the mode and mean in non-permeabilised LNCaP were 1.43 ± 0.04 and 1.72 ± 0.02 respectively, representing a mean-mode separation of approximately 20%. Comparatively, the mode reduced by $\sim 5\%$, while the mean reduced by $\sim 2\%$ after permeabilisation. While the reduction in tail mass indicated by the threshold fraction and 90th percentile may at first seem to be in contrast with the increase in mean-mode separation, this is explained by considering the relative shifts in mean and mode; a modal shift towards more circular nuclei can increase the relative mean-mode separation if a substantial elongated tail persists to reduce the effect of this shift on the mean. More simply, the data are consistent with permeabilisation producing a modest rounding of the typical (most probable) LNCaP nucleus, while leaving a pronounced elongated fraction of nuclei largely intact, such that the mean remains tail-dominated. Nevertheless, the observed changes in mean-mode separation are not large relative to the combined uncertainty, and the aspect ratio CV is remains unchanged between conditions (Figures 3.8 and 4.2). As such, while the observations here are commensurate with a weak permeabilisation-induced rounding of the central population, they do not support any strong claims that permeabilisation materially affects the global "tailedness" of aspect ratio in LNCaP beyond expected between-experiment variability.

Nuclear shape can be interpreted as a balance between externally applied cytoskeletal/adhesive stresses, and intrinsic nuclear structural resistance provided by lamina and chromatin [131, 402, 443]. In this framework, high aspect ratio nuclei represent states in which stress transmission to the nucleus is sufficient to overcome nuclear resistance, sustaining elongated or indented projections [444, 445]. As previously stated, detergent permeabilisation solubilises lipid membranes in a dose and exposure dependent manner, and can promote redistribution/extraction of soluble intracellular components even after fixation [16, 295, 441, 446]. Consequently, permeabilisation may reduce the effective mechanical constraints and alter nucleo-cytoplasmic force transmission, allowing the most mechanically stressed nuclei (contained in the tail) to relax towards more compact projected shapes.

In permeabilised PNT2-C2 (Figure 4.6), the distribution was more tightly concentrated at low values than in LNCaP, with KDE mode 1.29 ± 0.02 and comparatively limited tail mass. The mean was 1.436 ± 0.008, and exceeded the mode by approximately 10%, a substantially smaller fraction than in LNCaP (where mean-mode separation was $\sim 25\%$), consistent with a weaker contribution from extreme elongation to the mean. In the non-permeabilised baseline for PNT2-C2, the mean (1.443 ± 0.006) also exceeded the mode (1.32 ± 0.03) by $\sim 10\%$, consistent with the essentially unchanged absolute values of mean and modal aspect ratio across permeabilisation states. This stability was also observed for the 90th percentile (1.82 to 1.81), and > 2 fraction ($\sim 5\%$ both before and after permeabilisation), implying that permeabilisation does not have any substantial effect on nuclear aspect ratio in PNT2-C2, despite the dramatic increase in nuclear area. Bulk statistics, however, show that the aspect ratio CV for PNT2-C2 increases significantly after permeabilisation (Figures 3.8 and 4.2), and Figure 4.6(b) shows a slight increase in the spread of aspect ratio values, with the presence of occasional highly elongated/irregular nuclei (including potential segmentation edge cases in crowded fields) when compared to Figure 3.13. In practical terms, permeabilisation appears to slightly increase the frequency of "shape outliers" in PNT2-C2 without substantially shifting the typical nucleus away from near-circularity. The near constant mean and mode imply that these outliers preferentially inflate dispersion metrics such as CV,



(a) Histogram showing the aspect ratio in each PNT2-C2 nucleus. A KDE is overlaid on top of the histogram, with fitting applied to the KDE to determine the overall distribution of the data [258]. The best model (chosen by ICL [259]) was a single skew-normal distribution with a $= 9 \pm 1$, loc $= 1.069 \pm 0.008$, and scale $= 0.47 \pm 0.01$, with a peak at 1.19 ± 0.02. The KDE peak location was 1.29 ± 0.02.



(b) Images of detected nuclei selected from the low (i), modal (ii), and high (iii) regions of the corresponding histogram. Red ellipses are drawn purely for illustrative purposes to show the approximate region which was detected as a nucleus by the segmentation code. Aspect ratio was calculated as the ratio between major and minor ellipse axes. A 20 µm scale bar is shown in red at the bottom right corner of each image.

Figure 4.6: Histogram showing the nuclear aspect ratio in permeabilised PNT2-C2 cells stained with only DAPI ((a)), with some images of representative nuclei from the low, modal, and high regions of the histogram ((b)). The KDE is used as a basis for fitting, with the histogram being purely illustrative. The best fit was chosen using an Integrated Classification Likelihood (ICL) [259]. Uncertainties are calculated from the covariance matrix.

rather than exerting sufficient leverage to have any substantial effect on central tendency measures, whereas in LNCaP, the increase in mean-mode separation reflects a tail that remains a significant contributor to the mean, even when the tail has been slightly suppressed. This again emphasises that permeabilisation can preferentially perturb the tails and heterogeneity of morphology distributions in a way that is undetectable when only considering mean values.

Combining the results from bulk statistics and distributional analysis, the most significant morphological consequence of permeabilisation is a strong, cell line dependent remodelling of nuclear area distributions. LNCaP showed a modest reduction in mean area, resulting from the suppression of extreme high-area nuclei, with modal area remaining unchanged, whereas PNT2-C2 showed a large increase in mean area, driven by both an increase in modal area as well as a notable increase in tail mass. In contrast, nuclear eccentricity patterns remained relatively preserved in mean terms (with LNCaP nuclei continuing to exhibit greater aspect ratio than their PNT2-C2 counterparts), with permeabilisation acting primarily to modestly suppress the eccentricity of the most extreme LNCaP nuclei, and increase the shape heterogeneity in PNT2-C2 nuclei.

In the broader prostate-cancer morphometry context, where larger nuclei and increased pleomorphism are treated as malignancy-associated signals, this constitutes a significant pre-analytical source of bias that can shift the measured distributions independently of intrinsic biology [131, 217, 360, 366]. Detergent permeabilisation has been observed through this analysis to introduce changes in nuclear area that are similar in direction and magnitude to disease-associated morphometric differences, and can even exceed and invert them. This reinforces the idea that the permeabilisation conditions should be treated as an experimental factor that must be rigorously controlled or explicitly modelled when comparing nuclear morphology across prostate-derived cell lines.

4.1.2.2 Intensity Analysis

Here, bulk statistics for several DAPI intensity metrics are used to assess DNA associated fluorescence under fixed acquisition settings, and are interpreted comparatively across preparation conditions. As in Section 3.2.2, the following descriptors are considered:

1. The total DAPI intensity per nucleus, I_{DAPI} (Equation (2.9))
2. The DAPI density per nucleus, $\rho_{\mathrm{DAPI},i}$ (Equation (2.10))
3. The coefficient of variation (CV) of total DAPI intensity across nuclei (Equation (2.27))
4. The coefficient of variation (CV) of DAPI density across nuclei (Equation (2.27))
5. The coefficient of variation (CV) of DAPI pixel intensity within each nucleus, $\mathrm{CV}_{\mathrm{DAPI}}$ (Equation (2.13))

Throughout, mean±SE and mode±$\mathrm{mode}_{\mathrm{SE}}$ are reported, with direct comparisons to the non-permeabilised baseline (Section 3.2.2). As discussed previously (Section 2.3.5), mode refers to the peak position of the kernel density estimate (KDE) (Equation (2.34)), with a more detailed discussion of the mode in relation to the shape of the distribution given in Section 4.1.2.2.

Relative to the non-permeabilised baseline (Figure 3.14), permeabilisation induced a consistent reduction in the mean DAPI density of both cell lines, however the underlying drivers of this change differ

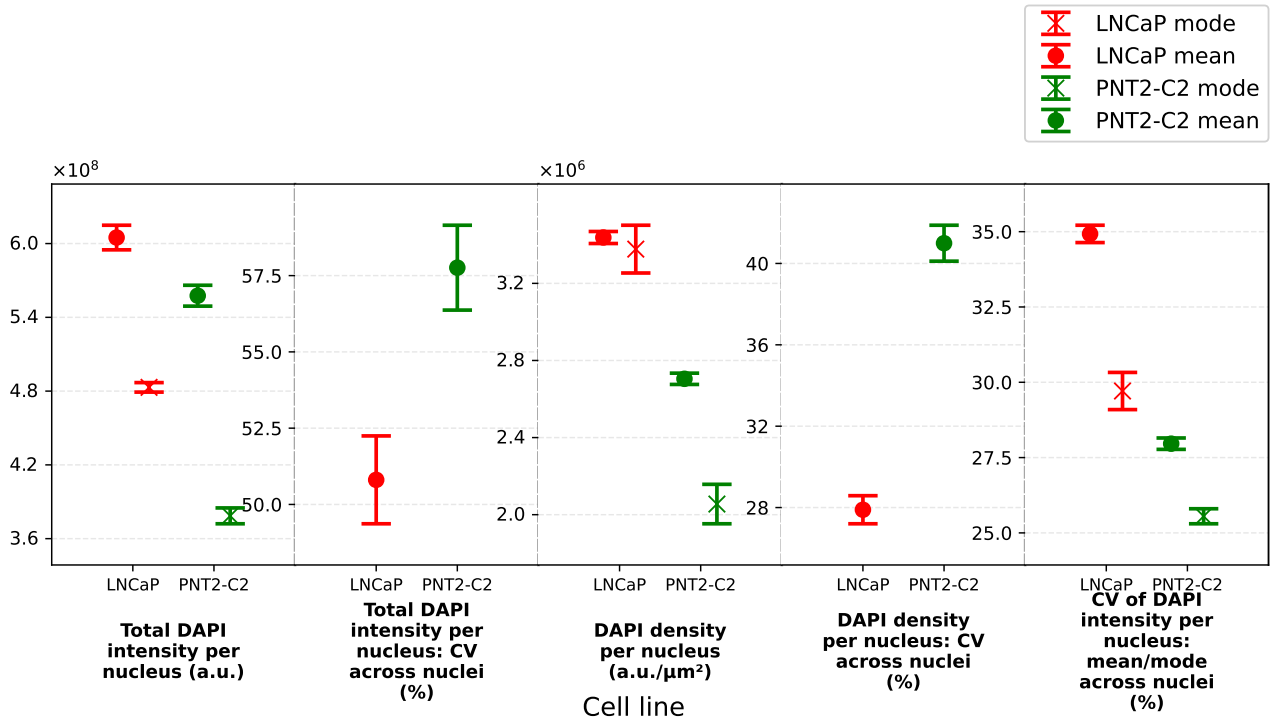


Figure 4.7: Plot showing key DAPI intensity summary statistics for permeabilised LNCaP and PNT2-C2 cells stained with only DAPI. Total DAPI intensity per nucleus was the sum of all pixel intensities in the nuclei mask for a given nucleus. DAPI density was the total intensity divided by nuclear area. CV of DAPI intensity per nucleus was the coefficient of variation of all pixel intensity values in the nuclei mask for a given nucleus. Mode here refers to the peak position of the KDE (Equation (2.34)), consistent with the distributional analyses in Section 4.1.2.2. Error bars represent standard error (calculated using the delta method for CV across nuclei and modes; Equations (2.28) and (2.29)).

once the relationship between density, total intensity, and area is considered. In LNCaP, total intensity decreased from $(8.1 \pm 0.2) \times 10^8$ a.u. to $(6.0 \pm 0.1) \times 10^8$ a.u. following permeabilisation (a relative decrease of $\sim 25\%$), while DAPI density decreased from $(4.34 \pm 0.06) \times 10^6$ a.u. µm^{-2} to $(3.44 \pm 0.03) \times 10^6$ a.u. µm^{-2} (approximately 20%), implying only a modest inferred change in nuclear area, i.e. the density decrease in LNCaP is largely driven by changes in intensity. In contrast, under the same conditions, PNT2-C2 showed a relatively small change in total intensity ($(5.75 \pm 0.05) \times 10^8$ a.u. to $(5.58 \pm 0.08) \times 10^8$ a.u., $\sim -2.5\%$) but a large reduction in DAPI density ($(4.28 \pm 0.02) \times 10^6$ a.u. µm^{-2} to $(2.70 \pm 0.03) \times 10^6$ a.u. µm^{-2}, $\sim -35\%$), implying a $\sim 1.5\times$ increase in effective area, a change which was corroborated by direct area measurements shown in Figures 3.8 and 4.2. This quantitative divergence strongly supports the previous interpretation from Section 4.1.2.1 that permeabilisation can have a strong effect on nuclear area in a cell line dependent manner, rather than solely acting as a neutral, stain-access step.

In the non-permeabilised condition, LNCaP exhibited substantially higher mean total DAPI intensity than PNT2-C2 (Figure 3.14), with a $\sim 40\%$ relative increase from PNT2-C2 to LNCaP, consistent with the established aneuploid tendency of LNCaP relative to immortalised non-tumourigenic prostate epithelium models such as PNT2-C2 [1, 2, 151, 165]. Following permeabilisation, this contrast was markedly reduced, with LNCaP only showing an approximate 10% increase in total DAPI intensity compared to PNT2-C2, however when considering that PNT2-C2 had larger nuclei on average than LNCaP following permeabilisation, this indicates that permeabilisation acts to preferentially suppress the observed DAPI intensity in LNCaP nuclei when compared to PNT2-C2.

The same trend between cell lines was not observed for DAPI density; in the non-permeabilised condition, DAPI density was relatively similar between cell lines, with the mean overlapping within uncertainty, and differences only being seen when considering the modes (where LNCaP had $\sim 15\%$ greater DAPI density). After permeabilisation, differences between cell lines were much more pronounced, with the mean DAPI density in LNCaP being approximately 30% greater than in PNT2-C2, with modal density being almost 60% greater in LNCaP. Considering these findings in conjunction with the results from Section 4.1.2.1, this suggests that permeabilisation increased the segmented area and inclusion of dim pixels more strongly in PNT2-C2. Notably, these results do not support a preparation-independent interpretation of DAPI density as direct surrogates for DNA density across conditions, because detergent permeabilisation is known to perturb macromolecular retention, epitope accessibility, and structural organisation, even after fixation[295, 296].

While permeabilisation increased the mean-mode separation for total DAPI intensity, particularly in PNT2-C2 (where non-permeabilised mean and mode were $(5.75 \pm 0.05) \times 10^8$ a.u. and $(4.55 \pm 0.02) \times 10^8$ a.u. respectively, with the mean exceeding the mode by $\sim 25\%$, and permeabilised mean and mode were $(5.58 \pm 0.08) \times 10^8$ a.u. and $(3.79 \pm 0.06) \times 10^8$ a.u., with mean exceeding the mode by almost 50%). The approximate doubling of mean-mode separation in PNT2-C2 following permeabilisation (with LNCaP showing a similar relative increase in mean-mode separation) indicates a substantial increase in right-skew (i.e. a minority of brighter nuclei raising the mean), despite the overall trend of decreasing absolute total intensity following permeabilisation. These observations are consistent with the pronounced broadening of internuclear dispersion also observed for PNT2-C2. Conversely, for DAPI density, LNCaP mode remained within uncertainty of the mean after permeabilisation (with a mean density of $(3.44 \pm 0.03) \times 10^6$ a.u. μm^{-2} and mode of $(3.4 \pm 0.1) \times 10^6$ a.u. μm^{-2}), whereas PNT2-C2 retained a considerable mean-mode separation, with the mean exceeding the mode by $\sim 30\%$ ($(2.70 \pm 0.03) \times 10^6$ a.u. μm^{-2} vs. $(2.1 \pm 0.1) \times 10^6$ a.u. μm^{-2}). Operationally, this suggests that after permeabilisation, the typical PNT2-C2 nucleus is substantially less bright than implied by the mean, again compatible with a long-tailed distribution driven by a subset of comparatively bright nuclei.

Permeabilisation also altered the dispersion across nuclei, and in a cell line dependent manner. For LNCaP, the CV across nuclei for total intensity did not change outside of uncertainty, while the CV across nuclei for DAPI density showed a modest decrease of $\sim 15\%$ from $(33 \pm 1)\,\%$ to $(27.9 \pm 0.7)\,\%$. In contrast, PNT2-C2 showed marked increases in both CV descriptors; total intensity CV increased from $(46.2 \pm 0.8)\,\%$ to $(58 \pm 1)\,\%$ (an approximate increase of 25%), and DAPI density CV increased from $(28.0 \pm 0.4)\,\%$ to $(41.0 \pm 0.9)\,\%$ (approximately 45%). This indicates that permeabilisation introduced substantial internuclear variability in PNT2-C2, whereas LNCaP remained comparatively stable or became slightly more homogeneous under the same conditions. Such preparation-dependent changes are consistent with detergent-driven perturbations to retention and accessibility that can vary with nuclear envelope integrity, chromatin state, and fixation efficiency [295, 296]. Importantly, these changes in CV should be interpreted as reflecting the measured fluorescence variability, and are not necessarily direct measures of biological variability.

Within-nucleus (intranuclear) heterogeneity, quantified as CV_{DAPI}, decreased in LNCaP after permeabilisation (with mean decreasing from $(38.2 \pm 0.4)\,\%$ to $(34.9 \pm 0.3)\,\%$, approximately 10%, and mode from $(41 \pm 1)\,\%$ to $(29.7 \pm 0.6)\,\%$, approximately 30%). In PNT2-C2, however, the mean CV_{DAPI} increased slightly from $(27.1 \pm 0.1)\,\%$ to $(28.0 \pm 0.2)\,\%$ approximately 3%, while the mode showed an almost equal and opposite effect (decreasing from $(26.3 \pm 0.2)\,\%$ to $(25.6 \pm 0.2)\,\%$, approximately

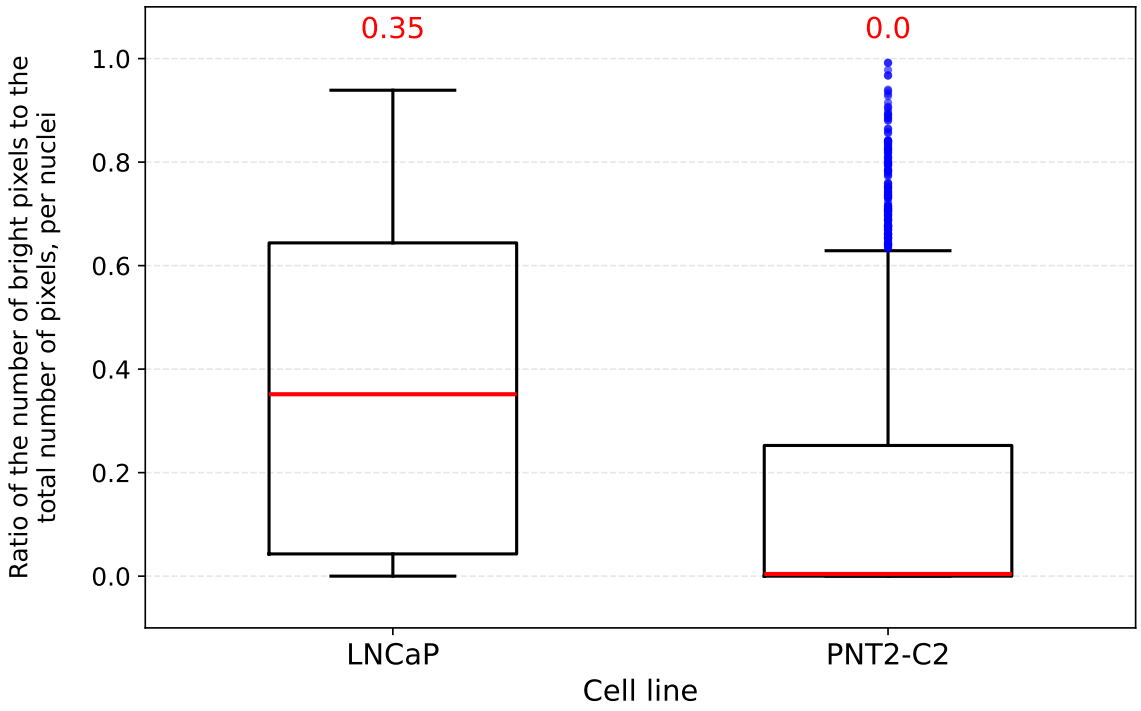
3%). Taken together with the bulk-intensity reductions, this suggests that detergent permeabilisation disproportionately reduced the prevalence of high-contrast intranuclear features in LNCaP, whereas the typical intranuclear heterogeneity of PNT2-C2 did not exhibit any considerable change, despite the strong reduction in absolute intensity. Mechanistically, DAPI fluorescence depends on minor-groove binding in A-T rich double-stranded DNA, and is highly sensitive to the local DNA environment and accessibility [8, 200, 423]. Detergent permeabilisation can alter nuclear structure and macromolecular context, and has previously been observed to introduce redistribution/extraction artefacts; as such, the observed CV_{DAPI} shifts are best interpreted as preparation-dependent changes in the measured intensity field, rather than direct evidence of altered chromatin compaction [16, 295, 447].

Bright/dark ratio box plots

As in Section 3.2.2, the bright and dark ratios (R_{bright} and R_{dark}) quantify, on a per-nucleus basis, the fraction of pixels above a fixed high-intensity threshold (bright, $I > 2.2000 \times 10^4$ a.u.) or below a fixed low-intensity threshold (dark, $I < 1.0000 \times 10^4$ a.u.). These threshold-based metrics are intrinsically sensitive to global intensity scaling and segmentation boundaries, therefore, interpretation across conditions assumes that threshold placement remains meaningful after permeabilisation, and should be contextualised using the corresponding bulk intensity and density shifts (Figure 4.7) to avoid over-interpreting threshold-driven redistributions.

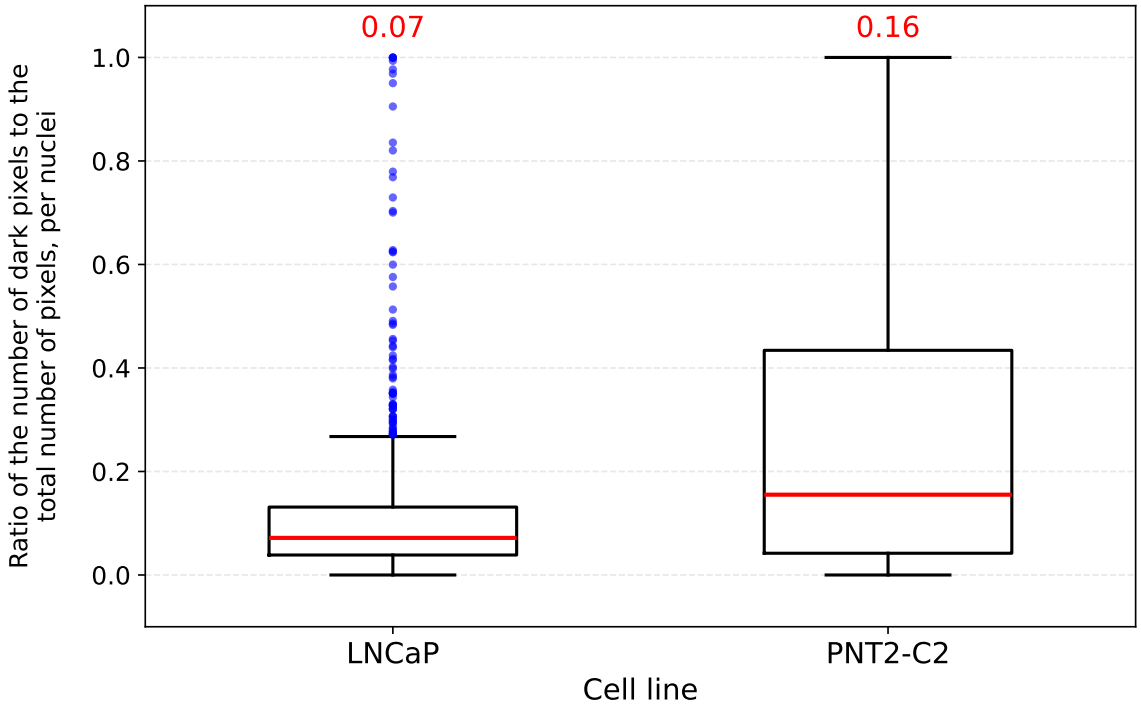
The bright/dark ratio box plots (Figure 4.8) show that permeabilisation produces a large redistribution of the intranuclear pixel intensity distribution defined by the fixed intensity thresholds used in the non-permeabilised baseline (Figure 3.15). Quantitatively, the median (IQR) R_{bright} decreased from 70.7% (42.5% – 82.8%) to 35.1% (4.3% – 64.4%) in LNCaP, and from 55.6% (20.2% – 79.1%) to 0.4% (0% – 25.2%) in PNT2-C2. In addition to the median shift, the IQR of R_{bright} increased in LNCaP (from 40.3% to 60.1%), but decreased in PNT2-C2 (from 58.9% to 25.2%), however, note that the apparent reduction for PNT2-C2 occurred alongside a collapse of the lower quartile to 0%, indicating that the permeabilised populations are dominated by low bright pixel fractions while retaining a broad upper tail. This interpretation is reinforced by the outlier structure; in permeabilised PNT2-C2 cells, number of bright ratio outliers increased drastically from 0 without permeabilisation to 138 after, despite the very low median, signifying a strongly positively skewed distribution in which most nuclei contain a negligible bright pixel fraction and a minority subpopulation retains a substantial number of bright pixels.

This was concomitant with changes in R_{dark}, with the median (IQR) increasing from 4.2% (0.9%−10.2%) to 7.1% (3.9% – 13.1%) in LNCaP, and from 0.5% (0% – 2.5%) to 15.5% (4.2% – 43.4%) in PNT2-C2. The corresponding IQR showed similar disparity between cell lines as for the bright ratio, remaining relatively stable in LNCaP (showing a negligible reduction from 9.3% to 9.2%), while showing a marked increase in PNT2-C2 (rising from 2.5% to 39.2%), demonstrating another cell line specific change; permeabilisation universally increases the prevalence of dark pixels, while substantially increasing nucleus-to-nucleus variability in specifically the non-malignant cell line. Notably, the outlier burden for R_{dark} also changed in a cell line dependent manner. In LNCaP, the number of high outliers increased from 45 to 70, consistent with a modest enrichment of nuclei exhibiting unusually large dark pixel fractions (however, note that the number of nuclei between conditions was not identical; represented as a fraction of $N_{\max}$, the outlier percentages were 6.8% and 7.4% respectively, so while permeabilisation



LNCaP: Q1=0.04, Q3=0.64, IQR=0.60, L.Out=0, H.Out=0
PNT2-C2: Q1=0.00, Q3=0.25, IQR=0.25, L.Out=0, H.Out=138

(a) Boxplot showing the ratio of the number of bright pixels to the total number of pixels in each nucleus in permeabilised LNCaP (left) and PNT2-C2 (right) cells stained with only DAPI. The median is reported in red above each plot, with the quartiles, IQR, and number of outliers at each end reported underneath. Bright pixels are defined as pixels with an intensity greater than 2.2000×10^4 a.u.



LNCaP: Q1=0.04, Q3=0.13, IQR=0.09, L.Out=0, H.Out=70
PNT2-C2: Q1=0.04, Q3=0.43, IQR=0.39, L.Out=0, H.Out=0

(b) Boxplot showing the ratio of the number of dark pixels to the total number of pixels in each nucleus in permeabilised LNCaP (left) and PNT2-C2 (right) cells stained with only DAPI. The median is reported in red above each plot, with the quartiles, IQR, and number of outliers at each end reported underneath. Dark pixels are defined as pixels with an intensity lower than 1.0000×10^4 a.u.

Figure 4.8: Box plots showing bright and dark ratios (R_{bright} and R_{dark}) of permeabilised LNCaP (left) and PNT2-C2 (right) cells stained with only DAPI. Outliers are overlaid as blue dots.

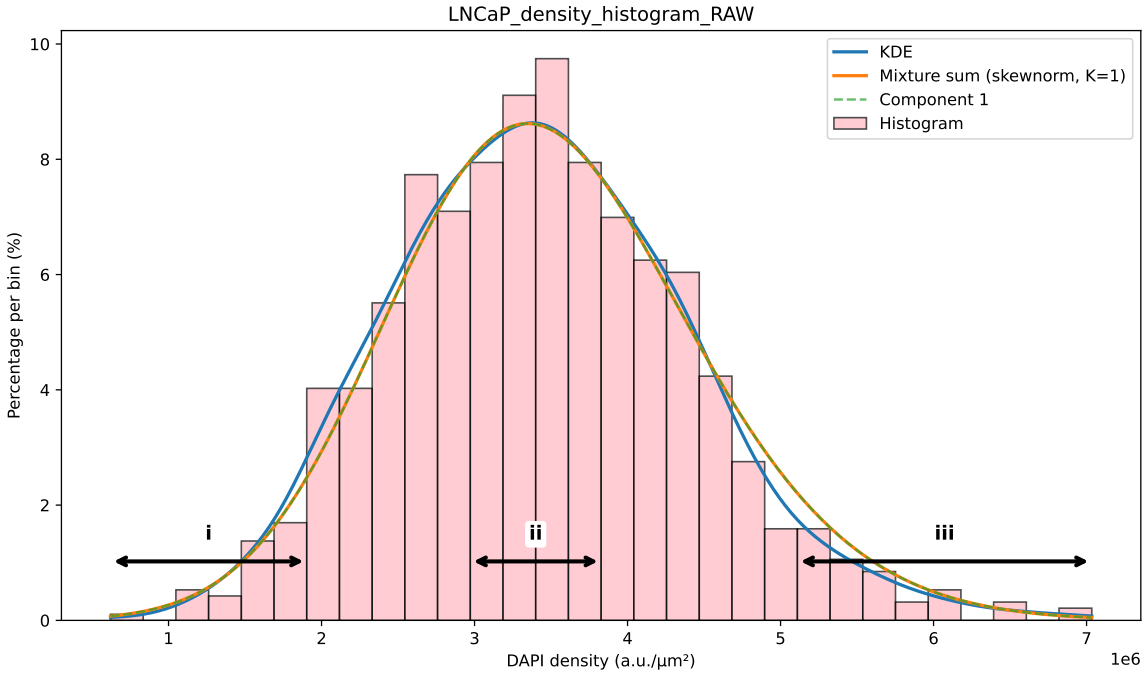
may have some impact, the observed change is much more modest when normalised to the total number of nuclei). In contrast, PNT2-C2 exhibited 239 high outliers (10.1% of $N_{\max}$) in the non-permeabilised condition, but none post-permeabilisation. This sharp reduction in the number of outliers may be driven by the large increase in both median and IQR, suggesting that the permeabilisation-associated increase in R_{dark} in PNT2-C2 reflects a broad rightward redistribution of the bulk population (i.e. many nuclei shifting towards higher dark pixel fractions), rather than being the result of a small number of extreme nuclei driving the change.

Collectively, these shifts in medians, spreads, and outlier structure indicate that permeabilisation does not simply rescale intensity, but alters the internal distribution of pixels across fixed bright and dark thresholds. The effect is qualitatively shared (reduced R_{bright}, increased R_{dark}) yet quantitatively divergent across cell lines, with PNT2-C2 showing the strongest suppression of bright pixels and the largest increase in dark-pixel variability, consistent with a pronounced, heterogeneous redistribution of intranuclear intensity following detergent exposure. In conjunction with the reduction in DAPI density (Figure 4.7), the joint decrease in R_{bright} and increase in R_{dark} strongly suggests a global downward shift in the intranuclear intensity distribution and/or increased inclusion of dim pixels to the nuclear masks during segmentation (consistent with the morphology changes observed in Section 4.1.2.1). Under fixed absolute intensity thresholds, such shifts will necessarily classify a larger proportion of pixels as "dark" and fewer as "bright", consistent with the particularly large density reduction observed in PNT2-C2. Although DAPI intensity can be modulated by DNA accessibility and chromatin environment, the threshold-based ratios used here cannot uniquely distinguish biological reorganisation from preparation induced artefacts, or changes in the accuracy of segmentation caused by obfuscation of nuclear boundaries [8, 201, 295]. As such, while it is clear that permeabilisation substantially reduces the measured high-intensity pixel burden and increases the low-intensity burden, with particularly strong effects on the non-malignant cell line, the cause of these changes cannot be isolated from this analysis alone. What can be stated without ambiguity, however, is that these preparation effects should be explicitly accounted for in any downstream comparison of chromatin-associated fluorescence between permeabilised and non-permeabilised datasets.

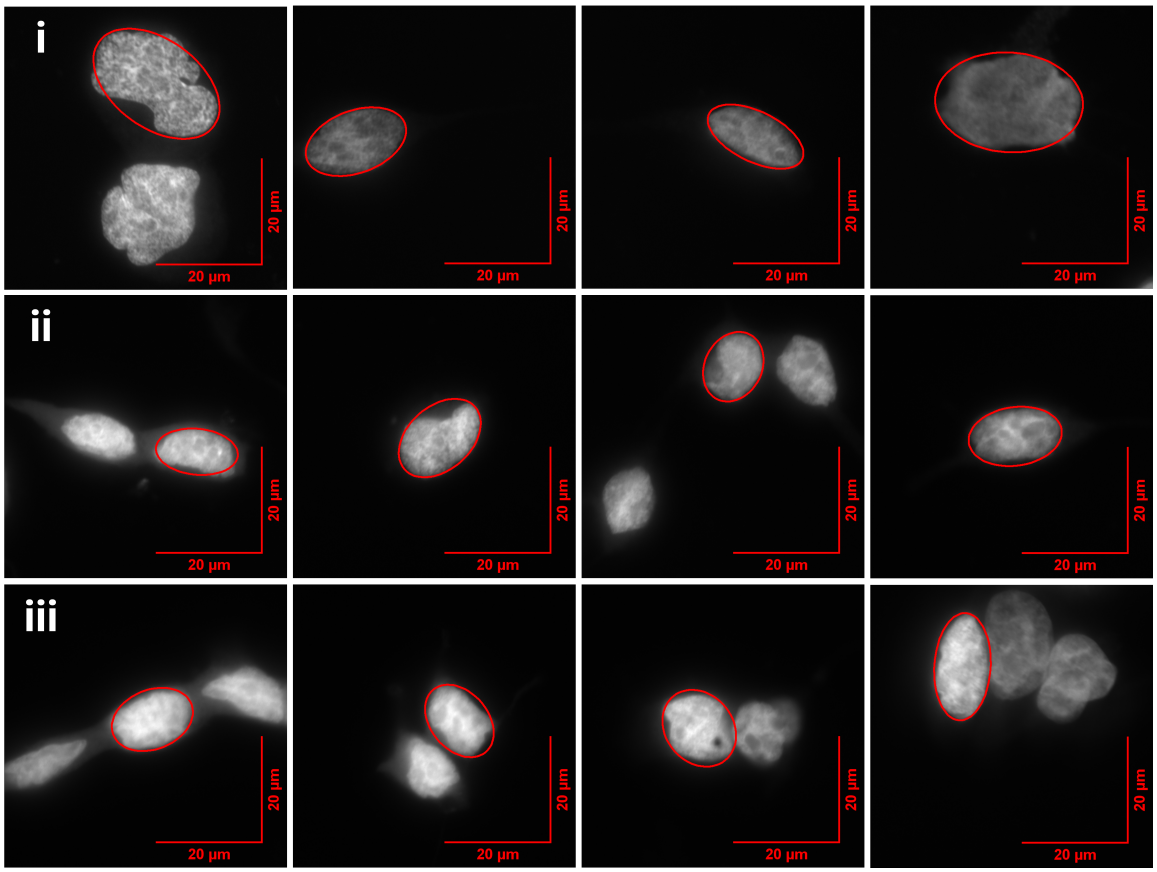
Analysis of histograms

As in the non-permeabilised baseline (Section 3.2.2), the primary motivation for analysing per-nucleus histograms is that bulk summaries can mask both inter and intranuclear heterogeneity, as well as potentially informative tail behaviour. This is particularly relevant for DAPI intensity based measures for two key reasons. Firstly, because DAPI fluorescence arises from DNA minor-groove binding, and is therefore sensitive to both DNA spatial distribution and accessibility [448], and secondly, Triton X-100 permeabilisation is a membrane-disrupting preparation step that can be associated with nuclear/chromatin perturbations [449, 450]. As such, distribution-level changes following permeabilisation can be interpreted as an operational readout of whether the preparation step modifies the population structure and/or tail fractions of the chosen intensity measures, beyond what is apparent from arithmetic means alone.

The permeabilised LNCaP density distribution shown in Figure 4.9 was best described by a single skew-normal model, with the KDE mode occurring at $(3.4 \pm 0.1) \times 10^6$ a.u. µm^{-2}. Relative to non-permeabilised LNCaP (Figure 3.16), which had a modal value of $(4.2 \pm 0.2) \times 10^6$ a.u. µm^{-2}, there was



(a) Histogram showing the DAPI density in each LNCaP nucleus. A KDE is overlaid on top of the histogram, with fitting applied to the KDE to determine the overall distribution of the data [258]. The best model (chosen by ICL [259]) was a single skew-normal distribution with a $= 1.37 \pm 0.08$, loc $= (2.6354 \pm 0.0008) \times 10^6$, and scale $= (1.3173 \pm 0.0008) \times 10^6$ with a peak at $(3.347 \pm 0.004) \times 10^6$ a.u. µm^{-2}. The KDE peak location was $(3.4 \pm 0.1) \times 10^6$ a.u. µm^{-2}.



(b) Images of detected nuclei selected from the low (i), modal (ii), and high (iii) regions of the corresponding histogram. Red ellipses are drawn purely for illustrative purposes to show the approximate region which was detected as a nucleus by the segmentation code. DAPI density was calculated as the sum of the intensity of all pixels in the nuclei mask divided by the nuclear area. A 20 µm scale bar is shown in red at the bottom right corner of each image.

Figure 4.9: Histogram showing the DAPI density in permeabilised LNCaP cells stained with only DAPI ((a)), with some images of representative nuclei from the low, modal, and high regions of the histogram ((b)). The KDE is used as a basis for fitting, with the histogram being purely illustrative. The best fit was chosen using an Integrated Classification Likelihood (ICL) [259]. Uncertainties are calculated from the covariance matrix.

an approximately 20% reduction in the modal DAPI density of nuclei. Importantly, the tail is strongly attenuated: only $\sim 1\%$ of permeabilised LNCaP nuclei had a density exceeding 6×10^6 a.u. µm^{-2}, and the 90th percentile was $\sim 4.7 \times 10^6$ a.u. µm^{-2}. In the non-permeabilised dataset, $\sim 15\%$ of nuclei had a density $>6 \times 10^6$ a.u. µm^{-2}, with a 90th percentile of $\sim 6.2 \times 10^6$ a.u. µm^{-2}. This is a significant change, an approximately 10× reduction in the extreme high-density tail constitutes strong distribution-level evidence that detergent permeabilisation either reduces the measured high-intensity pixel burden, increases the effective segmented area, or both.

In permeabilised PNT2-C2 (Figure 4.10), the density distribution was best described by a single inverse-Gaussian model, with a mode of $(2.1 \pm 0.1) \times 10^6$ a.u. µm^{-2}. Compared to non-permeabilised PNT2-C2 (Figure 3.17), the modal shift was substantially larger than in LNCaP (with an approximate 45% reduction to $(3.62 \pm 0.09) \times 10^6$ a.u. µm^{-2}). Additionally, the same reduction in tail mass is observed; $\sim 1\%$ of permeabilised PNT2-C2 nuclei had an area $>6 \times 10^6$ a.u. µm^{-2}, and the 90th percentile was $\sim 4.2 \times 10^6$ a.u. µm^{-2}, whereas non-permeabilised nuclei had $\sim 10\%$ of nuclei above 6×10^6 a.u. µm^{-2}, and a 90th percentile of $\sim 6 \times 10^6$ a.u. µm^{-2}. This strong reduction in the modal DAPI density in PNT2-C2 is consistent with the bulk-result divergence described earlier (Section 4.1.2.2): total DAPI intensity experiences relatively little change, while DAPI density decreases markedly, implying a dominant contribution from increased area (i.e. nuclear expansion or segmentation dilation) rather than a proportional loss in total signal.

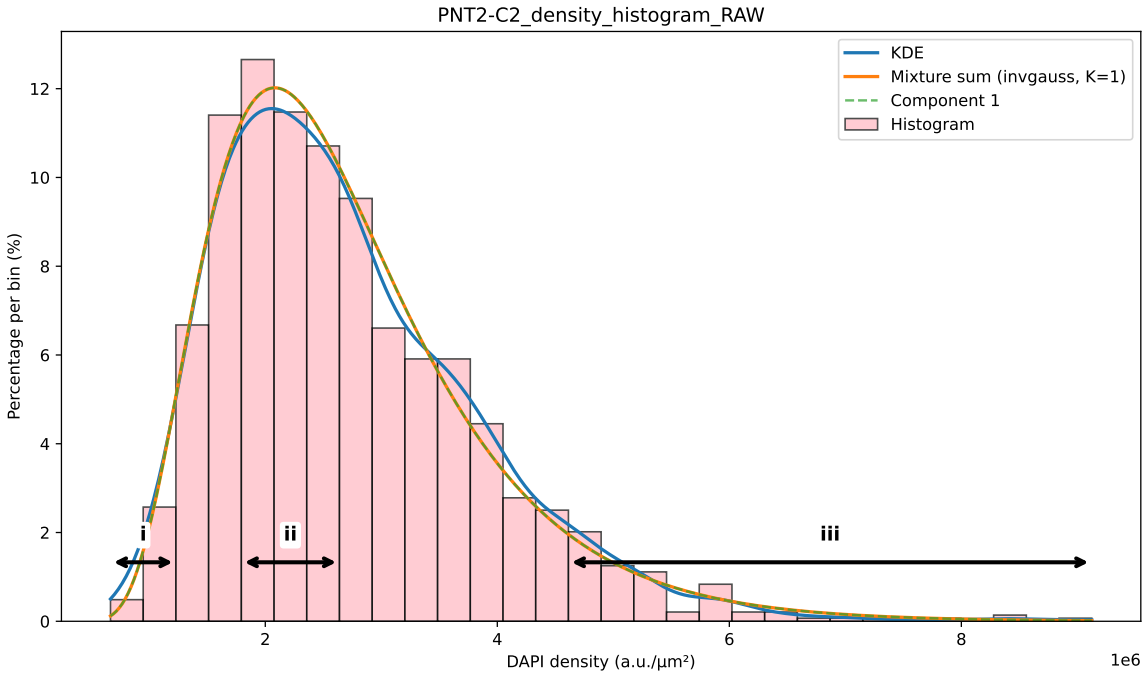
Following detergent permeabilisation, LNCaP retained a higher density mode than PNT2-C2 ($(3.4 \pm 0.1) \times 10^6$ a.u. µm^{-2} vs. $(2.1 \pm 0.1) \times 10^6$ a.u. µm^{-2} respectively), showing the non-permeabilised ordering not only persists, but is accentuated as a result of permeabilisation (with the difference in mode between cell lines rising from $\sim 15\%$ in to $\sim 65\%$). This is consistent with the malignant/aneuploid status of LNCaP, and documented subpopulation structure, but the present analysis adds a new preparation-dependent nuance; permeabilisation suppresses the extreme high-density tail in both cell lines, while simultaneously widening the separation in their modal densities, indicating that "tail fraction" and "mode separation" can respond differently to the same chemical perturbation [14, 155].

Histograms of per-nucleus DAPI intensity CV (Figures 4.11 and 4.12) provide a distributional measure of intranuclear texture/heterogeneity, defined as

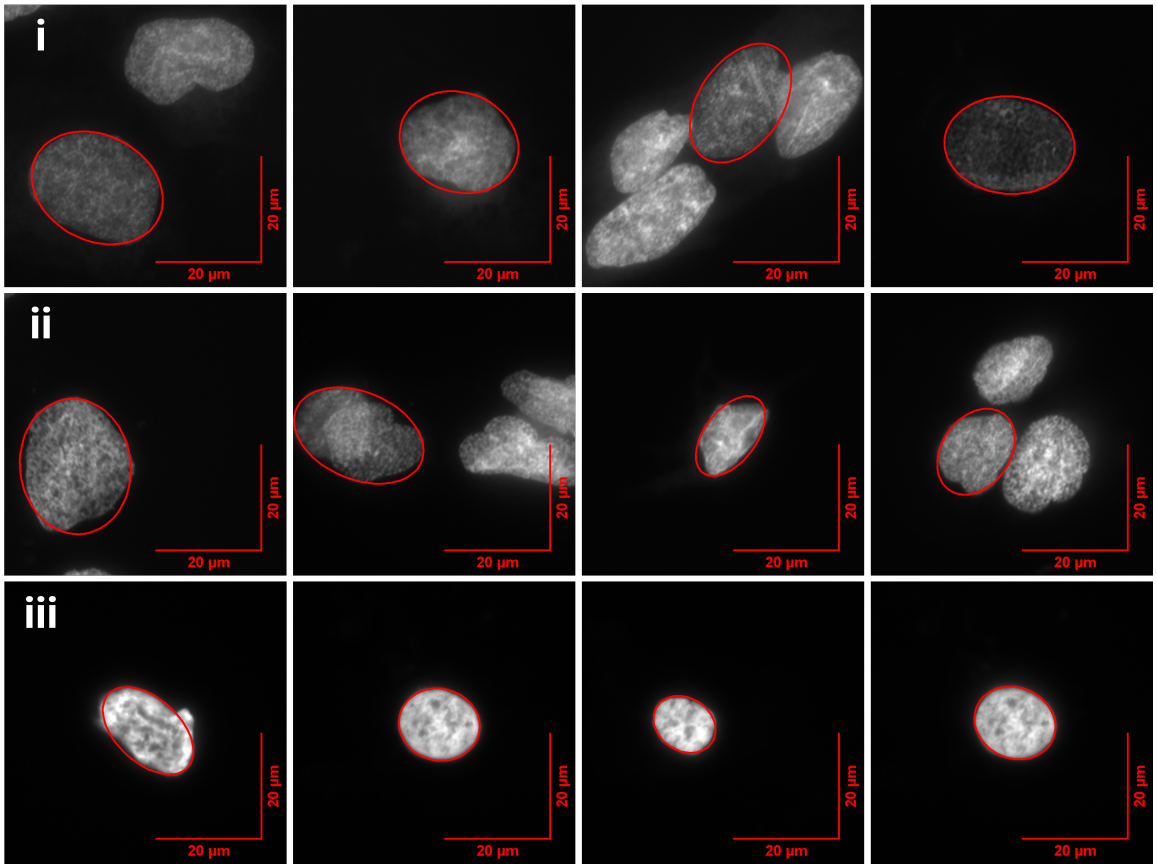
$$\mathrm{CV}_{\mathrm{DAPI}} = \frac{\sigma(I)}{\mu(I)} \times 100,$$

where I denotes the set of pixel intensities within a given segmented nucleus. While $\mathrm{CV}_{\mathrm{DAPI}}$ is not uniquely specific to chromatin compaction, it is an operationally useful metric for detecting whether nuclei contain strongly coexisting bright and dark subregions, i.e. high-contrast texture [134, 208].

Permeabilised LNCaP $\mathrm{CV}_{\mathrm{DAPI}}$ remained bimodal in the parametric characterisation, with a two component Gaussian fit best describing the data, however the distribution was shifted such that the previously dominant peak (iii in Figures 3.18 and 4.11) which encompassed approximately 80% of the data in the non-permeabilised dataset reduced in weighting to $\sim 55\%$, with a corresponding increase in the complementary Gaussian component (ii) from $\sim 20\%$ to $\sim 45\%$, yielding near parity in the fitted component contributions. Consistent with this, the rightmost component showed only a marginal reduction in fitted amplitude alongside a slight narrowing, whereas the amplitude of the leftmost

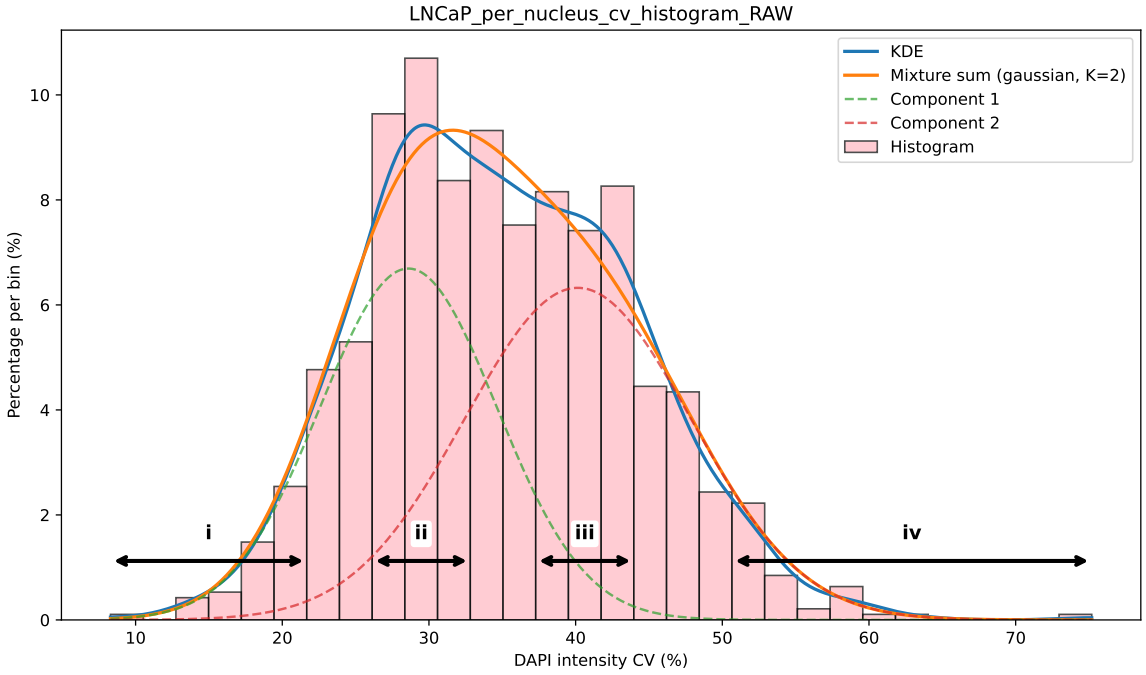


(a) Histogram showing the DAPI density in each PNT2-C2 nucleus. A KDE is overlaid on top of the histogram, with fitting applied to the KDE to determine the overall distribution of the data [258]. The best model (chosen by ICL [259]) was a single inverse Gaussian distribution with $\mu = 0.178 \pm 0.002$, loc $= 0$, and scale $= (1.525\,490 \pm 0.000\,002) \times 10^7$, with a peak at $(2.08 \pm 0.02) \times 10^6$ a.u. μm^{-2}. The KDE peak location was $(2.1 \pm 0.1) \times 10^6$ a.u. μm^{-2}.

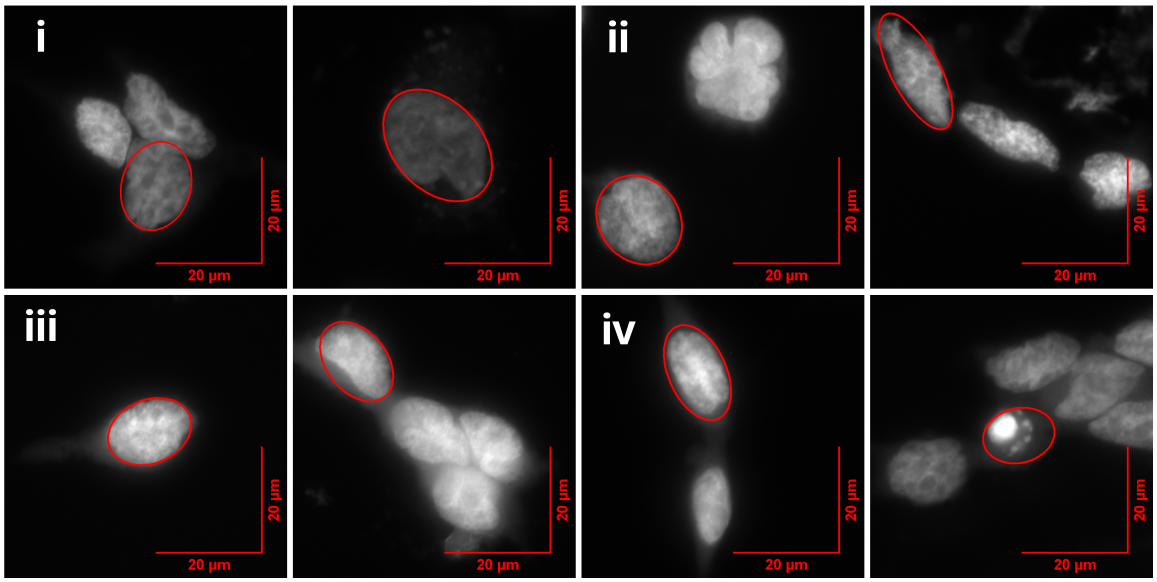


(b) Images of detected nuclei selected from the low (i), modal (ii), and high (iii) regions of the corresponding histogram. Red ellipses are drawn purely for illustrative purposes to show the approximate region which was detected as a nucleus by the segmentation code. DAPI density was calculated as the sum of the intensity of all pixels in the nuclei mask divided by the nuclear area. A 20 μm scale bar is shown in red at the bottom right corner of each image.

Figure 4.10: Histogram showing the DAPI density in permeabilised PNT2-C2 cells stained with only DAPI ((a)), with some images of representative nuclei from the low, modal, and high regions of the histogram ((b)). The KDE is used as a basis for fitting, with the histogram being purely illustrative. The best fit was chosen using an Integrated Classification Likelihood (ICL) [259]. Uncertainties are calculated from the covariance matrix.



(a) Histogram showing the CV of DAPI intensity in each LNCaP nucleus. A KDE is overlaid on top of the histogram, with fitting applied to the KDE to determine the overall distribution of the data [258]. The best model (chosen by ICL [259]) was a Gaussian distribution with two components, the first had $\mu = 28.6 \pm 0.3$ and $\sigma = 6.0 \pm 0.2$, with a peak at $(28.6 \pm 0.3)\,\%$, the second had $\mu = 40.1 \pm 0.3$ and $\sigma = 7.7 \pm 0.2$, with a peak at $(40.1 \pm 0.3)\,\%$. The KDE peak location was $(29.7 \pm 0.6)\,\%$.



(b) Images of detected nuclei selected from the low (i), modal/first subpopulation (ii), second subpopulation (iii), and high (iv) regions of the corresponding histogram. Red ellipses are drawn purely for illustrative purposes to show the approximate region which was detected as a nucleus by the segmentation code. CV was calculated as 100 times the standard deviation of all pixel intensities in the nuclei mask divided by the mean. A 20 µm scale bar is shown in red at the bottom right corner of each image.

Figure 4.11: Histogram showing the CV of DAPI intensity in permeabilised LNCaP cells stained with only DAPI ((a)), with some images of representative nuclei from the low, modal/first subpopulation, second subpopulation, and high regions of the histogram ((b)). The KDE is used as a basis for fitting, with the histogram being purely illustrative. The best fit was chosen using an Integrated Classification Likelihood (ICL) [259]. Uncertainties are calculated from the covariance matrix.

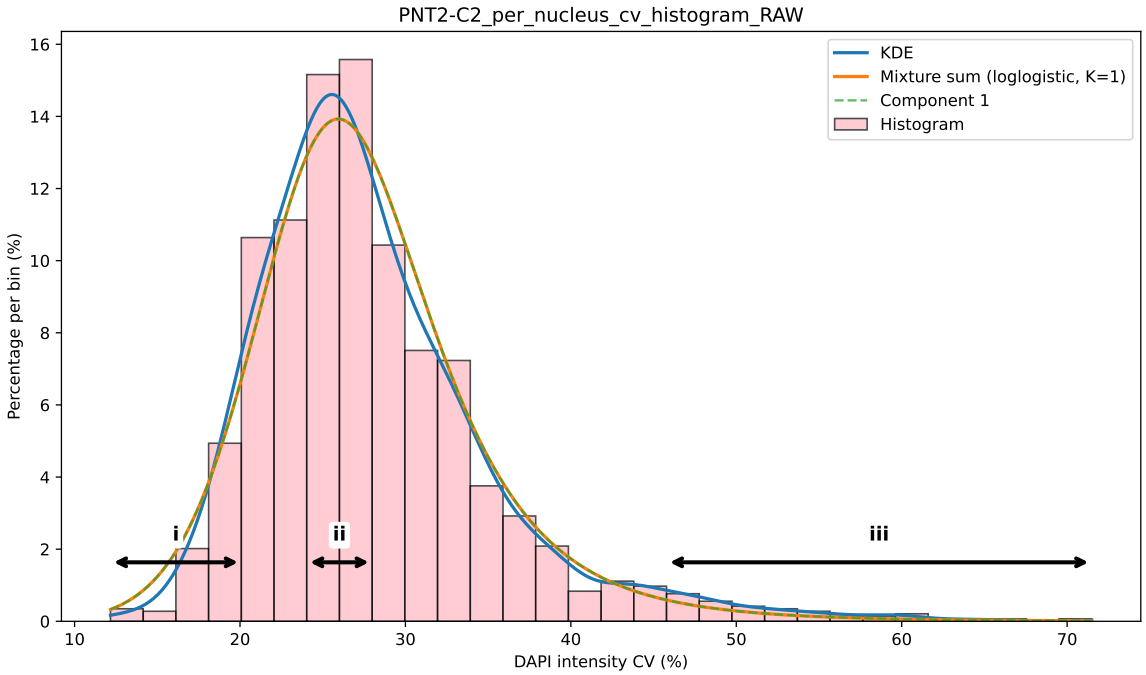
component increased substantially, such that the two became comparable in prominence. The positions of each component remained largely unchanged following permeabilisation, with the leftmost peak showing the largest shift from $(24.3 \pm 0.4)\,\%$ to $(28.6 \pm 0.3)\,\%$, a relative increase of $\sim 20\%$, and the rightmost peak decreasing from $(41.4 \pm 0.4)\,\%$ to $(40.1 \pm 0.3)\,\%$, a $\sim 3\%$ relative reduction. While the change in the position of the leftmost peak may seem substantial in relative terms, this is influenced by lower starting value; an absolute increase of $\sim 4\%$ is small when considering the distribution as a whole. Regardless, the combined effects of permeabilisation served to swap the position of the KDE mode from closely reflecting the rightmost peak in non-permeabilised cells (with a KDE mode of $(41 \pm 1)\,\%$), to instead match the position of the leftmost peak in permeabilised cells (with KDE mode $= (29.7 \pm 0.6)\,\%$), a relative decrease of $\sim 30\%$ in the modal intranuclear heterogeneity. Collectively, these changes represent a significant enrichment of the lower-CV subpopulation in LNCaP cells.

Another key distribution-level change is observed in the tail; $\sim 30\%$ of permeabilised LNCaP nuclei had $\mathrm{CV}_{\mathrm{DAPI}} > 40\%$, with a 90th percentile of 46.8%, whereas approximately 45% of nuclei in the non-permeabilised baseline had a DAPI intensity CV above 40%, and a 90th percentile of 52.1%. This indicates that detergent permeabilisation reduces the proportion of strongly texture-rich/high-contrast nuclei in LNCaP, consistent with the increased prominence of the low $\mathrm{CV}_{\mathrm{DAPI}}$ subpopulation.

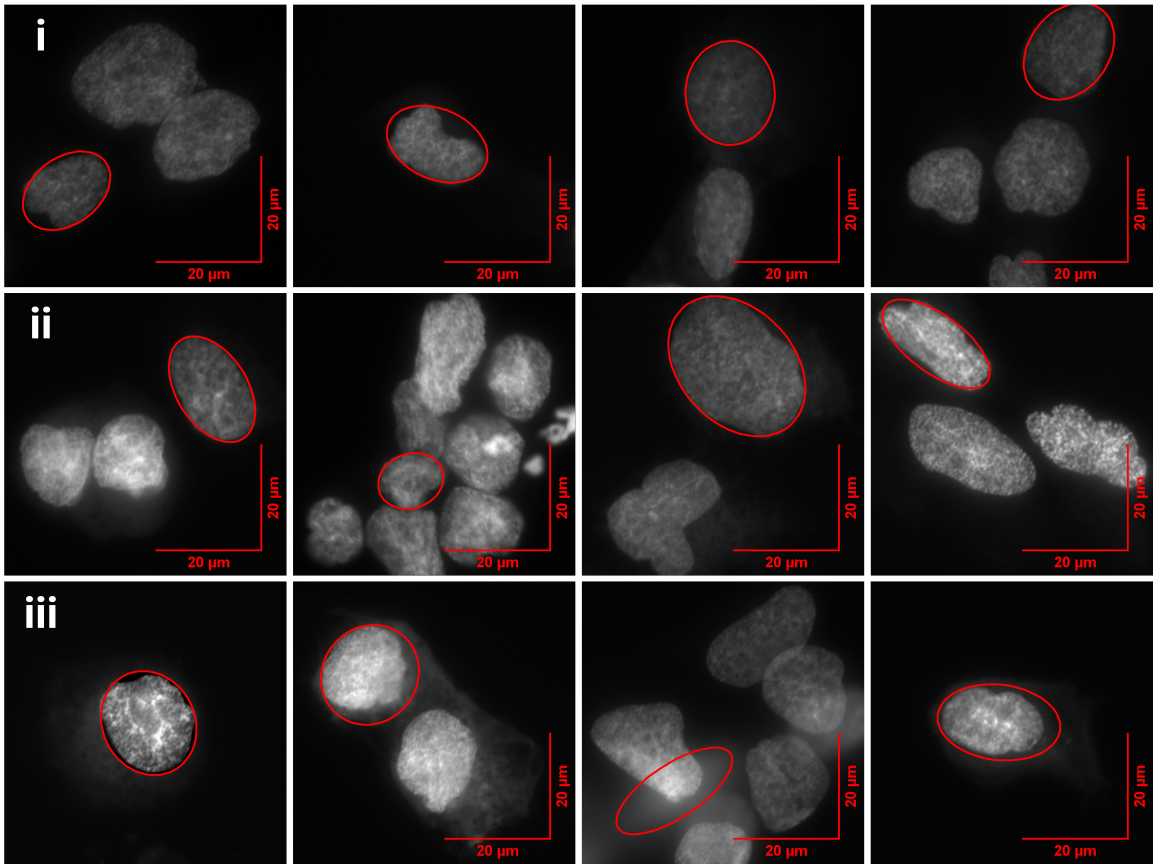
For PNT2-C2, the $\mathrm{CV}_{\mathrm{DAPI}}$ distribution was well-described by a single log-logistic model both pre and post permeabilisation (Figures 3.19 and 4.12). The modal value of $(25.6 \pm 0.2)\,\%$ changed only very slightly compared to the non-permeabilised baseline ($(26.3 \pm 0.2)\,\%$), a relative decrease of $\sim 2.5\%$. Despite this, there was a notable increase in tail mass following permeabilisation, with $\sim 6\%$ of permeabilised PNT2-C2 nuclei with $\mathrm{CV}_{\mathrm{DAPI}} > 40\%$, compared to only $\sim 2\%$ in the non-permeabilised condition. The 90th percentile showed a similar trend, rising from 33.4% in non-permeabilised cells to 36.7% following permeabilisation. This indicates that permeabilisation acts to introduce the proportion of highly heterogeneous detected PNT2-C2 nuclei, without substantially affecting the "typical" nucleus, i.e. a primarily tail-driven effect.

In the non-permeabilised condition, the separation between LNCaP and PNT2-C2 $\mathrm{CV}_{\mathrm{DAPI}}$ modes is large ($\sim 40\%$ in LNCaP vs. $\sim 25\%$ in PNT2-C2), consistent with malignancy-associated chromatin texture changes reported in the wider prostate cancer literature [425, 451]. Following permeabilisation, this separation narrows substantially (with KDE peaks of $\sim 30\%$ in LNCaP vs. $\sim 25\%$ in PNT2-C2), indicating that the preparation step minimises the cell line difference in modal intranuclear heterogeneity. The present study therefore provides evidence that careful consideration of sample preparation steps should be applied to the interpretation of DAPI-derived measures of chromatin texture; the magnitude of the difference between malignant and healthy $\mathrm{CV}_{\mathrm{DAPI}}$ explicitly depends on whether detergent permeabilisation is applied, even when everything else remains unchanged between conditions.

The distribution-level changes detailed above are consistent with the bright/dark ratio box plots reported previously (Figure 4.8). Specifically, permeabilisation acts to approximately half the median bright-pixel fraction in LNCaP (decreasing from 71% to 35%), and drives the PNT2-C2 median from 56% all the way to 0%, indicating that after permeabilisation, most PNT2-C2 nuclei do not contain any (or a negligible amount) of pixels above the fixed "bright" threshold of 22000 a.u. In parallel, the dark-pixel fraction increased from 4% to 7% in LNCaP, and from 1% to 16% in PNT2-C2. The redistribution of pixels away from the bright field provides a concrete, within-nucleus corroboration for both the strong suppression of the high-density tail in both cell lines (Figures 4.9 and 4.10), and the



(a) Histogram showing the CV of DAPI intensity in each PNT2-C2 nucleus. A KDE is overlaid on top of the histogram, with fitting applied to the KDE to determine the overall distribution of the data [258]. The best model (chosen by ICL [259]) was a single log-logistic distribution with $c = 7.4 \pm 0.2$, $\text{loc} = 0$, and $\text{scale} = 26.9 \pm 0.2$, with a peak at $(25.9 \pm 0.2)\,\%$. The KDE peak location was $(25.6 \pm 0.2)\,\%$.



(b) Images of detected nuclei selected from the low (i), modal (ii), and high (iii) regions of the corresponding histogram. Red ellipses are drawn purely for illustrative purposes to show the approximate region which was detected as a nucleus by the segmentation code. CV was calculated as 100 times the standard deviation of all pixel intensities in the nuclei mask divided by the mean. A 20 µm scale bar is shown in red at the bottom right corner of each image.

Figure 4.12: Histogram showing the CV of DAPI intensity in permeabilised PNT2-C2 cells stained with only DAPI ((a)), with some images of representative nuclei from the low, modal, and high regions of the histogram ((b)). The KDE is used as a basis for fitting, with the histogram being purely illustrative. The best fit was chosen using an Integrated Classification Likelihood (ICL) [259]. Uncertainties are calculated from the covariance matrix.

differing response in DAPI intensity CV between cell lines; LNCaP loses a large fraction of extremely high-contrast nuclei, while PNT2-C2 experiences a small increase in the proportion of nuclei with mixed intensity fields, capable of driving an increase in "tailedness".

While prostate cancer is known to exhibit altered chromatin organisation and heterogeneity, comparatively few studies explicitly quantify (at the single nucleus resolution) how a common immunofluorescence preparation step such as detergent permeabilisation reshapes the entire distribution of DAPI-derived density and texture measures within both malignant and non-malignant prostate cell models. The results presented above demonstrate that permeabilisation by exposure to Triton X-100 (as used in this pipeline) strongly suppresses extreme high-density nuclei in both malignant and normal cell lines, but has a divergent effect on intranuclear heterogeneity tails, emphasising the sentiment echoed throughout Section 4.1, that sample preparation steps can induce profound changes in not only mean signal levels, but also the prevalence of rare, high-heterogeneity subpopulations, that are invisible to mean-based reporting.

4.1.3 Interim conclusions

This section of Chapter 4 established a quantitatively controlled, permeabilised DAPI-only reference condition that isolates preparation-induced effects from marker-induced effects, thereby providing the necessary control baseline for subsequent interpretation of data utilising the additional nucleolar stain. Convergence testing confirmed that all reported descriptors retained ideal SEM scaling ($pow \approx -1/2$; Section 4.1.1), and that the available nucleus counts remain sufficient under the operational convergence criterion of %SEM $< 2.5\%$. That is not to say permeabilisation had no effect on the convergence of these metrics, however. Permeabilisation materially reshaped descriptor stability in a cell line dependent manner; in LNCaP, sampling requirements decreased for nuclear area ($N_{\text{conv}} : 365 \rightarrow 340$, $\sim 10\%$) and DAPI density ($190 \rightarrow 140$, $\sim 25\%$), whereas in PNT2-C2, N_{conv} increased for all descriptors, particularly strongly in DAPI density ($145 \rightarrow 300$, $\sim 105\%$). Critically, the relative stability ordering for intensity-normalised descriptors was inverted following permeabilisation (with PNT2-C2 requiring substantially larger N than LNCaP for DAPI density), demonstrating that detergent treatment cannot be treated as a neutral preparatory step for intensity-linked measures.

Considering the measured values directly, permeabilisation produced large, non-uniform shifts in nuclear morphology and DAPI-derived intensity structure. The most prominent morphological outcome was a reversal in nuclear area ordering; LNCaP mean nuclear area decreased by $\sim 10\%$ ($(195 \pm 3)\,\mu m^2 \rightarrow (179 \pm 2)\,\mu m^2$), whereas PNT2-C2 increased by $\sim 55\%$ ($(136 \pm 1)\,\mu m^2 \rightarrow (211 \pm 2)\,\mu m^2$), resulting in PNT2-C2 nuclei being $\sim 20\%$ larger than LNCaP post-permeabilisation (Section 4.1.2.1). Distributional modelling further indicated heavier-tailed area behaviour after permeabilisation (being best described by log-logistic fits rather than log-normals), consistent with increased probability mass at large areas (Section 4.1.2.1). In terms of intensity measures, both cell lines exhibited reduced DAPI density means and modes, but resulting from different dominant causes; LNCaP showed a substantial reduction in total DAPI intensity ($\sim 25\%$) with a comparatively modest inferred area change, whereas PNT2-C2 showed minimal change in the total intensity ($\sim -2.5\%$) but a large density reduction ($\sim 35\%$), consistent with a dominant increase in effective segmented area (Section 4.1.2.2). Across both lines, permeabilisation strongly suppressed extreme high-density tails (only $\sim 1\%$ of nuclei above 6×10^6 a.u. μm^{-2}, compared to $\sim 10 - 15\%$ pre-permeabilisation; Figures 4.9 and 4.10), and reshaped intranuclear texture distributions such that the modal CV_{DAPI} separation between malignant

and non-malignant models was reduced (with LNCaP exhibiting a marked enrichment of the lower CV_{DAPI} subpopulation, and PNT2-C2 showing a primarily tail-driven increase in high CV_{DAPI} nuclei; Figures 4.11 and 4.12). Collectively, these findings provide direct, distribution-level evidence that permeabilisation can attenuate, invert, or selectively reweight apparent cell line differences in both morphology and DAPI-derived intensity/texture.

Practically, this permeabilised DAPI-only characterisation of commonly used prostate cancer models constitutes new, preparation-specific knowledge, that must condition the interpretation of the subsequent anti-fibrillarin analyses. In the subsequent section, DAPI-derived morphology, density, and texture must therefore be interpreted with this in mind, rather than treating them as direct measures of DNA content and chromatin architecture. The DAPI-only permeabilised results thus provide the necessary reference against which marker-specific changes in nuclear prevalence, size, and organisation can be separated from detergent-driven redistribution, segmentation-boundary shifts, and tail suppression effects introduced prior to immunolabelling.

4.2 DAPI and anti-fibrillarin analysis

This section extends the permeabilised DAPI-only analysis (Section 4.1) to the full immunolabelling workflow (DAPI + anti-fibrillarin, AF488) with two linked objectives. First, the same DAPI-derived morphology and intensity descriptors are briefly re-evaluated under dual-stain conditions to test whether antibody incubation and associated handling steps introduce any additional, statistically meaningful perturbation beyond the permeabilisation effect already characterised. Second, and more importantly, anti-fibrillarin labelling is used to quantify nucleolar content directly and comparatively between the malignant LNCaP model and the non-malignant PNT2-C2 model.

Fibrillarin is a nucleolar protein localised predominantly to the dense fibrillar component (DFC) and is widely used as an immunofluorescent marker of nucleoli [17, 219, 417, 429]. Accordingly, in this work nucleoli were operationally defined as discrete intranuclear regions positive for anti-fibrillarin immunofluorescence, with the nuclear compartment defined by DAPI-based segmentation. This definition is consistent with the established subnucleolar organisation (fibrillar centre, DFC, granular component), and with routine approaches in which fibrillarin signal provides both a spatial identifier for nucleoli and a quantitative readout for nucleolus-associated fluorescence [139, 452, 453]. Note, however, that while nucleoli are referred to for brevity, fibrillarin is preferentially localised within the DFC, and as such all nucleolar measures reported here may best be interpreted as DFC-associated nucleolar measures rather than as analogous representations of the entire nucleolar footprint. Additionally, under strong nucleolar stress, fibrillarin relocalisation can alter apparent nucleolar morphology, such that changes in fibrillarin associated area or intensity may reflect reorganisation as well as changes in nucleolar size [224].

On this basis, the central hypothesis is that LNCaP exhibits increased nucleolar prevalence and/or size relative to PNT2-C2, measurable as shifts in fibrillarin-positive nucleolar area and intensity-derived metrics. This hypothesis is motivated by the well established observation that increased nucleolar size and number are features of prostate cancer (and malignancy more broadly), commonly interpreted as reflecting increased ribosomal biogenesis and protein synthesis demand [18, 72, 73, 138, 212, 454].

Throughout this section, the shorthand "FIB" is used to denote fibrillarin-related signals and descriptors

(e.g. I_{FIB} as anti-fibrillarin AF488 intensity). In addition to absolute nucleolar measures, ratio descriptors are also used to normalise inter-sample variability and link nucleolar signal to nuclear context, including DAPI:FIB intensity and nucleolar:nuclear area ratios per nucleus. This work follows on from, and provides an orthogonal validation of, previous Raman spectroscopic studies, which reported a protein-rich intranuclear component in LNCaP and attributed this to nucleolar remodelling; by explicitly segmenting and quantifying nucleoli via FIB labelling, the present analysis tests this mechanistic interpretation directly [127].

4.2.1 Convergence testing

As previously (Sections 3.1 and 4.1.1), convergence was assessed using random repeated subsampling to estimate the standard error of the mean (SEM) as a function of the number of nuclei, N Sections 2.3.4 and 2.3.6, with SEM decay modelled using a power law of the form Equation (2.37), and γ calculated using Equation (2.38). N_{conv} was defined using the same percentage standard error criterion as throughout (Equation (2.39)).

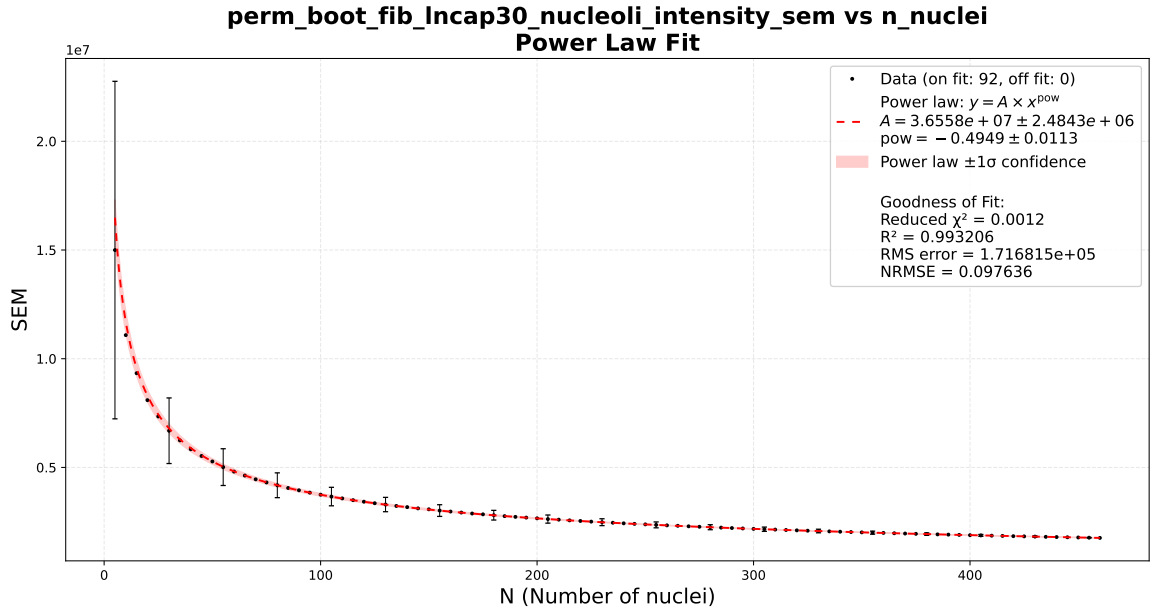
Following the addition of the FIB channel, two additional convergence measures were introduced to support downstream nucleolar analysis: the background-suppressed nucleolar intensity I_{nucleoli} and the nucleoli:nucleus area ratio $R_{\mathrm{N:NA}}$, defined as described in Equations (2.21) and (2.22). These were included because they provide, respectively, an intensity-based and a morphology-normalised nucleolar readout specific to the DAPI+FIB workflow. Importantly, these fluorescence-derived descriptors are mechanistically closer to nucleolar protein localisation and concentration than to classical histological nucleolar prominence, which is known to correlate with ribosome biogenesis and proliferation, but is not sub-compartment specific [63, 136, 430]. One set of example convergence plots for the new nucleoli intensity metric are shown in Figure 4.13.

The fitted *pow* exponents extracted from the repeated random subsampling SEM fits are summarised in Figure 4.14. *pow* is expected to remain close to $-1/2$ under the central limit framework, indicating that the SEM decreases approximately as $N^{-1/2}$ as additional nuclei are included. For each descriptor and cell line, *pow* is shown for two fitting approaches: the primary weighted fit performed in linear space, and the corresponding linear regression performed in log space. The two estimates are included as an internal consistency check on exponent stability with respect to the fitting domain; agreement between the solid (linear-space) and hatched (log-space) bars supports that the observed scaling is not an artefact of the specific regression representation.

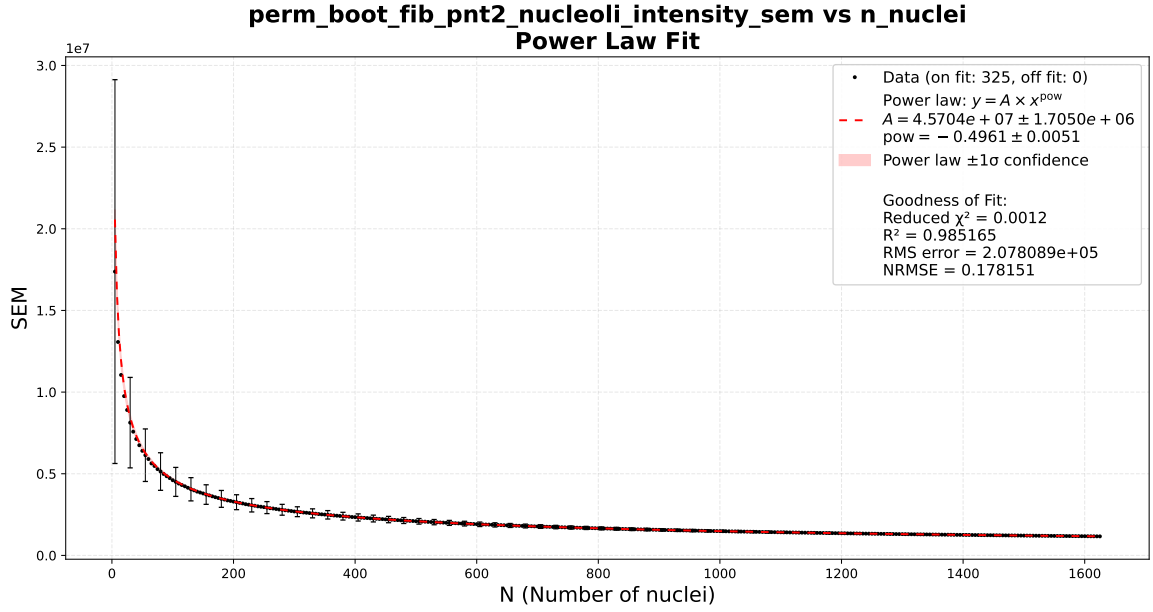
Metrics were chosen as described previously (Section 3.1.2), such that the primary differences between chosen descriptors are expressed not through systematic departures in *pow*, but through the convergence rate, γ, and the resulting sampling requirement to reach the convergence criterion of %SEM $<$ 2.5%. These fit parameters, together with goodness-of-fit statistics and the discrete convergence point, N_{conv}, are consolidated in Table 4.2 to provide a basis for interpreting both previously used DAPI-derived descriptors, as well as the newly introduced nucleolar measures prior to downstream biological comparison. Additional convergence data are presented in Section B.3.

4.2.1.1 Key quantitative outcomes

Across all descriptors, fitted exponents remained close to $pow = -0.5$(Table 4.2), supporting continued use of the SEM(N) $\propto N^{-1/2}$ framework outlined previously (Sections 3.1 and 4.1.1) under the dual-stain



(a) The average SEM of nucleoli intensity in permeabilised LNCaP cells stained with DAPI and anti-fibrillarin (DAPI+FIB), calculated from 10000 repeated random subsamples plotted against N. A power law fit is applied, with fit parameters of $A = (3.7 \pm 0.2) \times 10^7$, and $pow = -0.49 \pm 0.01$. Goodness of fit metrics were NRMSE = 0.0976, $\chi^2_\nu = 0.001\,22$, and $R^2 = 0.993$.



(b) The average SEM of nucleoli intensity in permeabilised PNT2-C2 cells stained with DAPI and anti-fibrillarin (DAPI+FIB), calculated from 10000 repeated random subsamples plotted against N. A power law fit is applied, with fit parameters of $A = (4.6 \pm 0.2) \times 10^7$, and $pow = -0.496 \pm 0.005$. Goodness of fit metrics were NRMSE = 0.178, $\chi^2_\nu = 0.001\,18$, and $R^2 = 0.985$.

Figure 4.13: Graphs showing the standard error of the mean (SEM) convergence of LNCaP ((a)) and PNT2-C2 ((b)) nucleoli intensity as a function of N, the number of nuclei included in the calculation. Error bars are drawn every five points, and indicate the resampling standard deviation of each datapoint. NRMSE refers to normalised root mean square error (normalised to the terminal SEM), χ^2_ν refers to reduced chi squared, and R^2 refers to the coefficient of determination. Fits use weighted least squares regression with inverse variance weighting applied, and uncertainties were calculated using covariance aware error propagation Equation (2.40).

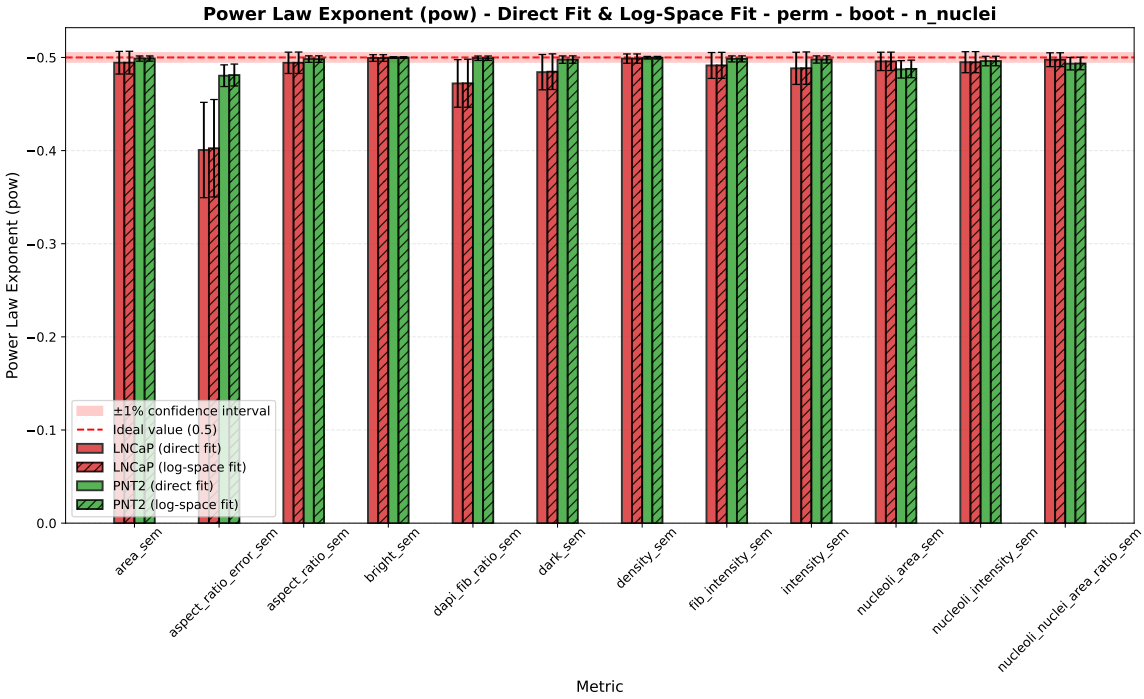


Figure 4.14: *pow* values from the repeated random subsampling power law fit for SEM in linear space (solid fill), and the linear fit in log space (hatched fill) for each metric and cell line. Cells were permeabilised, then stained with DAPI and anti-fibrillarin (DAPI+FIB). LNCaP is shown in red, and PNT2-C2 in green. The red dotted line indicates the ideal value of -0.5, with the shaded region representing the allowed 1% variation.

workflow. Crucially, the newly introduced nucleolar measures also satisfied this criterion: I_{nucleoli} yielded $pow = -0.49 \pm 0.01$ for LNCaP, and -0.496 ± 0.005 for PNT2-C2, while $R_{N:NA}$ yielded $pow = -0.498 \pm 0.007$ for LNCaP, and -0.493 ± 0.007 for PNT2-C2. The greatest observed deviation from ideal *pow* scaling was seen in LNCaP $R_{N:NA}$, however, this still lay only $\sim 2.5\%$ away from the ideal value of -0.5, well within the $\pm 5\%$ tolerance band adopted for practical equivalence. As such, the dominant limitation for nucleolar descriptors is not exponent instability, but rather the absolute sampling requirement needed to reach the $\%\text{SEM} < 2.5\%$ criterion.

For the DAPI-derived legacy descriptors (area, aspect ratio, and DAPI density), convergence points remained below a maximum of $N_{\text{conv}} = 345$ nuclei in both cell lines (Table 4.2). More specifically, compared to the DAPI-only values, N_{conv} in LNCaP increased by $\sim 1\%$ (the smallest possible step) for area, $\sim 20\%$ for aspect ratio, and $\sim 35\%$ for DAPI density. For PNT2-C2, an approximately 10% increase in N_{conv} was observed for area, a 5% decrease for aspect ratio, and a $\sim 15\%$ decrease for DAPI density. Importantly, the relative cell line ordering remained broadly intact when compared to the DAPI-only analysis (Table 4.1). No significant changes were anticipated following the addition of FIB staining, and so the changes observed here (which range from negligible to modest in absolute sampling terms) are attributed to a combination of expected inter-experimental variance, and the additional sample preparation required for immunostaining (such as extended incubation and repeated washing steps), rather than a biologically driven perturbation of the DAPI channel. More explicitly, the absence of a consistent shift in direction across descriptors and cell lines supports this interpretation. Conversely, the nucleolar intensity descriptors required substantially larger N to reach convergence, with the required increases varying in a cell line dependent manner. For I_{nucleoli}, LNCaP required 455 nuclei to reach convergence, essentially saturating the available sample size ($N_{\max} = 460$). PNT2-C2, on the other hand, required 795 nuclei to reach the same convergence threshold, but retained a comfortable

Cell line	Metric	γ	*pow*	NRMSE	χ^2_ν	N_{conv}
LNCaP ($N_{\max} = 460$)	Area	0.0115 ± 0.0008	-0.49 ± 0.01	0.119	0.001 46	345
	Aspect ratio	1.5 ± 0.1	-0.49 ± 0.01	0.121	0.001 55	235
	DAPI density	$(7.7 \pm 0.2) \times 10^{-7}$	-0.499 ± 0.005	0.0498	6.29×10^{-4}	190
	Nucleoli intensity	$(2.7 \pm 0.2) \times 10^{-8}$	-0.49 ± 0.01	0.0976	0.001 22	455
	Nucleoli Nuclei area ratio	27 ± 1	-0.498 ± 0.007	0.0870	0.001 11	230
PNT2-C2 ($N_{\max} = 1627$)	Area	0.0107 ± 0.0002	-0.499 ± 0.003	0.119	7.06×10^{-4}	345
	Aspect ratio	3.44 ± 0.09	-0.498 ± 0.003	0.136	7.99×10^{-4}	95.0
	DAPI density	$(6.34 \pm 0.06) \times 10^{-7}$	-0.500 ± 0.001	0.0535	2.96×10^{-4}	250
	Nucleoli intensity	$(2.19 \pm 0.08) \times 10^{-8}$	-0.496 ± 0.005	0.178	0.001 18	795
	Nucleoli Nuclei area ratio	9.3 ± 0.5	-0.493 ± 0.007	0.352	0.002 78	1050

Table 4.2: Summary of convergence analysis for permeabilised LNCaP and PNT2-C2 cells stained with DAPI and anti-fibrillarin (DAPI+FIB). Power law fit parameters and goodness of fit statistics refer to weighted fits performed to the mean SEM calculated through repeated random subsampling as a function of N, where γ is the rate of convergence (defined as $\gamma = \frac{1}{A}$), and *pow* is the fitted exponent. The convergence point, N_{conv}, is defined by the earliest sampled N for which the percentage standard error remains below 2.5% for all subsequent samples (reported as a discrete value without uncertainty). NRMSE refers to normalised root mean square error, normalised to terminal SEM, and χ^2_ν refers to reduced chi squared. $N_{\max}$ is the total number of nuclei available in each cell line.

margin relative to $N_{\max} = 1627$. This implies that the estimate of mean nucleolar signal in LNCaP is, by construction, close to the stability threshold under the chosen criterion, and therefore should be interpreted more cautiously in subsequent cell-line comparisons (particularly when the observed effects are small). Considered alongside the significantly larger number of nuclei required for nucleoli intensity convergence in PNT2-C2, the comparatively larger NRMSE (0.178 in PNT2-C2 vs 0.0976 in LNCaP) further suggests increased residual structure about the fitted scaling law, consistent with greater heterogeneity and/or heavier-tailed variability in nucleolar signal once background-suppressed intensities are integrated across nuclei.

A similar, more pronounced stability asymmetry was observed for the nucleoli:nuclei area ratio. In LNCaP, $R_{N:NA}$ converged at $N_{\text{conv}} = 230$, comparable to (and in some cases lower than) the DAPI-derived descriptors, supporting its use as a robust, normalised nucleolar morphology measure. In contrast, PNT2-C2 required $N_{\text{conv}} = 1045$, with markedly poorer fit-normalised error (NRMSE = 0.352 vs. 0.087) and elevated χ^2_ν relative to the non-malignant cell line (Table 4.2). This may indicate that although $R_{N:NA}$ is an appropriate and more stable alternative to absolute nucleolar area, meaningful estimation of PNT2-C2 nucleolar occupancy requires substantially larger sampling than for LNCaP, consistent with a broader and/or more outlier prone distribution of nucleolar segmentation outcomes in the non-malignant model under this workflow.

Taken together, these results justify the inclusion of I_{nucleoli} and $R_{N:NA}$ as the primary additional convergence tested nucleolar measures for the remainder of the section, while also providing clear guidance for the relative confidence placed on downstream interpretation. Specifically, where comparisons involve I_{nucleoli} in LNCaP, small effects should be interpreted carefully given the proximity of N_{conv} to N_{max}.

4.2.2 Morphological and intensity analysis

4.2.2.1 Morphology

As in the permeabilised DAPI-only reference condition (Section 4.1.2.1), nuclear morphology was quantified using the same DAPI-derived descriptors defined previously in Methods, namely projected nuclear area and nuclear aspect ratio (Equations (2.7) and (2.8)). The key purpose of re-evaluating these DAPI-derived morphology metrics under the DAPI+FIB workflow is not to re-establish nuclear pleomorphism (which was treated in detail previously), but rather to confirm that the immunolabelling workflow does not drastically change the DAPI-based morphology interpretation, and provide appropriate nuclear context for interpreting nucleolar morphology.

The central new contribution of the DAPI+FIB dataset here is the introduction of two nucleolar morphology measures derived from the anti-fibrillarin (FIB) channel. In this section, nucleolar area is defined as the total area of pixels in the FIB channel above the second Otsu threshold, within the nuclei mask defined by the DAPI channel. A normalised descriptor is additionally introduced as the nucleoli:nuclei area ratio, $R_{N:NA}$, defined previously (Section 4.2.1), and is interpreted as a nucleolar occupancy measure that captures how much of the nuclear footprint can be attributed to nucleoli. This is expected to be better aligned with the expected observations of nucleolar prominence than absolute nucleolar area alone, since nuclear size can vary substantially both biologically and due to preparation/segmentation effects [16, 63, 72, 131].

Bulk summaries of nuclear morphology under DAPI+FIB are shown in Figure 4.15, which is retained as the detailed within-condition reference for the DAPI+FIB dataset itself, including the exact mean, mode, and CV for nuclear area and aspect ratio. To make the preparation dependence easier to assess directly, paired comparison figures (Figures 4.16 and 4.17) collate the corresponding DAPI-derived morphology results across all three analytical conditions: non-permeabilised DAPI, permeabilised DAPI-only, and permeabilised DAPI+FIB. Throughout, the plotted mode is the peak position of the KDE, consistent with the histogram analyses in Sections 3.2.1, 4.1.2.1 and 4.2.2.1, and error bars represent standard error (calculated using the delta method where required; Equations (2.28) and (2.29)).

The key result from Figure 4.16 is that the apparent nuclear-size ordering is not preparation-invariant. In the non-permeabilised baseline, LNCaP nuclei were larger than PNT2-C2 by approximately 45% in mean area and $\sim 30\%$ in modal area (Figure 3.8). After permeabilisation, this ordering reversed: in the DAPI-only condition, nuclear area in PNT2-C2 exceeded that in LNCaP by approximately 20% in both mean and mode (Figure 4.2). Under the DAPI+FIB workflow, PNT2-C2 nuclei remained larger, but the gap narrowed to approximately 7% in the mean and 5% in the mode (Figure 4.15). Relative to the permeabilised DAPI-only reference, this narrowing was driven primarily by an upward shift in LNCaP nuclear area under DAPI+FIB, from $(179 \pm 2)\,\mu m^2$ to $(202 \pm 4)\,\mu m^2$ in the mean (an increase of $\sim 13\%$), and from $(152 \pm 3)\,\mu m^2$ to $(180 \pm 10)\,\mu m^2$ in the mode (an increase of $\sim 15\%$), whereas the mean

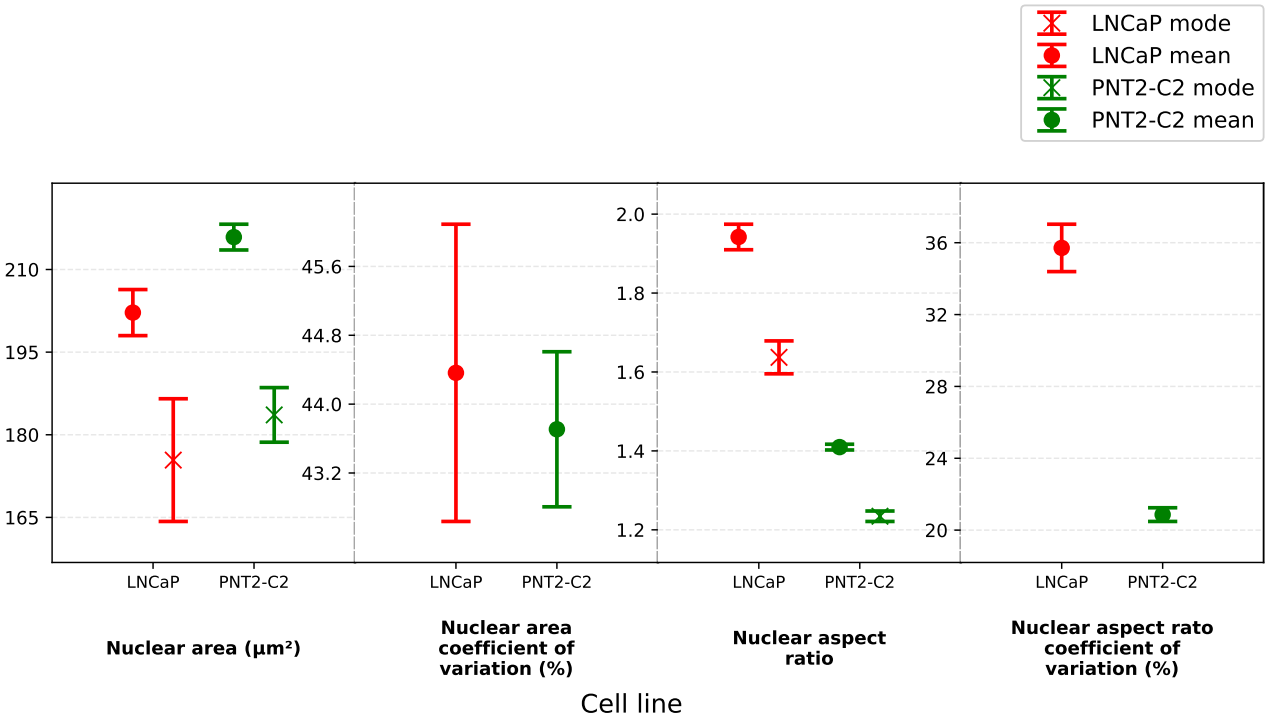
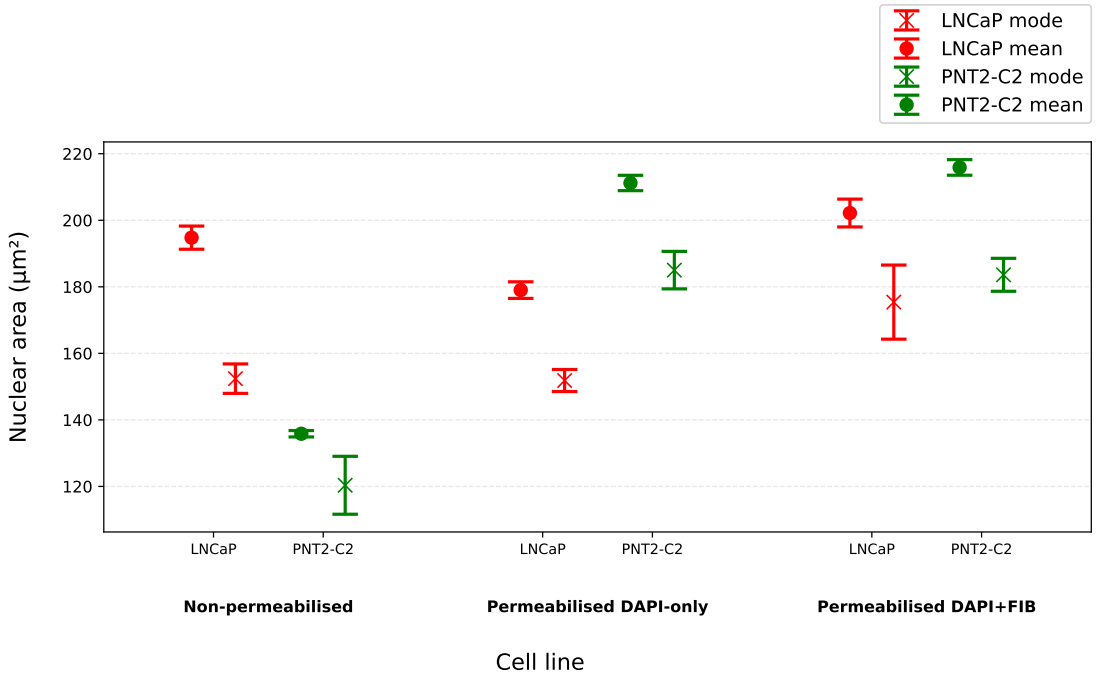


Figure 4.15: Plot showing the mean, mode and coefficient of variation (CV) of nuclear area and nuclear aspect ratio in permeabilised cells stained with DAPI and anti-fibrillarin (DAPI+FIB). Mode here refers to the peak position of the KDE (Equation (2.34)), consistent with the distributional analyses in Section 4.2.2.1. Error bars represent standard error (calculated using the delta-method where required for CV and mode estimates; Equations (2.28) and (2.29)). LNCaP nuclei had a mean area of $(202 \pm 4)\,\mu m^2$, modal area of $(180 \pm 10)\,\mu m^2$, area CV of $(44 \pm 2)\,\%$, mean aspect ratio of 1.94 ± 0.03, modal aspect ratio of 1.64 ± 0.04, and aspect ratio CV of $(36 \pm 1)\,\%$, while PNT2-C2 nuclei had a mean area of $(216 \pm 2)\,\mu m^2$, modal area of $(184 \pm 5)\,\mu m^2$, area CV of $(43.7 \pm 0.9)\,\%$, mean aspect ratio of 1.410 ± 0.007, modal aspect ratio of 1.23 ± 0.01, and aspect ratio CV of $(20.9 \pm 0.4)\,\%$.

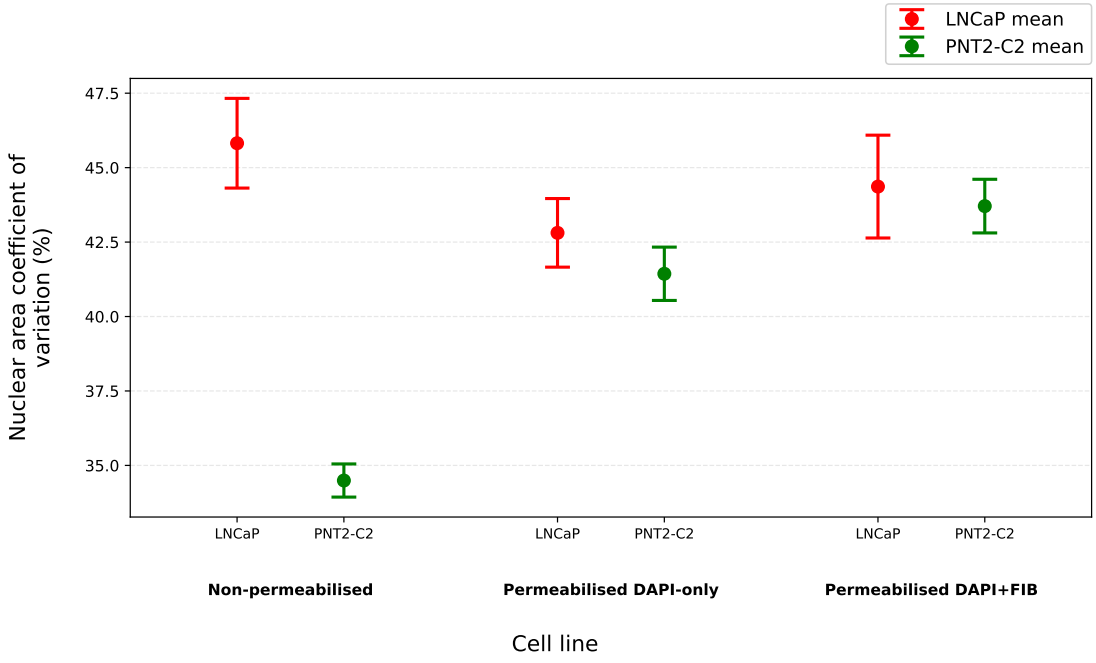
PNT2-C2 nuclear area only just differed outside of uncertainty, and modal area remained unchanged outside of uncertainty Figure 4.16. As such, Figure 4.16 makes the central point particularly clear: the perturbation resulting from detergent permeabilisation was sufficient to dominate the apparent biological separation in nuclear area, inverting the expected baseline ordering, while the additional FIB staining procedure partially relaxed this inversion, but did not restore the non-permeabilised pattern.

The corresponding area CV comparison shown in Figure 4.16 reinforces this interpretation; in the non-permeabilised state, LNCaP showed substantially greater relative area dispersion than PNT2-C2 ($(46 \pm 2)\,\%$ vs. $(34.5 \pm 0.6)\,\%$, a relative increase of $\sim 25\%$) Figure 3.8. After permeabilisation, however, the two cell lines became effectively commensurate in area CV ($(43 \pm 1)\,\%$ in LNCaP vs. $(41.4 \pm 0.9)\,\%$ in PNT2-C2; Figure 4.2), and remained so under DAPI+FIB ($(44 \pm 2)\,\%$ vs. $(43.7 \pm 0.9)\,\%$; Figure 4.15). More explicitly, the effect of permeabilisation on nuclear area is therefore not limited to a shift in central tendency alone; it also erodes observed differences in the relative heterogeneity in area between the malignant and non-malignant models (in this case by increasing the variance in PNT2-C2 nuclei).

The aspect ratio comparison shown in Figure 4.17 reveals a markedly different behaviour. Unlike nuclear area, nuclear aspect ratio ordering was stable across all three preparations: LNCaP remained the more elongated model throughout. In the non-permeabilised baseline (Figure 3.8), aspect ratio in LNCaP nuclei exceeded that in PNT2-C2 by approximately 20% in mean and 10% in mode. In the permeabilised DAPI-only condition (Figure 4.2), the corresponding differences were approximately 17%

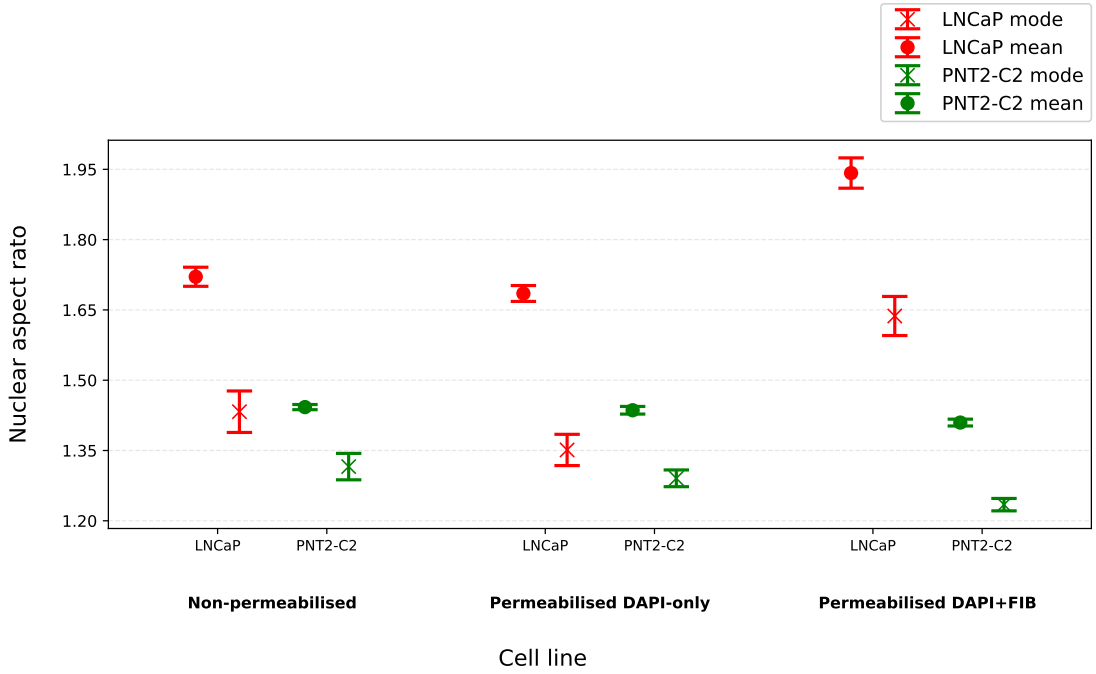


(a) Nuclear area.

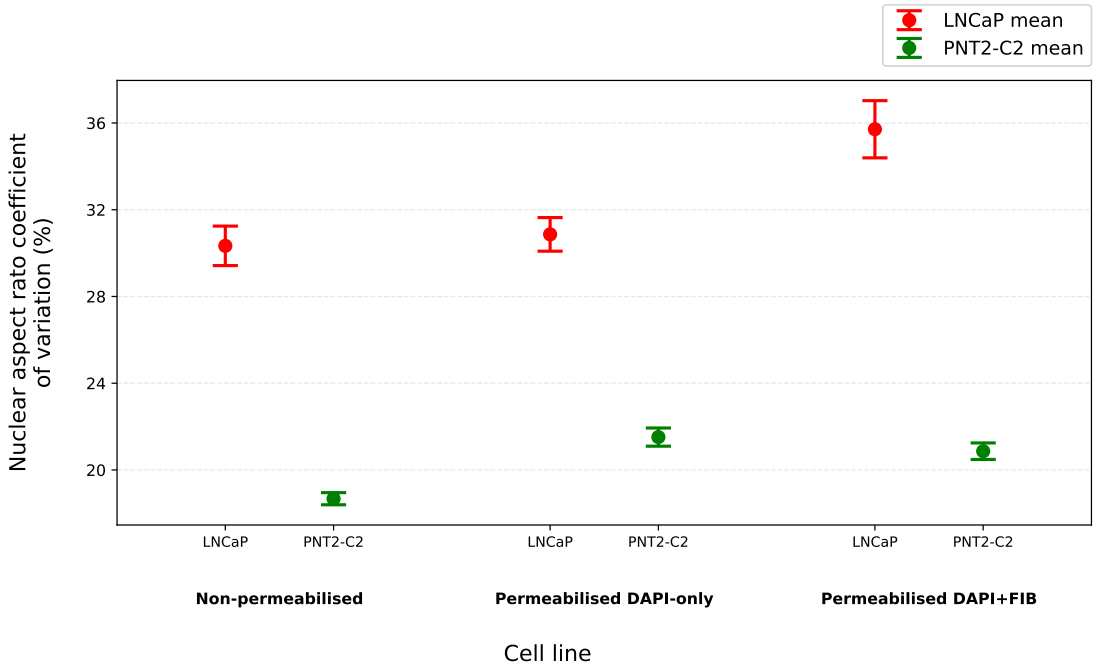


(b) Nuclear area CV across nuclei.

Figure 4.16: Cross-condition comparison of DAPI-derived nuclear area descriptors across non-permeabilised DAPI, permeabilised DAPI-only, and permeabilised DAPI+FIB preparations. Mean and mode summarise the same pooled nucleus-level distributions examined in Sections 3.2.1, 4.1.2.1 and 4.2.2.1, with the mode defined as the KDE peak location (Equation (2.34)). Error bars represent standard error.



(a) Nuclear aspect ratio.



(b) Nuclear aspect ratio CV across nuclei.

Figure 4.17: Cross-condition comparison of DAPI-derived nuclear aspect ratio descriptors across non-permeabilised DAPI, permeabilised DAPI-only, and permeabilised DAPI+FIB preparations. Mean and mode summarise the same pooled nucleus-level distributions examined in Sections 3.2.1, 4.1.2.1 and 4.2.2.1, with the mode defined as the KDE peak location (Equation (2.34)). Error bars represent standard error.

and 5%. Under DAPI+FIB, however, the separation between cell lines increased substantially, with LNCaP nuclei exceeding those in PNT2-C2 by approximately 40% in the mean aspect ratio, and 35% in the mode (Figure 4.15). This increase was driven almost entirely by LNCaP, whose mean aspect ratio rose from 1.69 ± 0.02 in the DAPI-only condition to 1.94 ± 0.03 under DAPI+FIB (an increase of $\sim 15\%$), while PNT2-C2 changed only slightly and in the opposite direction (from 1.436 ± 0.008 to 1.410 ± 0.007, a decrease of $\sim 2\%$; Figures 4.2 and 4.15). The practical consequence is that while nuclear area is strongly preparation-dependent, and can even reverse its between-line ordering following permeabilisation, nuclear elongation remains directionally robust, and was in this specific instance accentuated by the dual-stain workflow Figure 4.17(a).

Aspect ratio CV showed the same qualitative stability Figure 4.17(b). LNCaP nuclei consistently exhibited a higher shape dispersion than PNT2-C2 nuclei in all three conditions: $(30.3 \pm 0.9)\,\%$ vs. $(18.7 \pm 0.3)\,\%$ before permeabilisation (Figure 3.8), $(30.9 \pm 0.8)\,\%$ vs. $(21.5 \pm 0.4)\,\%$ after permeabilisation (Figure 4.2), and $(36 \pm 1)\,\%$ vs. $(20.9 \pm 0.4)\,\%$ under DAPI+FIB (Figure 4.15). Expressed in relative terms, the LNCaP aspect ratio CV exceeded that of PNT2-C2 by approximately 60%, 45%, and 70% in the three conditions respectively Figure 4.17(b). This is scientifically useful because it shows that one of the most biologically relevant findings in the present study, the association of the malignant model with increased nuclear elongation and greater shape heterogeneity, is not an artefact of one specific preparation branch. Rather, it persists across all three analytical conditions, even when the apparent nuclear-size ordering does not.

At the same time, the comparison plots should not be treated in isolation from the distributional analyses. Where the mean and modal measures diverge, and particularly in cases where the high-end tails are heavy, some contribution from segmentation edge cases or other imaging-related failures remains plausible alongside genuine biological heterogeneity, and as such, the paired comparison shown in Figures 4.16 and 4.17 is intended as a concise summary of preparation dependence, while the histograms and kernel density estimates provide detailed distribution-level interpretation.

With that DAPI-derived cross-condition context established, the nucleolar morphology measures can now be considered directly.

Turning to the nucleolar measures, a key and somewhat unexpected bulk finding is that absolute nucleolar area is essentially indistinguishable between LNCaP and PNT2-C2 within uncertainty. Specifically, LNCaP and PNT2-C2 have almost identical mean nucleolar areas ($(18.5 \pm 0.4)\,\mu m^2$ vs. $(18.6 \pm 0.3)\,\mu m^2$), and KDE modes which show slightly greater separation, but are still well within uncertainty of one another ($(15.2 \pm 0.7)\,\mu m^2$ vs. $(15.3 \pm 0.5)\,\mu m^2$). This is especially notable because nucleolar enlargement and prominence are widely described hallmarks of prostate cancer, and nucleolar alterations are mechanistically linked to oncogenic drivers (such as MYC) that are active in prostate neoplasia [17, 18, 63, 136]. The present dataset therefore suggests that either the anticipated difference in nucleolar size is weak/absent under these culture and preparation conditions, or the current operational definition is not fully sensitive to the structural nucleolar changes that drive nucleolar prominence.

In contrast to absolute nucleolar area, the nucleolar occupancy measure $R_{N:NA}$ showed a clear difference between cell lines at the level of central tendency. LNCaP exhibited a higher mean $R_{N:NA}$ than PNT2-C2 (0.093 ± 0.001 vs. 0.0869 ± 0.0009), corresponding to a relative increase of $\sim 5\%$ with respect to the PNT2-C2 mean. The modes are separated by an even greater margin (0.093 ± 0.002 vs. 0.084 ± 0.002,

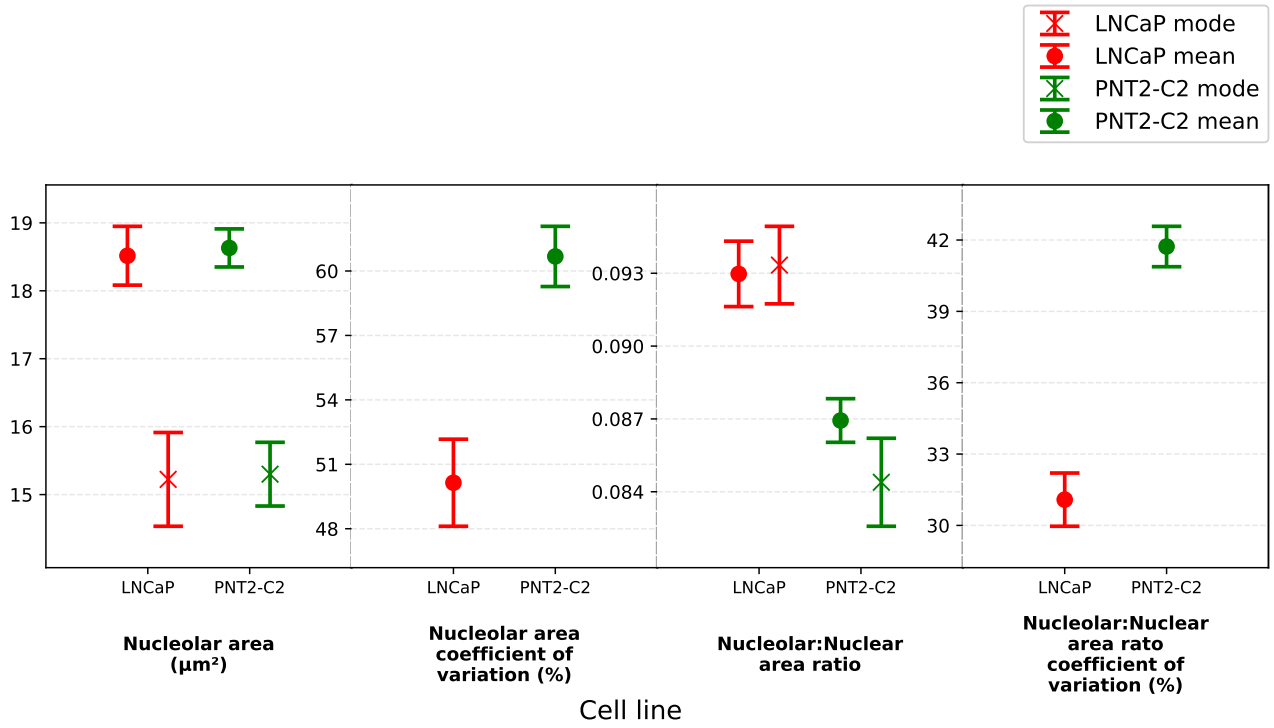


Figure 4.18: Plot showing the mean, mode and coefficient of variation (CV) of nucleolar area and nucleolar:nuclear area ratio ($R_{N:NA}$) in permeabilised cells stained with DAPI and anti-fibrillarin (DAPI+FIB). Mode here refers to the peak position of the KDE (Equation (2.34)), consistent with the distributional analyses in Section 4.2.2.1. Error bars represent standard error (calculated using the delta-method where required for CV and mode estimates; Equations (2.28) and (2.29)). LNCaP nucleoli had a mean area of $(18.5 \pm 0.4)\,\mu m^2$, modal area of $(15.2 \pm 0.7)\,\mu m^2$, area CV of $(50 \pm 2)\,\%$, mean $R_{N:NA}$ of 0.093 ± 0.001, modal $R_{N:NA}$ of 0.093 ± 0.002, and $R_{N:NA}$ CV of $(31 \pm 1)\,\%$, while PNT2-C2 nucleoli had a mean area of $(18.6 \pm 0.3)\,\mu m^2$, modal area of $(15.3 \pm 0.5)\,\mu m^2$, area CV of $(61 \pm 1)\,\%$, mean $R_{N:NA}$ of 0.0869 ± 0.0009, modal $R_{N:NA}$ of 0.084 ± 0.002, and $R_{N:NA}$ CV of $(41.7 \pm 0.8)\,\%$.

a $\sim 10\%$ modal increase in LNCaP). By definition, this pattern is consistent with the combination of near equal nucleolar area and smaller nuclear area in LNCaP, yielding increased nucleolar occupancy despite the absence of a detectable absolute nucleolar area increase. Interpreted biologically, this is compatible with LNCaP exhibiting a modestly increased relative nucleolar investment, consistent with the general link between malignant growth programmes and increased nucleolar activity/ribosome biogenesis demand [138, 141].

An additional distinction lies in the heterogeneity; PNT2-C2 shows substantially higher dispersion than LNCaP in both nucleolar descriptors. Nucleolar area CV increased from $(50 \pm 2)\,\%$ in LNCaP to $(61 \pm 1)\,\%$ in PNT2-C2, and $R_{N:NA}$ CV increased from $(31 \pm 1)\,\%$ to $(41.7 \pm 0.8)\,\%$. This indicates that the non-malignant PNT2-C2 model (which is SV40-immortalised rather than truly normal epithelium) exhibits a broader spread of nucleolar morphological states on a per-nucleus basis. Whether the increased dispersion reflects genuine biological heterogeneity or increased susceptibility to segmentation artefacts is addressed further in the histogram based analyses below.

Distributional analysis using histograms

As previously, nucleus-level histograms were used to interrogate distributional shape, explicitly separating central tendency (KDE mode) from tail behaviour (mean-mode separation, high-end threshold fractions, and 90th percentiles). Mean-mode separation is treated as a direct quantitative

marker of the tail leverage; when $\bar{x} > x_{\text{mode}}$, a small number of high-end events must be present with sufficient magnitude to shift the mean above the most probable value. Note that the number of bins for DAPI based morphology histograms here was chosen to mirror that from previous sections, however, as N_{max} was reduced in DAPI+FIB LNCaP, 50 bins was slightly excessive in this case. As the histogram is purely visual (with all analysis and fitting based on the KDE), the decision was made to keep a consistent level of binning so that histograms are directly comparable and equivalent in terms of the y-axis. As in the earlier histogram analyses, all distributions were checked by traceback to individual nuclei during interpretation, however in this section representative image panels are shown only where visual inspection adds substantial interpretive value to the discussion.

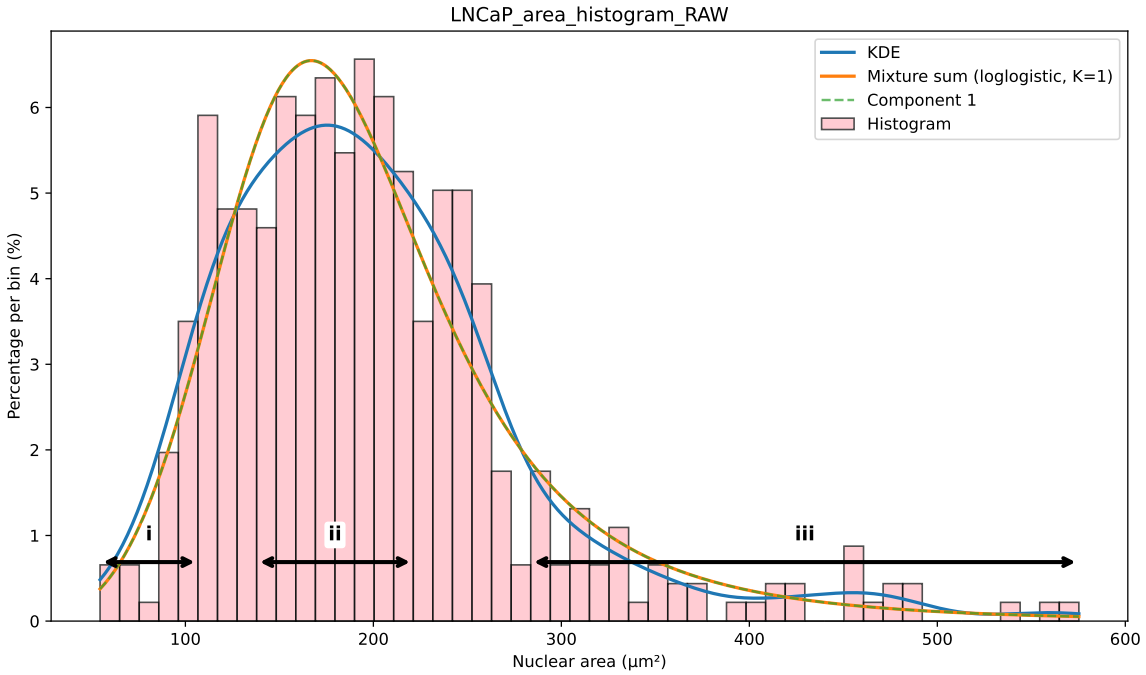


Figure 4.19: Histogram showing the nuclear area in permeabilised LNCaP cells stained with DAPI and anti-fibrillarin (DAPI+FIB). A KDE is overlaid on top of the histogram, with fitting applied to the KDE to determine the overall distribution of the data [258]. The best model (chosen by ICL [259]) was a single log-logistic distribution with c = 4.4 ± 0.2, loc = 0, and scale = 185 ± 3, with a peak at $(167 \pm 3)\,\mu\text{m}^2$. The KDE peak location was $(180 \pm 10)\,\mu\text{m}^2$. Uncertainties are calculated from the covariance matrix.

In both cell lines, nuclear area remained right-skewed after FIB staining (Figures 4.19 and 4.20). The mean exceeded the mode by $\sim 15\%$ in LNCaP, and by $\sim 18\%$ in PNT2-C2, indicating that in both lines the mean nuclear area is materially influenced by a minority of high-area nuclei. Using fixed upper-tail thresholds, $\sim 10\%$ of LNCaP nuclei and $\sim 15\%$ of PNT2-C2 nuclei exceed $300\,\mu\text{m}^2$, with 90th percentiles of $296\mu\text{m}^2$ and $325\mu\text{m}^2$ respectively. These results preserve crucial qualitative interpretation from the DAPI-only condition; PNT2-C2 exhibits a larger nuclear area distribution with a heavier high-area tail, and LNCaP retains a substantial right tail consistent with pleomorphic/nucleomegalic subpopulations and/or segmentation artefacts in crowded fields.

Similarly, aspect ratio distributions remained strongly right-tailed, however the parametric distribution which best described LNCaP changed from a skew-normal to a log-normal (Figures 4.21 and 4.22). The mean exceeded the mode by $\sim 20\%$ in LNCaP, while PNT2-C2 showed a smaller separation of $\sim 15\%$. Consistent with this, $\sim 40\%$ of LNCaP nuclei had an aspect ratio greater than 2 (with a 90th percentile of 2.78), whereas only $\sim 5\%$ of PNT2-C2 nuclei had an aspect ratio > 2 (here the 90th percentile was 1.76). These results reinforce the established conclusion that LNCaP contains a large

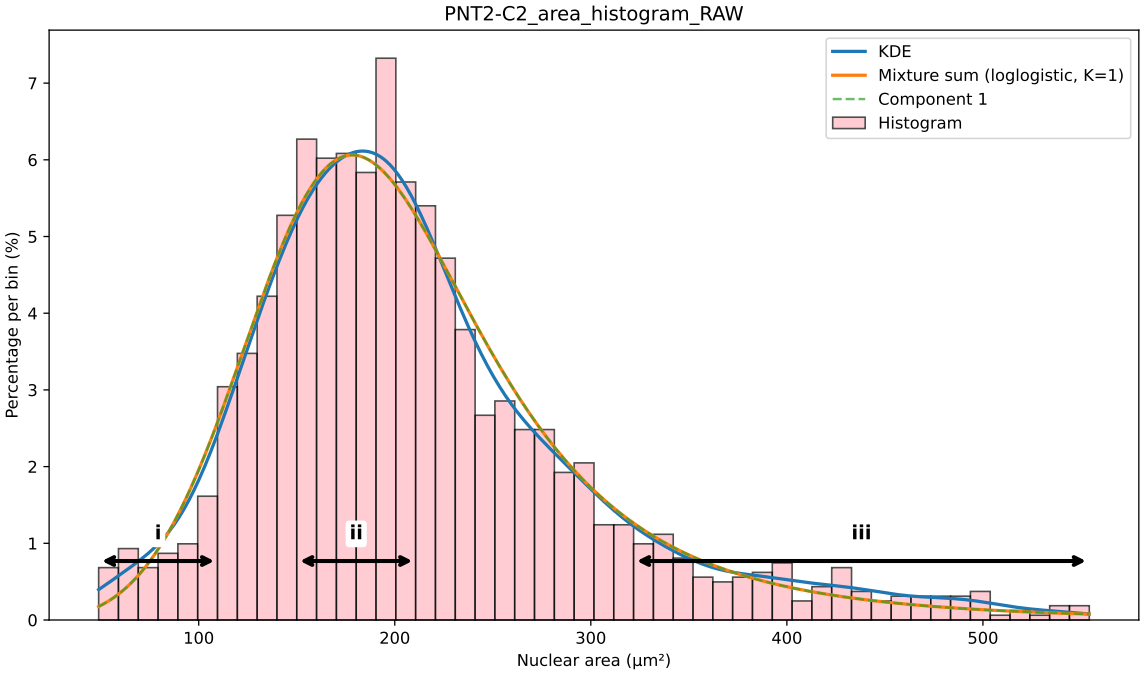


Figure 4.20: Histogram showing the nuclear area in permeabilised PNT2-C2 cells stained with DAPI and anti-fibrillarin (DAPI+FIB). A KDE is overlaid on top of the histogram, with fitting applied to the KDE to determine the overall distribution of the data [258]. The best model (chosen by ICL [259]) was a single log-logistic distribution with c = 4.50 ± 0.09, loc = 0, and scale = 197 ± 2, with a peak at $(178 \pm 2)\,\mu m^2$. The KDE peak location was $(184 \pm 5)\,\mu m^2$. Uncertainties are calculated from the covariance matrix.

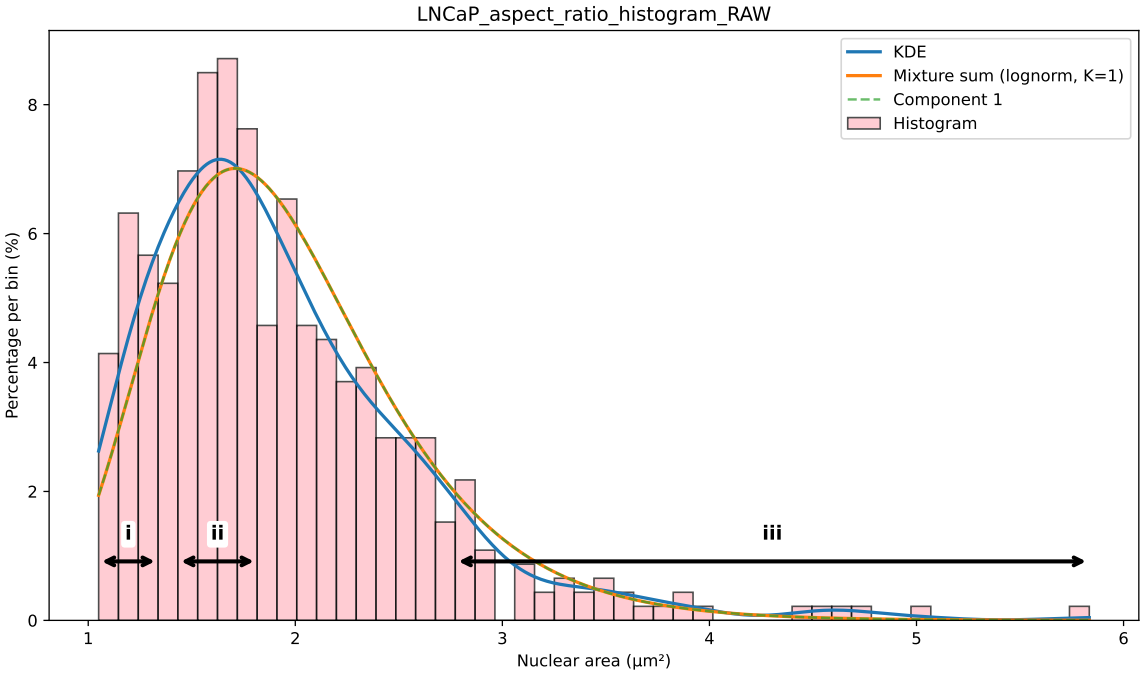


Figure 4.21: Histogram showing the nuclear aspect ratio in permeabilised LNCaP cells stained with DAPI and anti-fibrillarin (DAPI+FIB) A KDE is overlaid on top of the histogram, with fitting applied to the KDE to determine the overall distribution of the data [258]. The best model (chosen by ICL [259]) was a single log-normal distribution with s = 0.30 ± 0.01, loc = 0, and scale = 1.88 ± 0.03, with a peak at 1.71 ± 0.03. The KDE peak location was 1.64 ± 0.04. Uncertainties are calculated from the covariance matrix.

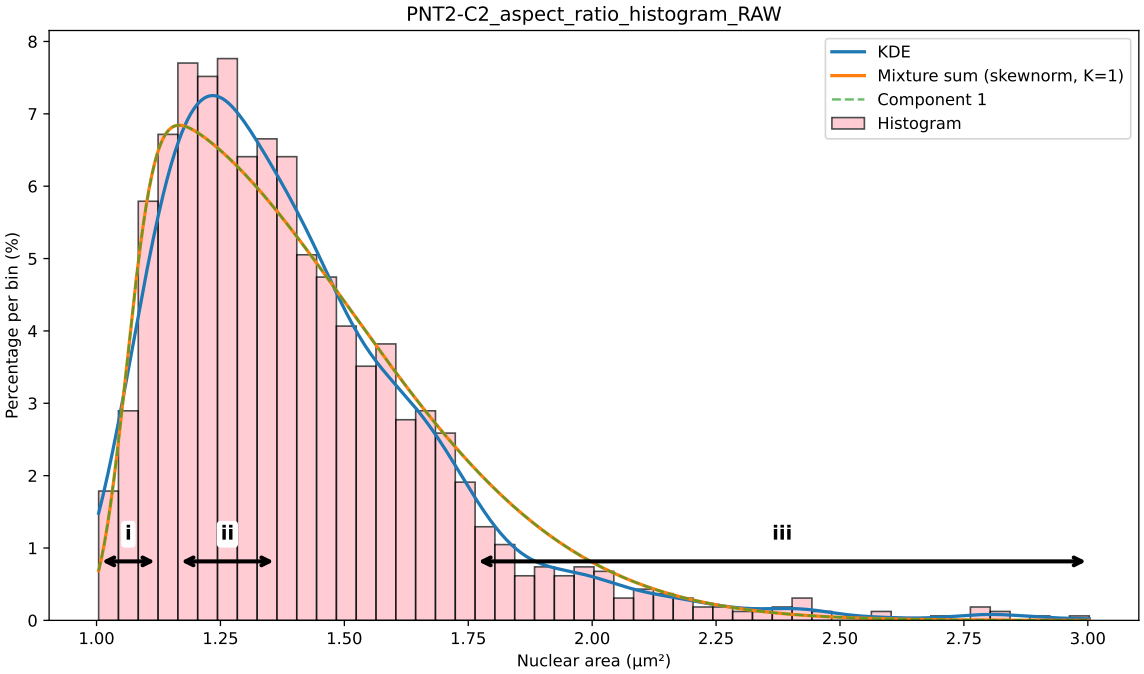


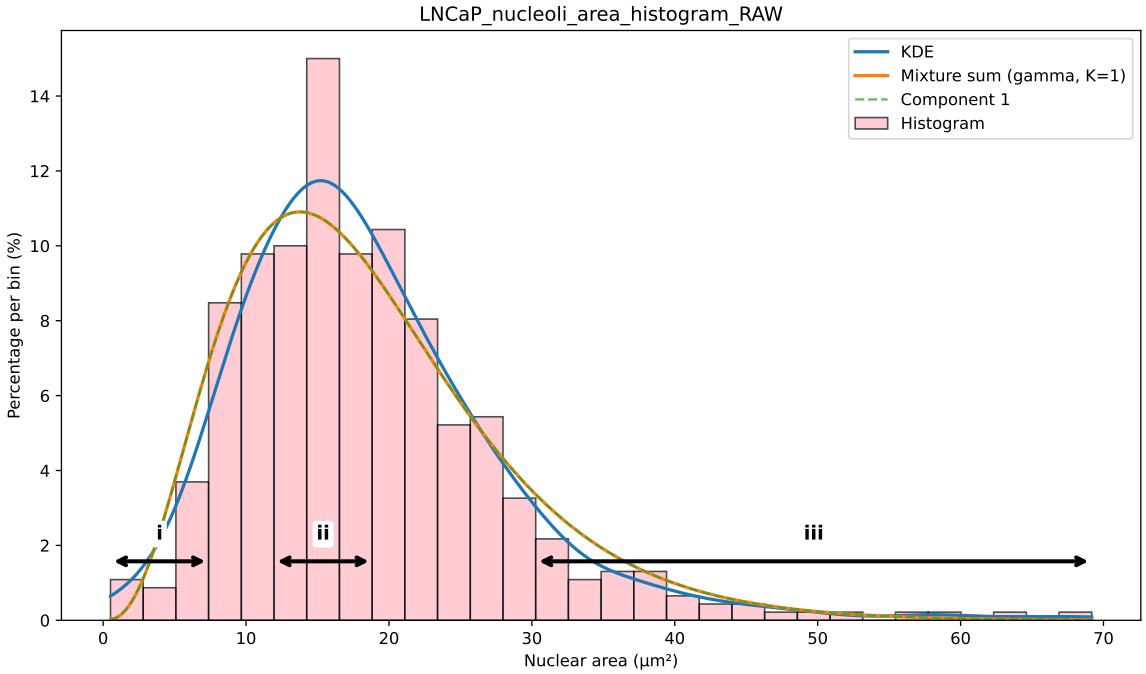
Figure 4.22: Histogram showing the nuclear aspect ratio in permeabilised PNT2-C2 cells stained with DAPI and anti-fibrillarin (DAPI+FIB). A KDE is overlaid on top of the histogram, with fitting applied to the KDE to determine the overall distribution of the data [258]. The best model (chosen by ICL [259]) was a single skew-normal distribution with a $= 10 \pm 1$, loc $= 1.061 \pm 0.006$, and scale $= 0.450 \pm 0.009$, with a peak at 1.17 ± 0.01. The KDE peak location was 1.23 ± 0.01. Uncertainties are calculated from the covariance matrix.

fraction of highly elongated/deformed nuclei relative to PNT2-C2, commensurate with the observations from Figure 4.15.

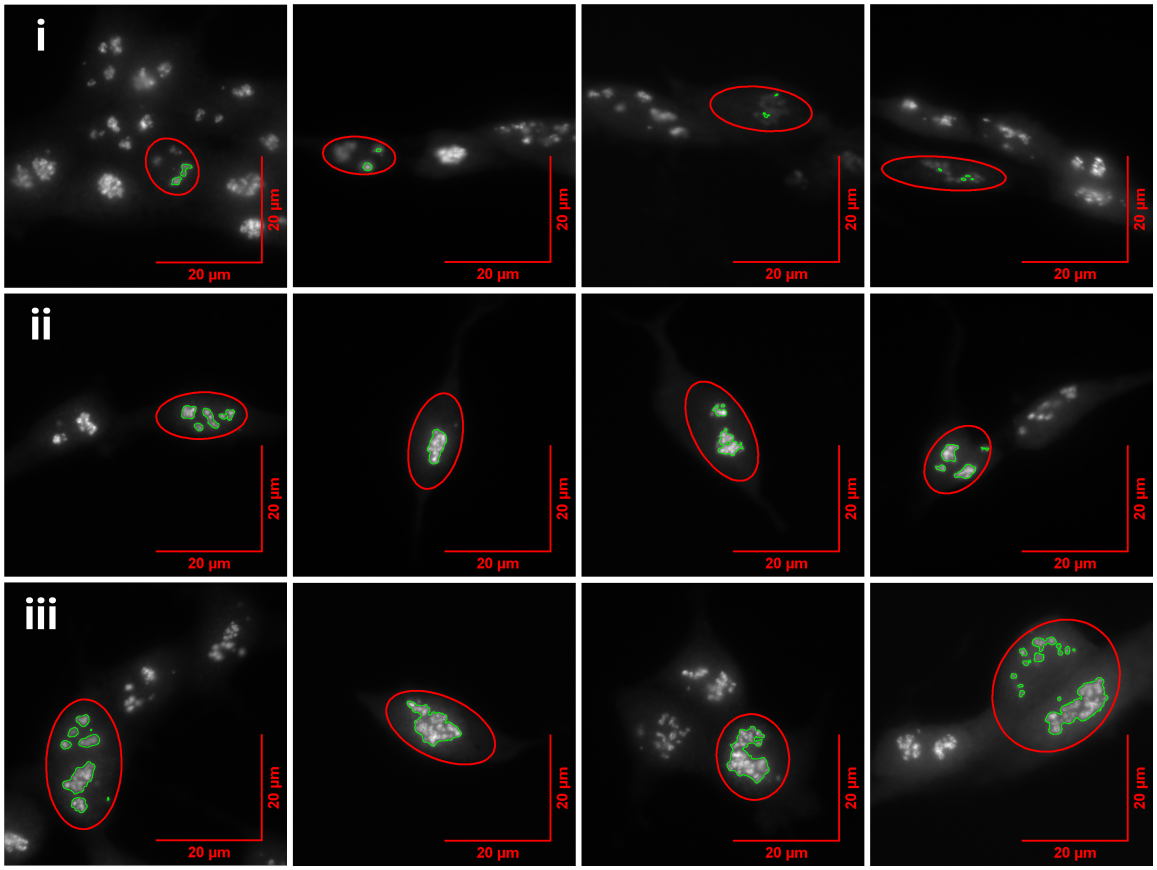
Under the present definition, both cell lines exhibited strongly right-skewed nucleolar area distributions (Figures 4.23 and 4.24), with near-identical bulk central values but materially different dispersion and tail prevalence. The type of parametric distribution which best described the data also differed between cell lines, with LNCaP being best described by a Gamma distribution, and PNT2-C2 by a log-logistic, suggesting heavier tailed behaviour in PNT2-C2. In LNCaP, the mean nucleolar area ($(18.5 \pm 0.4)\,\mu m^2$) exceeded the mode ($(15.2 \pm 0.7)\,\mu m^2$) by $\sim 22\%$. PNT2-C2 showed an almost identical separation ($(18.6 \pm 0.3)\,\mu m^2$ vs. $(15.3 \pm 0.5)\,\mu m^2$, a relative increase of $\sim 22\%$), indicating that the arithmetic mean is substantially tail-driven in both cell lines, and therefore should not be interpreted as the "typical" nucleolar footprint in isolation.

Despite the similarity in mean-mode separation, the tail mass showed greater difference between cell lines, with approximately 3% of LNCaP nuclei exhibiting nucleolar area $>40\,\mu m^2$, compared to 5.5% in PNT2-C2 (a $\sim 1.8\times$ increase in the prevalence of large nucleoli in PNT2-C2). The 90th percentiles similarly show a modest high-end increase in PNT2-C2 ($32.0\mu m^2$ compared to $30.1\mu m^2$ in LNCaP). Taken together with the higher nucleolar area CV ($\sim 60\%$ in PNT2-C2 vs. $\sim 50\%$ in LNCaP), this indicates that the non-malignant model exhibits more frequent extreme fibrillarin-positive nucleolar footprints, even though the typical (modal) nucleolar area is essentially indistinguishable between cell lines.

This result runs counter to the simplest expectation that prostate cancer models should necessarily exhibit larger nucleoli than non-malignant comparators. There are two mechanistically plausible interpretations consistent with both prostate cancer biology and the present imaging pipeline; either

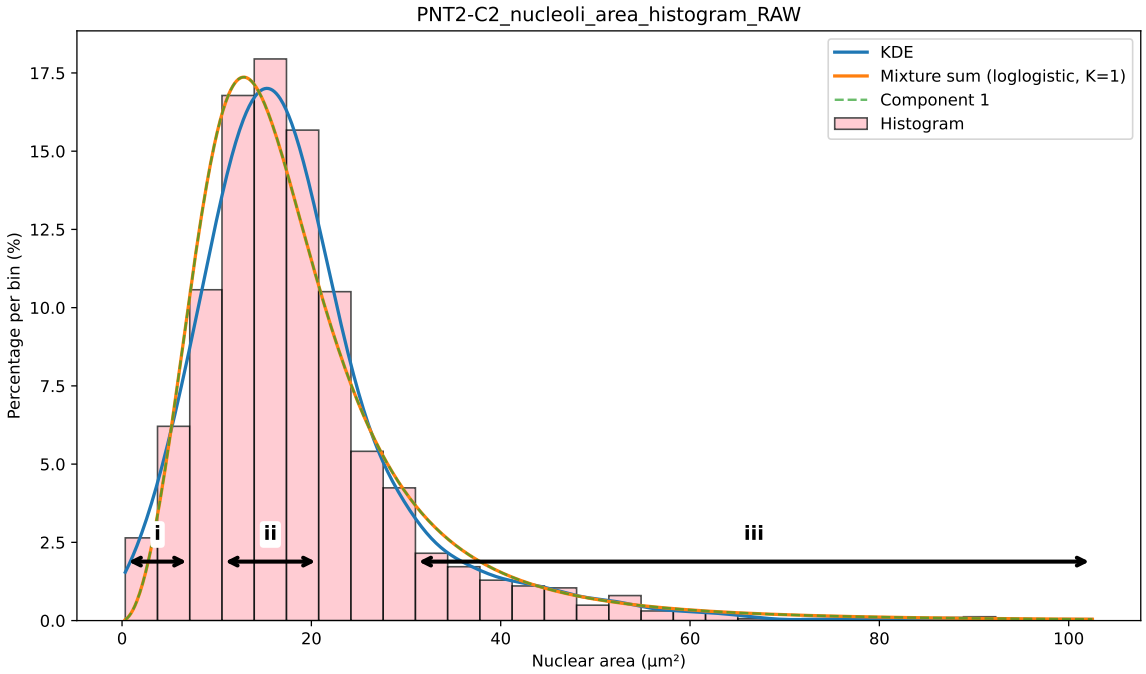


(a) Histogram showing the nucleolar area in each LNCaP nucleus. A KDE is overlaid on top of the histogram, with fitting applied to the KDE to determine the overall distribution of the data [258]. The best model (chosen by ICL [259]) was a single gamma distribution with a $= 3.9 \pm 0.2$, loc $= 0$, and scale $= 4.8 \pm 0.3$, with a peak at $(13.8 \pm 0.4)\,\mu m^2$. The KDE peak location was $(15.2 \pm 0.7)\,\mu m^2$.

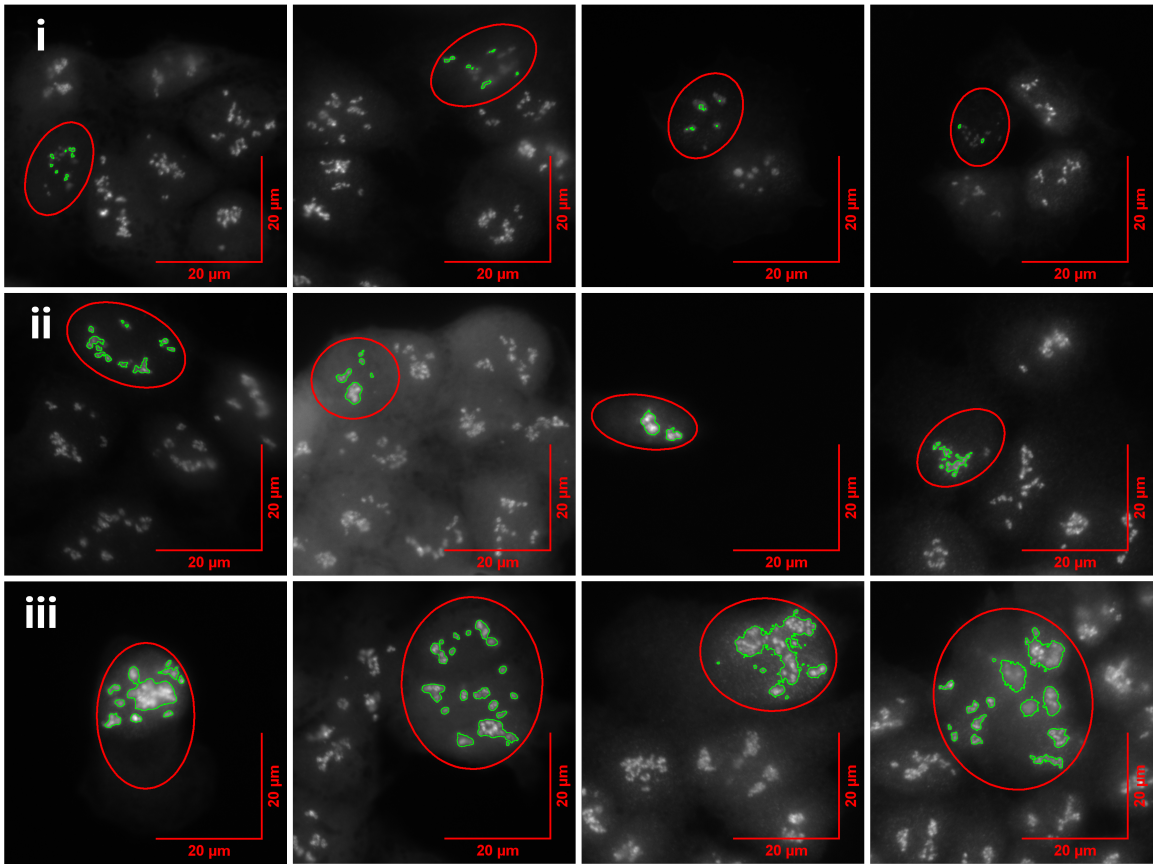


(b) Images of detected nuclei selected from the low (i), modal (ii), and high (iii) regions of the corresponding histogram. Red ellipses are drawn purely for illustrative purposes to show the approximate region which was detected as a nucleus by the segmentation code, with regions detected as nucleoli outlined in green. Nucleolar area was calculated using pixel counts above the second Otsu threshold in the FIB channel within the nuclei mask. A 20 µm scale bar is shown in red at the bottom right corner of each image.

Figure 4.23: Histogram showing the nucleolar area in permeabilised LNCaP cells stained with DAPI and anti-fibrillarin (DAPI+FIB) ((a)), with some images of representative nuclei from the low, modal, and high regions of the histogram ((b)). The KDE is used as a basis for fitting, with the histogram being purely illustrative. The best fit was chosen using an Integrated Classification Likelihood (ICL) [259]. Uncertainties are calculated from the covariance matrix.



(a) Histogram showing the nucleolar area in each PNT2-C2 nucleus. A KDE is overlaid on top of the histogram, with fitting applied to the KDE to determine the overall distribution of the data [258]. The best model (chosen by ICL [259]) was a single log-logistic distribution with c = 2.96 ± 0.06, loc = 0, and scale = 16.3 ± 0.2, with a peak at $(12.9 \pm 0.2)\,\mu m^2$. The KDE peak location was $(15.3 \pm 0.5)\,\mu m^2$.



(b) Images of detected nuclei selected from the low (i), modal (ii), and high (iii) regions of the corresponding histogram. Red ellipses are drawn purely for illustrative purposes to show the approximate region which was detected as a nucleus by the segmentation code, with regions detected as nucleoli outlined in green. Nucleolar area was calculated using pixel counts above the second Otsu threshold in the FIB channel within the nuclei mask. A 20 µm scale bar is shown in red at the bottom right corner of each image.

Figure 4.24: Histogram showing the nucleolar area in permeabilised PNT2-C2 cells stained with DAPI and anti-fibrillarin (DAPI+FIB) ((a)), with some images of representative nuclei from the low, modal, and high regions of the histogram ((b)). The KDE is used as a basis for fitting, with the histogram being purely illustrative. The best fit was chosen using an Integrated Classification Likelihood (ICL) [259]. Uncertainties are calculated from the covariance matrix.

the true biological difference in whole-nucleolus size between LNCaP and PNT2-C2 is truly small under these culture conditions, particularly since PNT2-C2 is SV40-immortalised and therefore can maintain high proliferative demand and ribosome biogenesis activity, despite being non-tumourigenic, and/or the nucleolar segmentation may systematically under-report nucleolar enlargement if the relevant structural change occurs primarily outside the dense fibrillar component (e.g. altered granular component organisation), such that whole nucleolar prominence does not map one-to-one onto the fibrillarin-positive area [136, 138, 219]. Another possible route to these observations is that the nucleolar state was significantly perturbed by sample preparation steps, obscuring real biological differences as a result; given the observations from the DAPI channel, this assessment is not unexpected (though not supported directly, since it was not possible to measure nucleolar state directly prior to permeabilisation in this study).

A visual inspection of detected nucleoli (Figures 4.23(b) and 4.24(b)) reveals that a not insignificant fraction of the smallest nucleolar area measurements arise from segmentation failure, rather than genuinely small nucleoli. Nucleolar detection was performed using per-image intensity thresholding (Otsu's method) on the FIB channel; consequently, when a given field of view contains exceptionally bright nucleoli, the resulting threshold can exclude lower intensity nucleolar pixels elsewhere within the same image. This manifests as regions that capture only the brightest nucleolar core, effectively biasing the nucleolar area downwards. Qualitatively, this underestimation mode appears more frequently in LNCaP, consistent with a larger intra-image dynamic range in fibrillarin signal, and therefore a greater propensity for faint nucleolar signal to fall below the chosen threshold.

At the opposite extreme, the largest nucleolar area values are occasionally driven by upstream nuclear undersegmentation in the DAPI channel, where multiple adjacent nuclei are merged into a single nuclear object. In such cases, fibrillarin-positive regions from multiple true nuclei are aggregated into one measurement, inflating both the apparent nuclear area and the associated summed nucleolar area.

Beyond these clear failure modes, the exemplar images also suggest qualitative differences in fibrillarin signal topology between the cell lines. In LNCaP, nucleoli are frequently detected as one (or a few) relatively contiguous, irregularly shaped domains per nucleus, with each typically being larger than those found in PNT2-C2. In contrast, in PNT2-C2 the FIB signal is more often split across multiple individual discrete regions distributed across the nuclear interior (with larger connected domains comparatively much rarer than in LNCaP). Importantly, these pattern differences can reflect genuine biological organisation (e.g. differences in nucleolar coalescence/compaction and the spatial distribution of DFC-associated signal), but can also reflect threshold dependent fragmentation and connectivity effects. As such, the nucleolar area reported here should be interpreted conservatively; the qualitative overlays motivate treating area-based conclusions with caution where segmentation artefacts are suspected.

The nuclear area normalised occupancy ratio provided a clearer separation between cell lines than absolute nucleolar area. In LNCaP, $R_{N:NA}$ was relatively tightly centred, and approximately symmetric (Figure 4.25), with almost identical mean and mode (0.093 ± 0.001 vs. 0.093 ± 0.002), giving no mean-mode separation outside of uncertainty. In contrast to absolute nucleolar area, the LNCaP nucleolar occupancy ratio was not strongly tail-dominated; rather, the arithmetic mean is closely representative of the typical nucleus. Functionally, this supports the interpretability of $R_{N:NA}$ as a

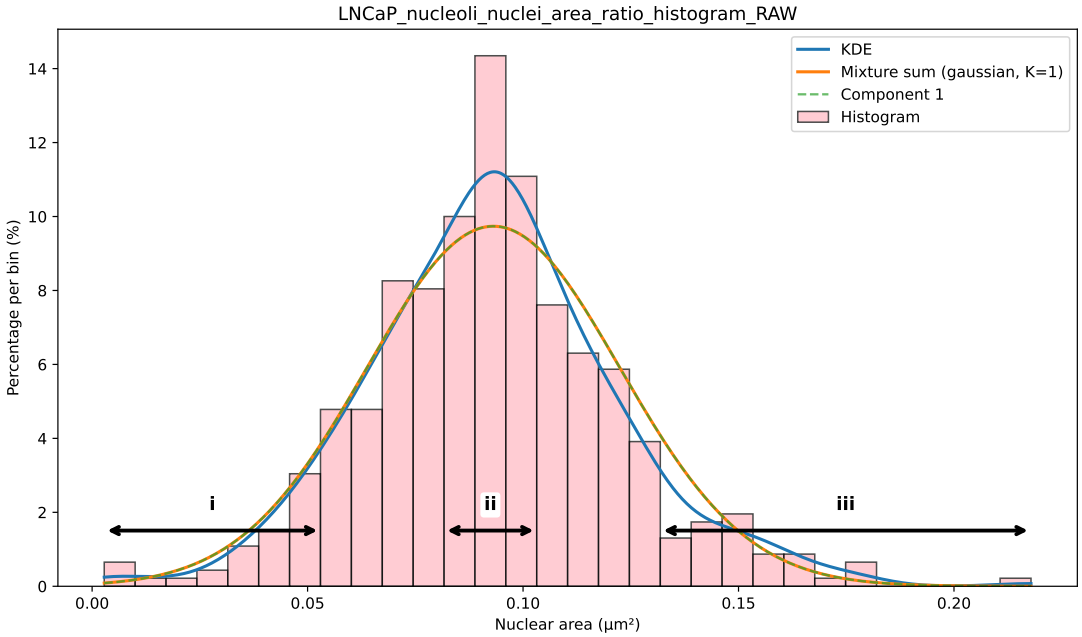


Figure 4.25: Histogram showing the ratio of nucleolar area to nuclear area in permeabilised LNCaP cells stained with DAPI and anti-fibrillarin (DAPI+FIB). A KDE is overlaid on top of the histogram, with fitting applied to the KDE to determine the overall distribution of the data [258]. The best model (chosen by ICL [259]) was a single Gaussian distribution with $\mu = 0.093 \pm 0.001$, and $\sigma = 0.029 \pm 0.001$, with a peak at 0.093 ± 0.001. The KDE peak location was 0.093 ± 0.002. Uncertainties are calculated from the covariance matrix.

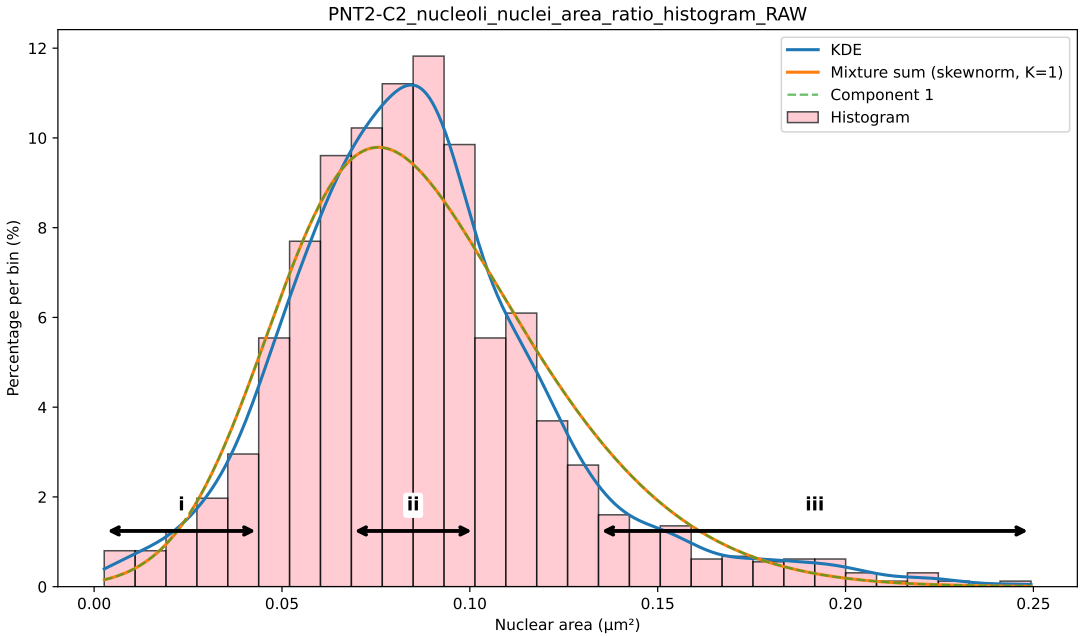


Figure 4.26: Histogram showing the ratio of nucleolar area to nuclear area in permeabilised PNT2-C2 cells stained with DAPI and anti-fibrillarin (DAPI+FIB). A KDE is overlaid on top of the histogram, with fitting applied to the KDE to determine the overall distribution of the data [258]. The best model (chosen by ICL [259]) was a single skew-normal distribution with a $= 2.4 \pm 0.2$, loc $= 0.049 \pm 0.001$, and scale $= 0.052 \pm 0.001$, with a peak at 0.076 ± 0.001. The KDE peak location was 0.084 ± 0.002. Uncertainties are calculated from the covariance matrix.

robust, normalised nucleolar morphology metric in the malignant model.

PNT2-C2 on the other hand showed a markedly more skewed $R_{N:NA}$ distribution (Figure 4.26), with the mean (0.0869 ± 0.0009) exceeding the mode (0.084 ± 0.002) by $\sim 3\%$. This asymmetry is consistent with a heavier right-tail in PNT2-C2, corroborated by the higher $R_{N:NA}$ CV ($\sim 42\%$ vs. $\sim 31\%$ in LNCaP), change in parametric characterisation (with LNCaP best described by a Gaussian, and PNT2-C2 best described by a skew-normal), and the increased fraction of high occupancy outliers. Specifically, $\sim 5.5\%$ of PNT2-C2 nuclei had $R_{N:NA} > 0.15$, compared with $\sim 3.5\%$ in LNCaP. Importantly, however, the 90th percentiles were essentially identical between cell lines (~ 0.13), indicating that the difference in "extreme" high occupancy between cell lines is concentrated predominantly beyond the 90th percentile (i.e. in the most extreme $\sim 5\%$ tails), rather than reflecting a broad upward shift of the high-end distribution.

Despite the heavier PNT2-C2 tail, both the mean and mode $R_{N:NA}$ were higher in LNCaP. The modal separation is particularly informative, with the most probable nucleolar occupancy in LNCaP exceeding that of PNT2-C2 by $\sim 11\%$, supporting the interpretation that for the dominant population of nuclei, the malignant model exhibits greater relative nucleolar occupancy, a trend which may have been even more obvious were it not for the various segmentation artefacts. Biologically, this is consistent with the framework in which nucleolar remodelling in cancer supports elevated ribosomal biogenesis capacity and protein synthesis demand, and with evidence that nucleolar programmes in prostate cancer are driven by oncogenic transcriptional programmes such as MYC, with fibrillarin implicated as a functional node [17, 138, 141]. In both tissue and cell line studies, prominent/enlarged nucleoli are a widely recognised feature of prostate cancer, and can correlate with aggressive disease states, though they are not universally present across all carcinomas [18, 63, 72, 224]. The present data therefore support a cautious but mechanistically plausible interpretation; LNCaP exhibits a modest increase in nucleolar occupancy, even when absolute nucleolar area is not significantly increased.

The above nucleolar conclusions are conditional on the operational definition implemented in the segmentation workflow. In particular, the absence of a difference in absolute nucleolar area between LNCaP and PNT2-C2 may reflect (at least) three technical factors, each of which would act to erode between-line differences:

1. Fibrillarin is predominantly contained within the dense fibrillar component; whole-nucleolus prominence may be driven by other compartments, or by reorganisation rather than proportional DFC expansion [138, 219].

2. Multi-Otsu thresholding with two thresholds on the scale of whole images will tend to prioritise high contrast nucleolar cores and under-segment more diffuse nucleolar signal, potentially biasing estimates towards brighter and more homogeneous "cores", even when nucleoli differ in overall size. Coupled with the observed segmentation failures in the very low region of nucleolar area histograms (especially for LNCaP), the increased fraction of low area nucleoli (thus reduced fraction of larger nucleoli assuming perfect segmentation) may act to artificially reduce both the mean and modal nucleolar area, obscuring real biological differences.

3. The heavy tailed structure and increased outlier fractions of nucleolar area plots (particularly in PNT2-C2) indicate that some portion of the extreme right tail includes segmentation edge cases. The upper tail should thus be interpreted as both biologically and technically informative;

it constrains the prevalence of rare, large nucleolar footprints, but also provides an upper bound on segmentation failures.

These caveats motivate two higher-level conclusions that remain robust despite the stated uncertainties. First, $R_{N:NA}$ is empirically the more discriminative and interpretable nucleolar morphology endpoint in this dataset, and yields a clear shift in central tendency while remaining close to mean-mode agreement in LNCaP. Second, the simple statement "LNCaP has larger nucleoli than PNT2-C2" is not supported by the measured nucleolar area under this workflow; instead, the data support a more nuanced view in which LNCaP exhibits higher typical nucleolar occupancy, while PNT2-C2 exhibits greater nucleolar heterogeneity and a heavier extreme-outlier tail. While these conclusions are biologically plausible given that PNT2-C2 is SV40-immortalised and can therefore sustain nucleolar activity associated with proliferative demand despite being non-malignant, and that nucleolar alterations in prostate neoplasia can manifest as changes in number/organisation and in nucleolar gene programmes rather than purely as a simple increase in size, it may be more likely that under the stated workflow, the analysis was simply not accurate/sensitive enough to separate any real biological changes from those introduced through sample preparation and imaging/segmentation [3, 17, 18].

4.2.2.2 Intensity analysis

Bulk summaries of DAPI-derived intensity descriptors under the DAPI+FIB workflow are shown in Figure 4.27. As previously (Sections 3.2.2 and 4.1.2.2), mean and modal values are reported for total nuclear DAPI intensity, I_{DAPI}, alongside the area normalised DAPI density defined in Equation (2.10), together with dispersion measures defined either across nuclei (CV of I_{DAPI} and CV of DAPI density, Equation (2.27)) or within each nucleus ($\mathrm{CV}_{\mathrm{DAPI}}$, Equation (2.13)).

As was the case for nuclear morphology, the detailed DAPI intensity summaries for the DAPI+FIB preparation in Figure 4.27 are retained as the within-condition reference for the present dataset, including the exact mean, mode, and CV values for the DAPI-derived intensity descriptors under the dual-stain workflow. The comparison figures (Figures 4.28 to 4.30) collate the corresponding DAPI-derived intensity results across the non-permeabilised, permeabilised DAPI-only, and permeabilised DAPI+FIB conditions in order to more directly observe the effects of detergent permeabilisation. As previously, the plotted mode is the peak position of the kernel density estimate (KDE; see Section 2.3.5 and Equation (2.34)) for the corresponding pooled nucleus-level distribution, consistent with the histogram analyses in Sections 3.2.2, 4.1.2.2 and 4.2.2.2, and error bars represent standard error (which again was calculated using the delta method where required; Equations (2.28) and (2.29)).

The first result from Figure 4.28 is that the apparent total DAPI-intensity ordering is not preparation-invariant. In the non-permeabilised baseline (Figure 3.14), LNCaP nuclei showed an approximately 40% increased mean total DAPI intensity relative to PNT2-C2, while the modal separation was larger still, with the typical LNCaP nucleus exceeding the typical PNT2-C2 nucleus by approximately 60%, consistent with the established aneuploid tendency of LNCaP [2, 14] and with the larger baseline nuclear area in the malignant model (Figure 3.8). After permeabilisation, this contrast was markedly reduced; in the DAPI-only condition, total DAPI intensity per nucleus in LNCaP remained only slightly higher in the mean ($\sim 10\%$) and moderately higher in the mode ($\sim 30\%$ Figure 4.7). After FIB staining, this ordering then reversed when considering the means, where total intensity per PNT2-C2 nucleus exceeded that in LNCaP by approximately 15% ($(8.8 \pm 0.1) \times 10^8$ a.u. vs. $(7.7 \pm 0.2) \times 10^8$ a.u.), and modal

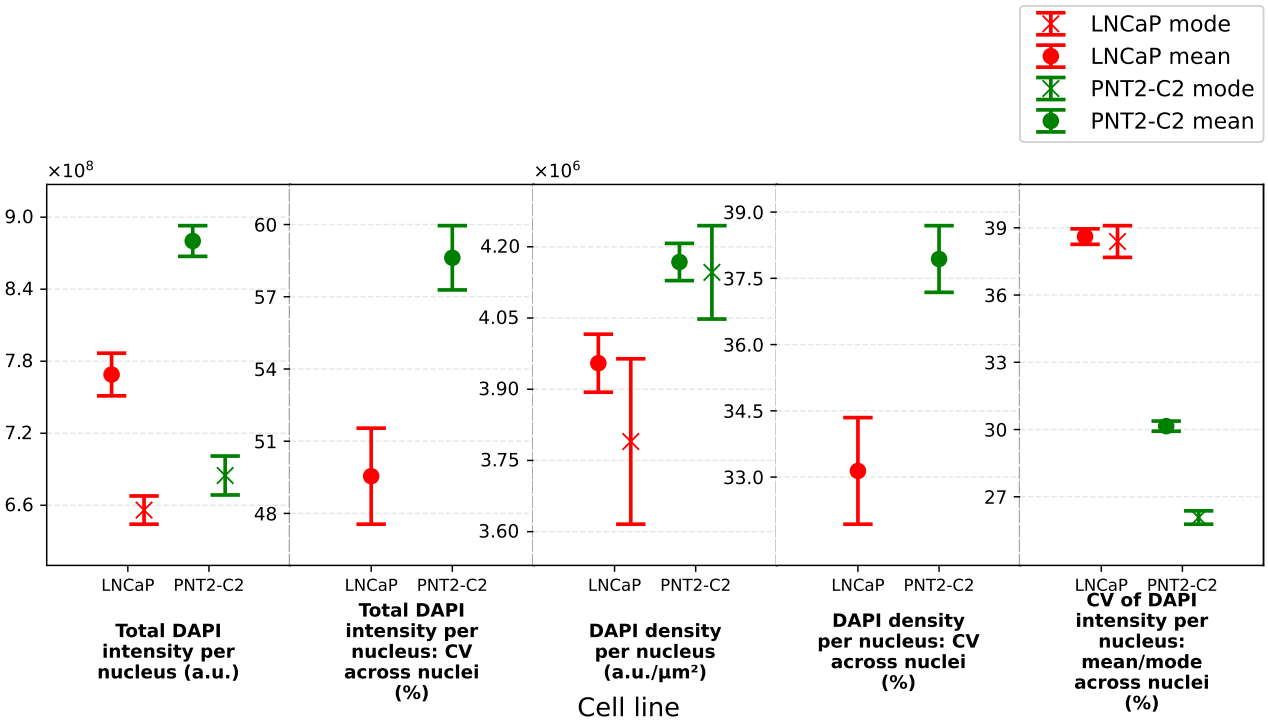


Figure 4.27: Plot showing key DAPI intensity summary statistics for permeabilised LNCaP and PNT2-C2 cells stained with DAPI and anti-fibrillarin (DAPI+FIB). Total DAPI intensity per nucleus was the sum of all pixel intensities in the nuclei mask for a given nucleus. DAPI density was the total intensity divided by nuclear area. CV of DAPI intensity per nucleus was the coefficient of variation of all pixel intensity values in the nuclei mask for a given nucleus. Mode here refers to the peak position on the x-axis of a kernel density estimate for each metric (see Sections 2.3.5 and 4.2.2.2). Error bars represent standard error (calculated using the delta method for CV across nuclei and modes; Equations (2.28) and (2.29).)

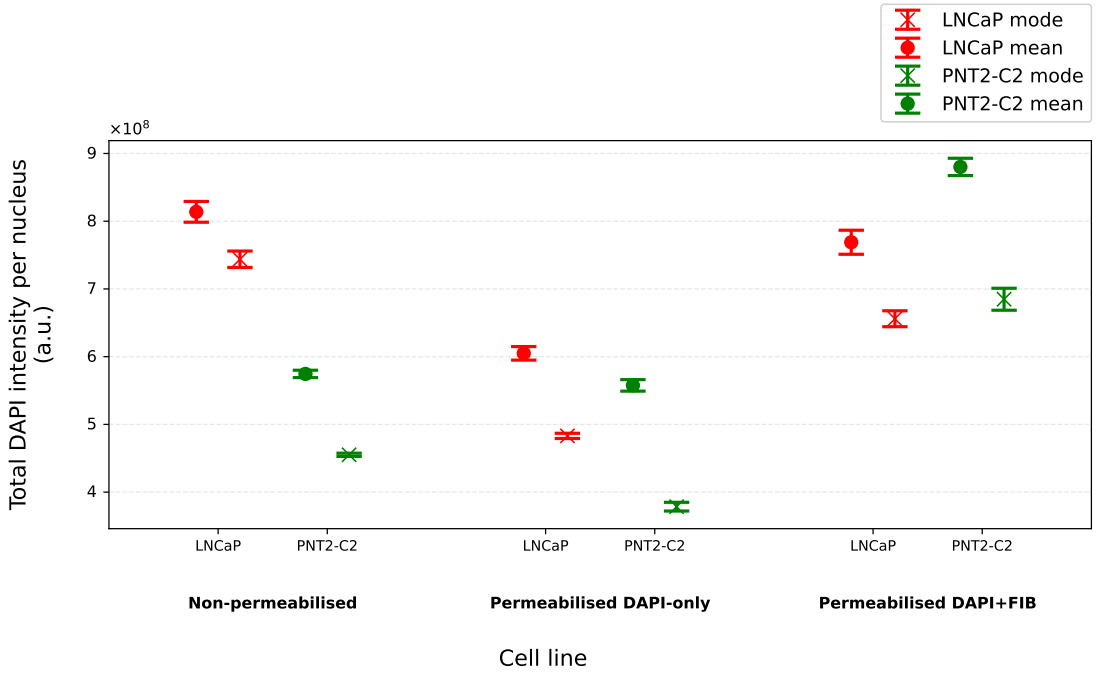


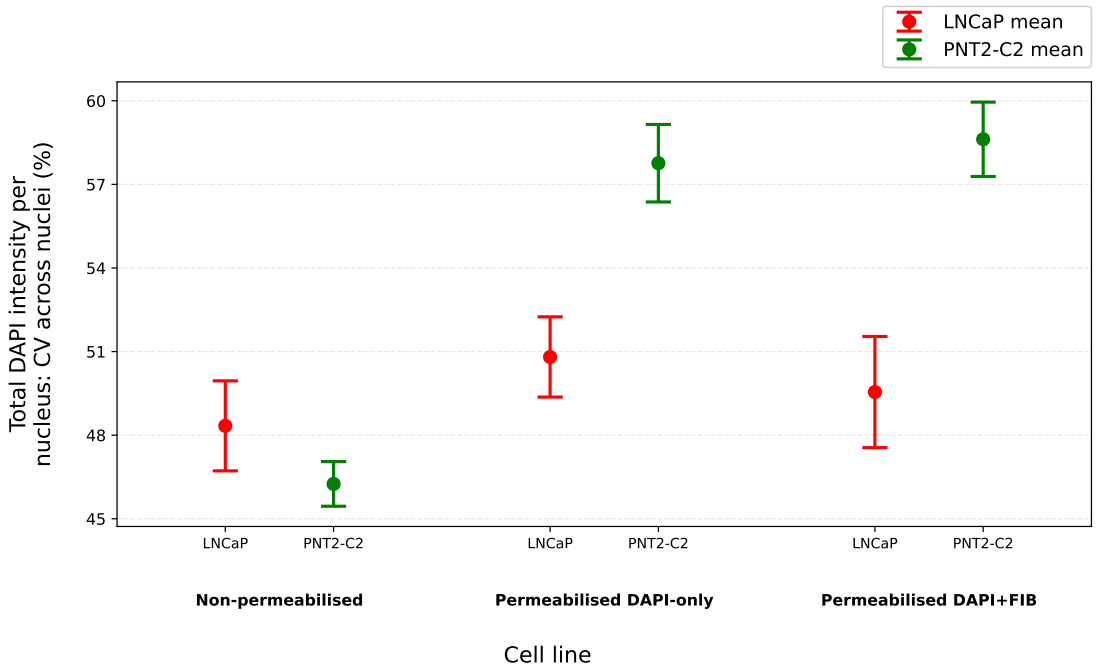
Figure 4.28: Cross-condition comparison of total nuclear DAPI intensity, I_{DAPI}, across non-permeabilised DAPI, permeabilised DAPI-only, and permeabilised DAPI+FIB preparations. Mean and mode summarise the same pooled nucleus-level distributions examined in Sections 3.2.2, 4.1.2.2 and 4.2.2.2, with the mode defined as the KDE peak location (Equation (2.34)). Error bars represent standard error.

values became effectively indistinguishable, not differing outside of uncertainty ($(6.8 \pm 0.2) \times 10^8$ a.u. in PNT2-C2 vs. $(6.6 \pm 0.1) \times 10^8$ a.u. in LNCaP; Figure 4.27). Relative to the permeabilised DAPI-only reference, this reversal was driven primarily by the approximately 2× greater increase in PNT2-C2 mean total DAPI intensity under DAPI+FIB than in LNCaP observed in Figure 4.28 ($\sim 60\%$) vs. ($\sim 30\%$). The fact that the modal difference also effectively disappeared indicates that this is not purely a tail-driven effect in a small minority of nuclei; rather, the dominant population itself shifted between conditions. As such, absolute I_{DAPI} appears to be strongly influenced by the differences in preparation across conditions in this study (and at the very least is directly affected by the changes in nuclear size observed in Figure 4.16) and should thus be interpreted carefully [295, 296].

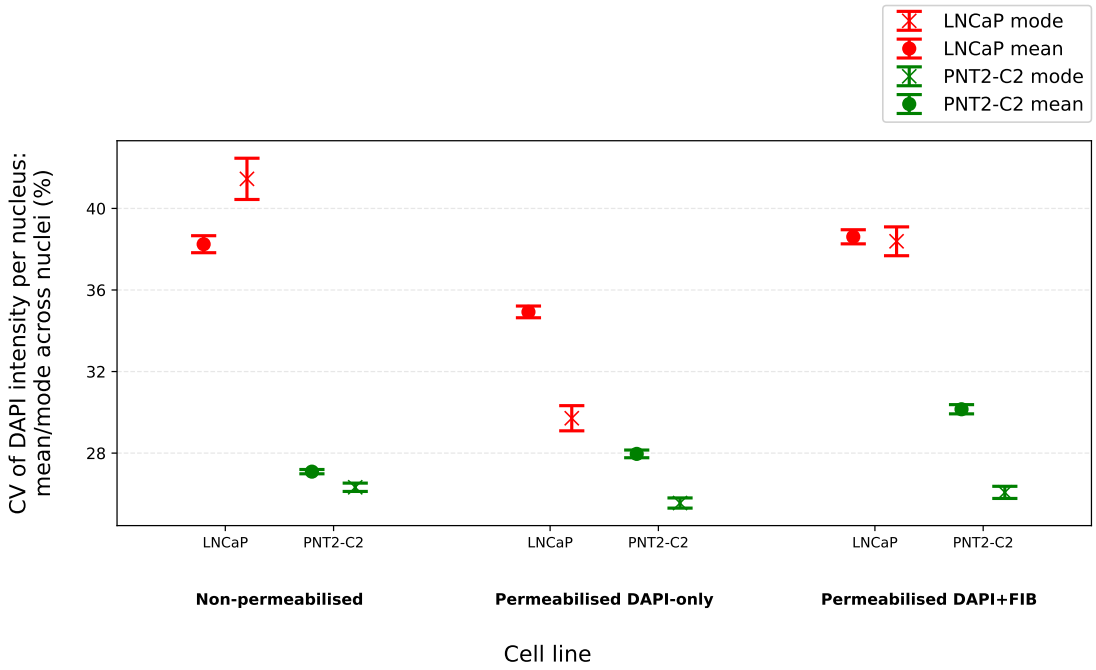
The heterogeneity measures shown in Figure 4.29 behave more consistently than the absolute intensity scale in Figure 4.28, however this consistency is metric-specific. For the internuclear CV of total DAPI intensity (i.e. CV across nuclei), only one value was reported per cell line, as this descriptor is necessarily defined across nuclei within each condition (Equation (2.27)). In the non-permeabilised baseline (Figure 3.14), the total DAPI intensity CV across nuclei did not differ outside of uncertainty, after permeabilisation, however, PNT2-C2 nuclei had a notably higher ($\sim 15\%$) relative dispersion in total DAPI intensity than LNCaP when stained with only DAPI ($(58 \pm 1)\,\%$ vs. $(51 \pm 1)\,\%$; Figure 4.7), and this ordering remained under DAPI+FIB ($(59 \pm 1)\,\%$ vs. $(50 \pm 2)\,\%$; Figure 4.27), where the relative increase grew to $\sim 20\%$ (Figure 4.29(a)). The impact of preparation effects therefore clearly extend beyond a shift in central tendency alone, and also increase the measured spread of total DAPI intensity across nuclei in the non-malignant comparator, with this broader dispersion remaining evident even after the additional immunolabelling steps.

In contrast, the intranuclear texture metric $\mathrm{CV}_{\mathrm{DAPI}}$ preserved the same cell-line ordering across all three conditions, and did so at the level of both the mean and mode, though the magnitude of the differences between cell lines was not consistent Figure 4.29(b). In the non-permeabilised baseline (Figure 3.14), LNCaP exceeded PNT2-C2 by approximately 40% in the mean ($(38.2 \pm 0.4)\,\%$ vs. $(27.1 \pm 0.1)\,\%$) and by approximately 60% in the mode ($(41 \pm 1)\,\%$ vs. $(26.3 \pm 0.2)\,\%$). After permeabilisation (Figure 4.7), this contrast was substantially weakened, however the ordering was preserved, with LNCaP exceeding PNT2-C2 by approximately 25% in the mean ($(34.9 \pm 0.3)\,\%$ vs. $(28.0 \pm 0.2)\,\%$) and approximately 15% in the mode ($(29.7 \pm 0.6)\,\%$ vs. $(25.6 \pm 0.2)\,\%$). Under DAPI+FIB staining (Figure 4.27), the separation then increased again (though not quite returning to the level in the non-permeabilised baseline), with LNCaP exceeding PNT2-C2 by approximately 30% in the mean ($(38.6 \pm 0.3)\,\%$ vs. $(30.2 \pm 0.2)\,\%$) and approximately 45% in the mode ($(38.4 \pm 0.7)\,\%$ vs. $(26.1 \pm 0.3)\,\%$). Thus, although the magnitude of the separation in the relative dispersion of nuclear texture is preparation-dependent, the directional result remains robust across conditions: LNCaP remained the more heterogeneous model throughout Figure 4.29(b). Crucially, this distinguishes the preparation-sensitive absolute intensity scale from a more stable texture-level contrast between malignant and non-malignant nuclei.

The density comparison in Figure 4.30(a) shows that even after normalising to nuclear area, DAPI intensity remains strongly preparation-sensitive, evident through changes in both the mean and mode across conditions. In the non-permeabilised baseline, the mean DAPI density showed no change between cell lines outside of uncertainty, however the modal density in LNCaP was approximately 15% higher ($(4.2 \pm 0.2) \times 10^6$ a.u. µm^{-2} vs. $(3.62 \pm 0.09) \times 10^6$ a.u. µm^{-2}; Figure 3.14), showing that the typical density state already differed between the malignant and non-malignant lines, even where the means overlapped. Following permeabilisation, both summaries shifted to accentuate this difference, with mean DAPI

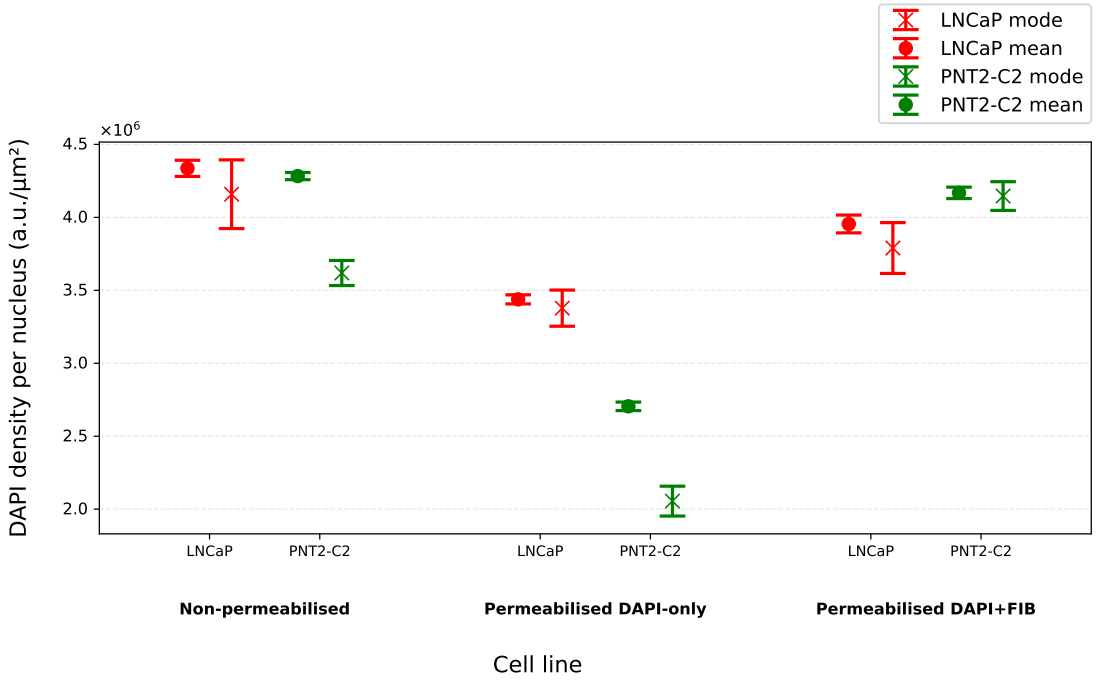


(a) CV of total DAPI intensity across nuclei.

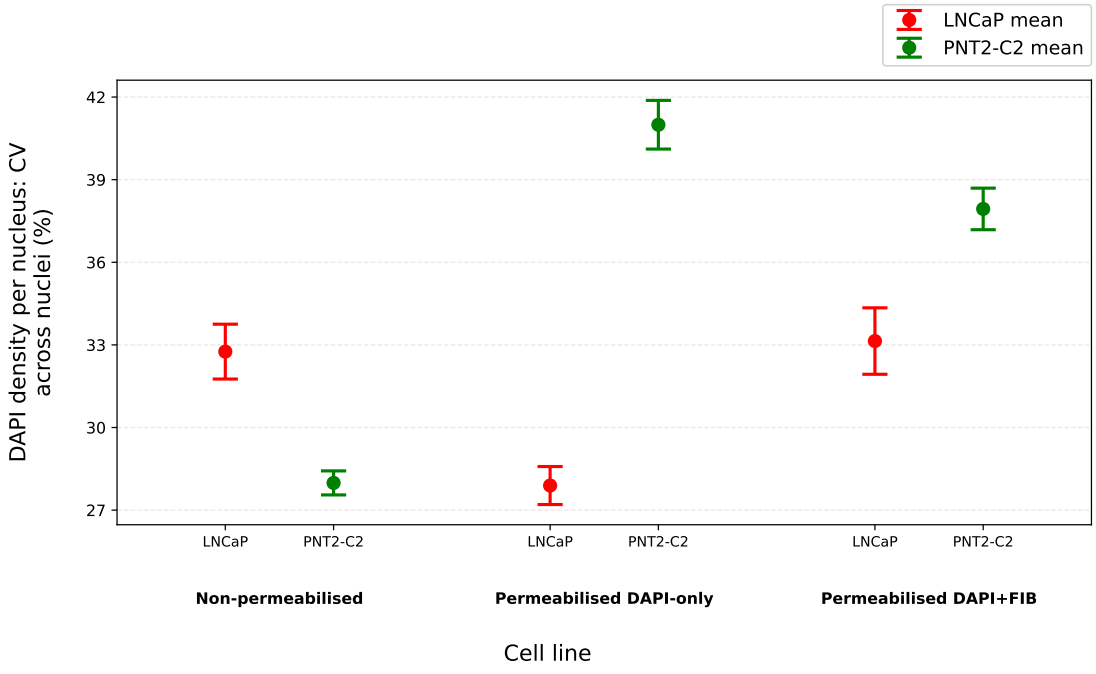


(b) Within-nucleus CV_{DAPI}.

Figure 4.29: Cross-condition comparison of DAPI intensity heterogeneity descriptors across non-permeabilised DAPI, permeabilised DAPI-only, and permeabilised DAPI+FIB preparations. For intranuclear CV_{DAPI}, both mean and mode are shown, with the mode defined as the KDE peak location (Equation (2.34)). Error bars represent standard error.



(a) DAPI density.



(b) CV of DAPI density across nuclei.

Figure 4.30: Cross-condition comparison of DAPI-density descriptors across non-permeabilised DAPI, permeabilised DAPI-only, and permeabilised DAPI+FIB preparations. For DAPI density, mean and mode are shown, with the mode defined as the KDE peak location (Equation (2.34)). Error bars represent standard error.

density in the permeabilised DAPI-only condition for LNCaP being approximately 30% higher relative to PNT2-C2 ($(3.44\pm0.03)\times10^6$ a.u. µm^{-2} vs. $(2.70\pm0.03)\times10^6$ a.u. µm^{-2}), while modal density was almost 65% higher in LNCaP ($(3.4\pm0.1)\times10^6$ a.u. µm^{-2} vs. $(2.1\pm0.1)\times10^6$ a.u. µm^{-2}; Figure 4.7). Under DAPI+FIB, this ordering reversed in both the mean and the mode: mean DAPI density in PNT2-C2 was $\sim 5\%$ higher than in LNCaP ($(4.17\pm0.04)\times10^6$ a.u. µm^{-2} vs. $(3.95\pm0.06)\times10^6$ a.u. µm^{-2}), and modal density was approximately 10% higher ($(4.1\pm0.1)\times10^6$ a.u. µm^{-2} vs. $(3.8\pm0.2)\times10^6$ a.u. µm^{-2}; Figure 4.27). Relative to the permeabilised DAPI-only reference, the modal shift is especially revealing: the typical PNT2-C2 nucleus approximately doubled in density (from $(2.1\pm0.1)\times10^6$ a.u. µm^{-2} to $(4.1\pm0.1)\times10^6$ a.u. µm^{-2}), whereas the corresponding increase in LNCaP was much more modest at $\sim 10\%$. More explicitly, even after normalising total intensity by nuclear area, the apparent between-line ordering remained conditional on preparation state. In the present study, DAPI density therefore cannot be treated as a preparation-independent surrogate for DNA density or chromatin compaction across conditions [8, 295, 296].

The DAPI-density CV across nuclei reinforces this conclusion Figure 4.30(b). In the non-permeabilised condition, LNCaP exhibited the higher density CV ($(33\pm1)\,\%$ vs. $(28.0\pm0.4)\,\%$, a relative increase of $\sim 20\%$; Figure 3.14). After permeabilisation however, this ordering reversed strongly, with PNT2-C2 exceeding LNCaP by approximately 50% ($(41.0\pm0.9)\,\%$ vs. $(27.9\pm0.7)\,\%$; Figure 4.7). Under DAPI+FIB, DAPI density CV remained higher in PNT2-C2, though the separation narrowed to approximately 15% ($(37.9\pm0.8)\,\%$ vs. $(33\pm1)\,\%$; Figure 4.27). Thus, the impact of differing sample preparation is not limited to the absolute density scale; it also changes the relative heterogeneity of DAPI density, with the strongest effect again observed in PNT2-C2 Figure 4.30.

Taken together, the three intensity-based comparison figures (Figures 4.28 to 4.30) show that the principal preparation-sensitive perturbation to the DAPI channel lies in the absolute intensity and density scales, and in some cases is sufficient to reverse the apparent biological ordering between cell lines. By contrast, the direction of the $\mathrm{CV}_{\mathrm{DAPI}}$ contrast was preserved, with LNCaP remaining the more heterogeneous model in terms of intranuclear texture throughout Figure 4.29(b). As for the morphological analysis (Section 4.2.2.1), these comparison plots are not intended to be interpreted in isolation from the histogram analyses. Where the mean and mode diverge, and particularly where heavy tailed behaviour or explicit subpopulation structure appears, the possible contribution of segmentation edge cases or other imaging-related failures cannot be completely excluded relative to genuine biological heterogeneity. As such, Figures 4.28 to 4.30 show the preparation dependence of each metric in a compact form, while the histogram sections remain necessary for detailed distribution-level interpretation.

With the preparation dependence of the DAPI-derived measures established, the additional FIB channel can now be considered directly. Under the DAPI+FIB branch of the workflow, nuclei were defined exclusively from the DAPI channel and the resulting DAPI-derived masks were transferred unchanged to the paired FIB image, such that all FIB and nucleolar measurements were treated as conditional intranuclear descriptors confined within the same nuclear geometry, rather than as independently segmented whole-field objects (Section 2.3.3, Figure 2.4, and Equations (2.16) to (2.19) and (2.21); [317]). The intensity-derived endpoints shown in Figure 4.31 therefore complement the nucleolar morphology descriptors in Figure 4.18 by quantifying, on a per-nucleus basis, both the total intranuclear fibrillarin signal I_{FIB} and the threshold-restricted nucleolar-assigned signal I_{nucleoli}, together with the nucleoli:DAPI ratio $R_{\mathrm{N:D}}$ (Equations (2.9), (2.17), (2.21) and (2.24)). This distinction

is important because, as discussed in Section 2.3.3, the FIB channel exhibited some dim background nuclear staining, and accordingly, I_{nucleoli} provided a more specific measure of nucleolar-assigned signal, whereas I_{FIB} captured the full FIB signal within each DAPI-defined nucleus. Throughout, the reported mode denotes the peak position of the weighted kernel density estimate (Section 2.3.5 and Equation (2.34); [258, 316, 337]), while error bars represent standard error (calculated using the delta method where required; Equations (2.28) and (2.29)).

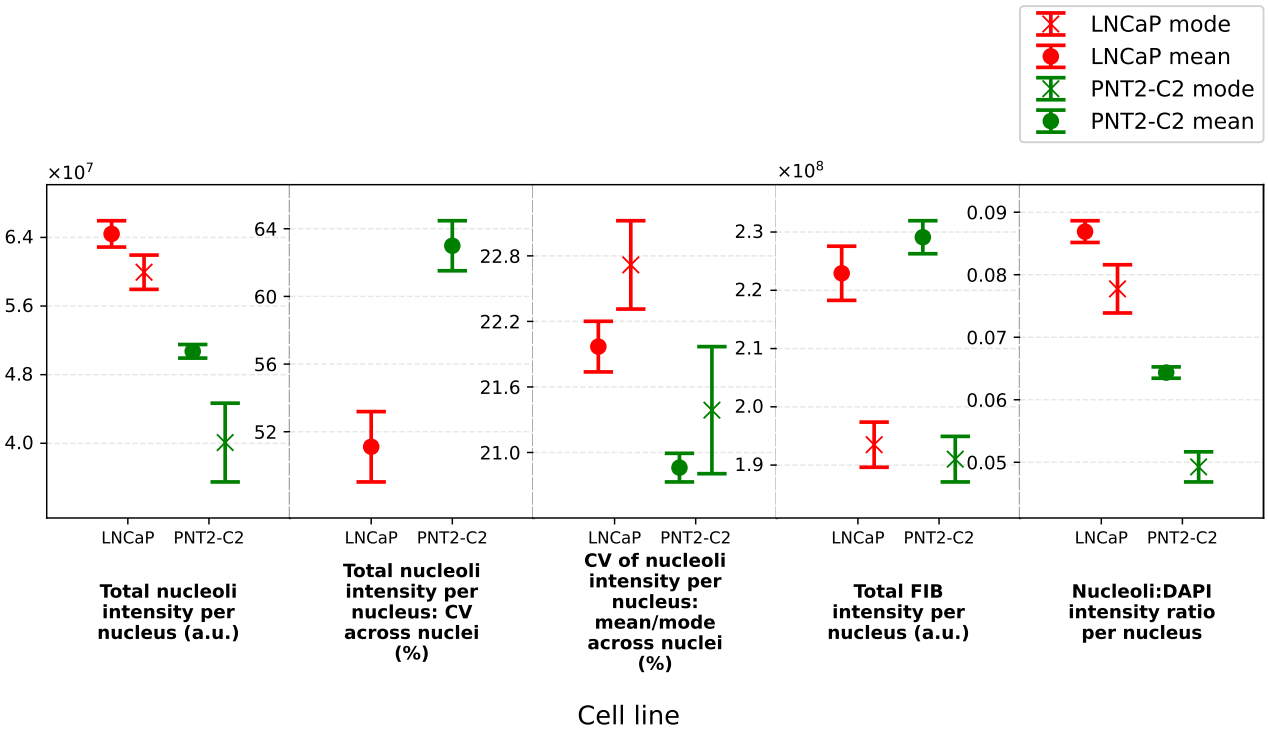


Figure 4.31: Plot showing the mean, mode and coefficient of variation (CV) of nucleolar and FIB-channel intensity measures in permeabilised cells stained with DAPI and anti-fibrillarin (DAPI+FIB). Total nucleoli intensity per nucleus was the sum of all pixel intensities in the FIB channel above the second Otsu threshold within the nuclei mask. The CV of nucleoli intensity per nucleus was the coefficient of variation of all nucleolar pixel intensity values for a given nucleus. Total FIB intensity was the sum of all pixel intensities in the FIB channel within the nuclei mask without any thresholding or processing, and the nucleoli:DAPI intensity ratio was the total nucleoli intensity divided by the total DAPI intensity for a given nucleus. Mode here refers to the peak position on the x-axis of a kernel density estimate for each metric (see Sections 2.3.5 and 4.2.2.2). Error bars represent standard error (calculated using the delta method for CV across nuclei and modes; Equations (2.28) and (2.29)).

At the level of central tendency, the background-suppressed nucleolar signal was markedly elevated in the malignant model. LNCaP exhibited a $\sim 25\%$ higher mean I_{nucleoli} than PNT2-C2 ($(6.4 \pm 0.2) \times 10^7$ a.u. vs. $(5.07 \pm 0.08) \times 10^7$ a.u.), while the KDE modes showed an even greater degree of separation, with the modal value being $\sim 50\%$ higher in LNCaP ($(6.0 \pm 0.2) \times 10^7$ a.u. vs. $(4.0 \pm 0.5) \times 10^7$ a.u.). This constitutes a key result relative to the nucleolar morphology findings; absolute nucleolar area was essentially indistinguishable between cell lines within uncertainty (Figure 4.18), yet the integrated nucleolar signal was substantially higher in LNCaP. The combination of comparable nucleolar area and elevated I_{nucleoli} implies a greater typical level of nucleolar FIB signal per unit area in LNCaP, consistent with an increase in fibrillarin-enriched dense fibrillar component content and/or a higher local concentration of nucleolar machinery, rather than a simple enlargement of the segmented nucleolar area. This is a mechanistically plausible distinction; fibrillarin localisation is compartment specific and nucleolar prominence in prostate malignancy does not always manifest as a proportional expansion of the fibrillarin-rich compartment alone [17, 63, 219, 224].

Importantly, this change was not mirrored by the raw, unthresholded nuclear FIB intensity, I_{FIB}, both the mean and modal values of which did not differ outside of uncertainty ($(2.23 \pm 0.05) \times 10^8$ a.u. and $(1.94 \pm 0.04) \times 10^8$ a.u. in LNCaP vs. $(2.29 \pm 0.03) \times 10^8$ a.u. and $(1.91 \pm 0.04) \times 10^8$ a.u. in PNT2-C2). As such, the principal difference between cell lines is not the total within-nucleus FIB signal, but rather the fraction of that signal which is specifically assigned to nucleolar pixels following background suppression. Considering this explicitly, by defining a derived "nucleolar fraction" as $I_{\mathrm{nucleoli}}/I_{\mathrm{FIB}}$, the ratio of the reported mean values yields was 0.289 in LNCaP and 0.221 in PNT2-C2, representing a $\sim 30\%$ relative increase in LNCaP. While this is not itself a primary endpoint, it provides a clear summary of the observed decoupling; in LNCaP a larger fraction of the nuclear FIB signal is concentrated into nucleolar-segmented pixels. In the context of prostate cancer, such a redistribution is compatible with increased ribosomal biogenesis demand and nucleolar functional upregulation in malignant cells [72, 136, 138, 141, 429].

Normalising the nucleolar signal to the DAPI channel further strengthens the interpretation that the effect is not a trivial consequence of global intensity scaling; the nucleoli:DAPI ratio $R_{N:D}$ is substantially higher in LNCaP than PNT2-C2 at both mean and mode (0.087 ± 0.002 and 0.078 ± 0.004 vs. 0.0644 ± 0.0009 and 0.049 ± 0.002), with the mean being $\sim 35\%$ higher in LNCaP, and the mode being $\sim 60\%$ higher. This is consistent with the joint behaviour of both nucleolar intensity and DAPI intensity separately; under DAPI+FIB, I_{nucleoli} is elevated in LNCaP, while I_{DAPI} is higher in PNT2-C2 (Section 4.2.2.2), and as such $R_{N:D}$ accentuates a nucleolar specific contrast between cell lines. Conceptually, this ratio can be viewed as an intranuclear normalisation that partially controls for inter-nuclear staining variability, and under the assumption that DAPI provides a stable proxy for DNA content, provides an interpretable measure of "nucleolar investment per nuclear DNA signal". The observed elevation of $R_{N:D}$ in LNCaP is therefore consistent with the established association between prostate malignancy and heightened nucleolar activity programmes [17, 63, 212].

Dispersion measures, however, indicate that the non-malignant model of PNT2-C2 remains more heterogeneous across nuclei for nucleolar signal magnitude. The CV of I_{nucleoli} across nuclei was higher in PNT2-C2 than LNCaP ($(63 \pm 1)\,\%$ vs. $(51 \pm 2)\,\%$), with a proportionally larger mean-mode separation in PNT2-C2 for both I_{nucleoli} and $R_{N:D}$ (i.e. a heavier right tail). This mirrors the results from Figure 4.18, where PNT2-C2 displayed a larger dispersion in nucleolar area and occupancy, and supports a consistent picture in which PNT2-C2 contains a broader mixture of nucleolar states, including a minority of high intensity events. A purely biological interpretation is not uniquely supported for this conclusion, as dispersion measures were also sensitive to the various modes of segmentation failures discussed previously.

Finally, the intranuclear texture descriptor, $\mathrm{CV}_{\mathrm{nucleoli}}$ was broadly similar between cell lines, but remained modestly higher in LNCaP ($\sim 5\%$ higher than PNT2-C2 in both mean and mode). This suggests that, within the segmented nucleolar regions, LNCaP nucleoli exhibit a slightly greater intranucleolar intensity heterogeneity, compatible with altered organisation and/or compartmentalisation of nucleolar components in malignant cells [17]. Taken together, the nucleolar intensity metrics provide a discriminative and mechanistically interpretable complement to the morphological results; although nucleolar area alone was not sufficient to separate the cell lines, LNCaP shows a clear elevation in nucleolar FIB signal, and in nucleolar signal normalised to DAPI, consistent with an increase in nucleolar functional activity without requiring gross nucleolar enlargement under the present operational definition.

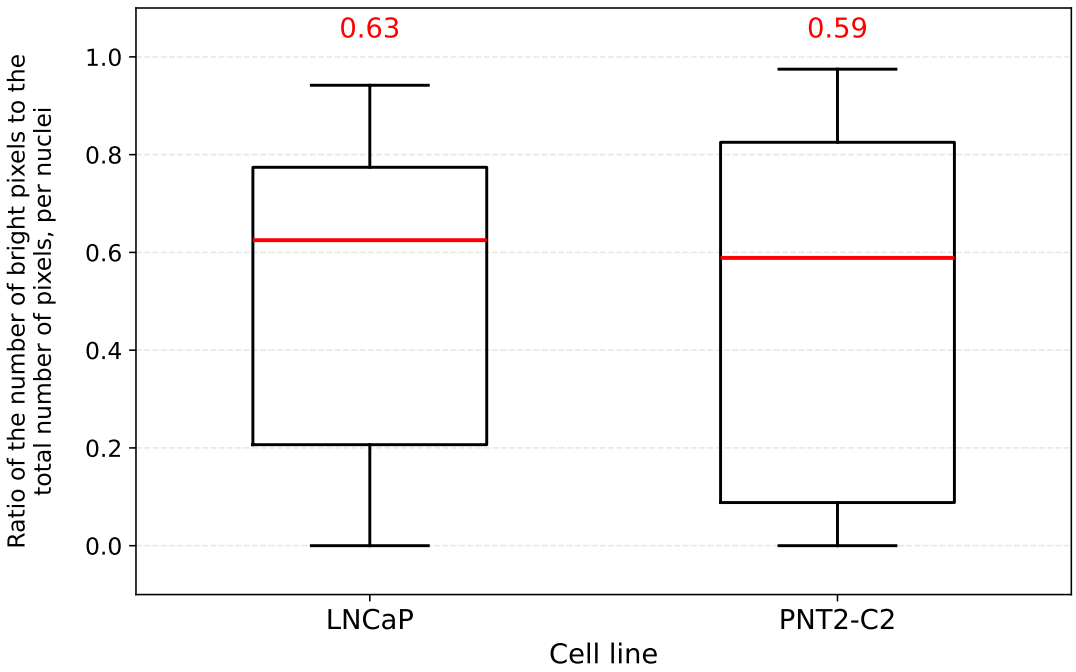
Bright/dark ratio box plots

As in Section 4.1.2.2, the bright and dark ratios (R_{bright} and R_{dark}) quantify the fraction of nuclear (DAPI) pixels above a fixed high-intensity threshold (bright, $I > 2.2000 \times 10^4$ a.u.) or below a fixed low-intensity threshold (dark, $I < 1.0000 \times 10^4$ a.u.) per nucleus. These threshold-based metrics are intrinsically sensitive to global intensity scaling and segmentation boundaries, and as such, comparison across conditions assumes that threshold placement remains meaningful and should be interpreted within the context of the bulk intensity summaries (Figures 4.7 and 4.27).

The bright/dark ratio box plots for the DAPI+FIB acquisition (Figure 4.32) indicate a substantial redistribution of DAPI pixel intensities across the fixed bright and dark thresholds when compared with the permeabilised DAPI-only reference (Figure 4.8). For the bright fraction, the median (IQR) increased from 35.1% (4.3% – 64.4%) to 62.5% (20.7% – 77.4%) in LNCaP, corresponding to a $\sim 80\%$ relative increase in the median, and increased from an effectively zero median in PNT2-C2 of 0.4% (0% – 25.2%) to 58.9% (8.8% – 82.5%). Here, the absolute shift of +58.5 % is the more interpretable effect size, given the near-zero baseline. Notably, the dispersion of R_{bright} behaved differently across cell lines; in LNCaP the IQR decreased slightly (from 60.1% to 56.7%), whereas in PNT2-C2 it expanded markedly (from 25.2% to 73.7%), indicating that the DAPI+FIB acquisition yields a much broader spread of bright-pixel fractions in PNT2-C2. Consequently, the large median separation in bright fractions between cell lines observed under permeabilised DAPI-only conditions (with LNCaP $\gg$ PNT2-C2) was largely eroded under DAPI+FIB staining, with medians becoming comparable (LNCaP 62.5% vs. PNT2-C2 58.9%).

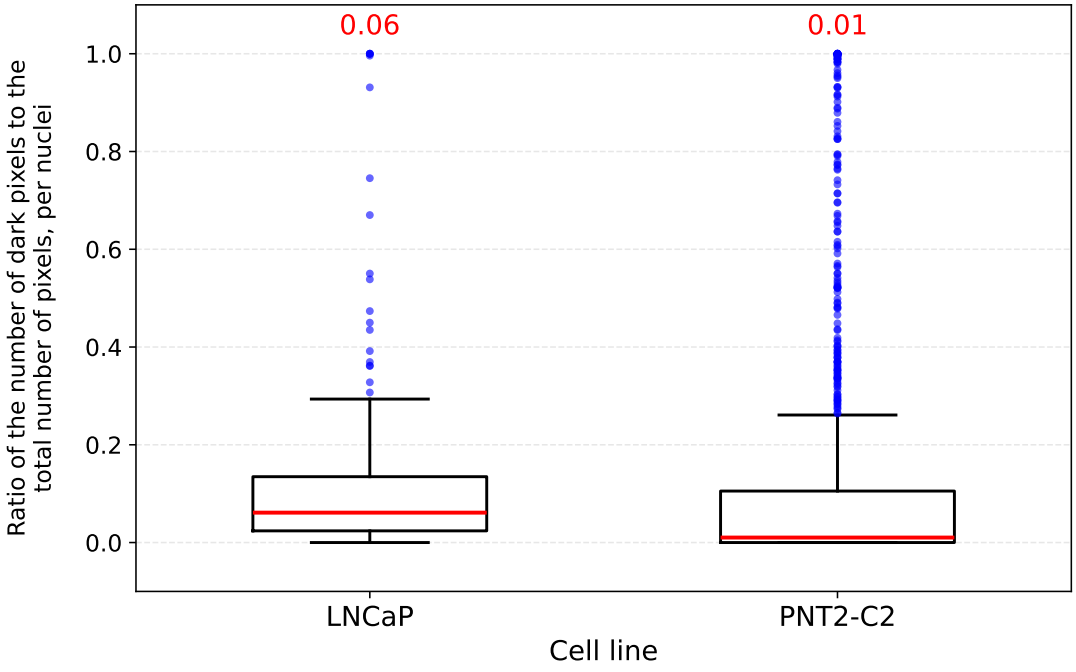
Changes in the dark fraction R_{dark} were directionally consistent with the above shift towards higher apparent DAPI intensities, but differed in magnitude between cell lines. In LNCaP, the median (IQR) decreased modestly from 7.2% (3.9% – 13.1%) to 6.1% (2.4% – 13.5%), a $\sim 15\%$ relative decrease, with a slight broadening of the IQR (from 9.2% to 11.1%). In contrast, PNT2-C2 displayed a pronounced reduction in R_{dark}, with the median (IQR) decreasing from 15.5% (4.2% – 43.4%) to 1.0% (0% – 10.5%), i.e. an absolute change of -14.5% (or a $\sim 95\%$ relative decrease), accompanied by a strong collapse of the IQR (from 39.2% to 10.5%). This behaviour reverses the DAPI-only ordering (where PNT2-C2 had a greater fraction of dark pixels relative to LNCaP) such that following FIB staining, PNT2-C2 exhibited a substantially smaller dark-pixel fraction than LNCaP.

The threshold-based redistributions should be interpreted cautiously; in the previous bulk analysis of DAPI intensity metrics (Figure 4.27), the DAPI+FIB acquisition showed clear evidence of differences in intensity scaling relative to the permeabilised DAPI-only dataset (notably, a higher total DAPI intensity and DAPI density per nucleus in PNT2-C2 under DAPI+FIB), implying that the fixed absolute thresholds used here are not guaranteed to interrogate comparable regions of the underlying pixel-intensity distributions across conditions. As such, the most operationally defensible inference from Figure 4.32 is that the DAPI+FIB DAPI channel is not directly comparable in terms of absolute intensity to the DAPI-only permeabilised reference, and that the large shifts (particularly the strong increase in R_{bright} and collapse of R_{dark}) in PNT2-C2 are more consistent with preparation-dependent intensity scaling than with a specific biological redistribution of chromatin; conclusions which are compatible with the shifts towards greater absolute DAPI intensity and density seen previously in this section (with relative increases of $\sim 55\%$ in both DAPI intensity and density for PNT2-C2).



LNCaP: Q1=0.21, Q3=0.77, IQR=0.57, L.Out=0, H.Out=0
PNT2-C2: Q1=0.09, Q3=0.83, IQR=0.74, L.Out=0, H.Out=0

(a) Boxplot showing the ratio of the number of bright pixels to the total number of pixels in each nucleus in permeabilised LNCaP (left) and PNT2-C2 (right) cells stained with DAPI and anti-fibrillarin (DAPI+FIB). The median is reported in red above each plot, with the quartiles, IQR, and number of outliers at each end reported underneath. Bright pixels are defined as pixels with an intensity greater than 2.2000×10^4 a.u.



LNCaP: Q1=0.02, Q3=0.13, IQR=0.11, L.Out=0, H.Out=19
PNT2-C2: Q1=0.00, Q3=0.11, IQR=0.11, L.Out=0, H.Out=157

(b) Boxplot showing the ratio of the number of dark pixels to the total number of pixels in each nucleus in permeabilised LNCaP (left) and PNT2-C2 (right) cells stained with DAPI and anti-fibrillarin (DAPI+FIB). The median is reported in red above each plot, with the quartiles, IQR, and number of outliers at each end reported underneath. Dark pixels are defined as pixels with an intensity lower than 1.0000×10^4 a.u.

Figure 4.32: Box plots showing bright and dark ratios (R_{bright} and R_{dark}) of permeabilised LNCaP (left) and PNT2-C2 (right) cells stained with DAPI and anti-fibrillarin (DAPI+FIB). Outliers are overlaid as blue dots.

Analysis of histograms

The per-nucleus histograms which follow provide a distribution-level complement to the bulk DAPI intensity summaries and bright/dark ratio box plots reported earlier (Figures 4.27 and 4.32). As in the DAPI-only baseline (Section 4.1.2.2) per-nucleus histograms were studied as mean values alone can hide inter and intranuclear heterogeneity, tail behaviour, and the presence of any subpopulations [174]. Throughout, analysis is based on the KDE-defined structure rather than the histogram, with the fitted parametric forms used as compact descriptors of the skewness and (where present) subpopulation structure.

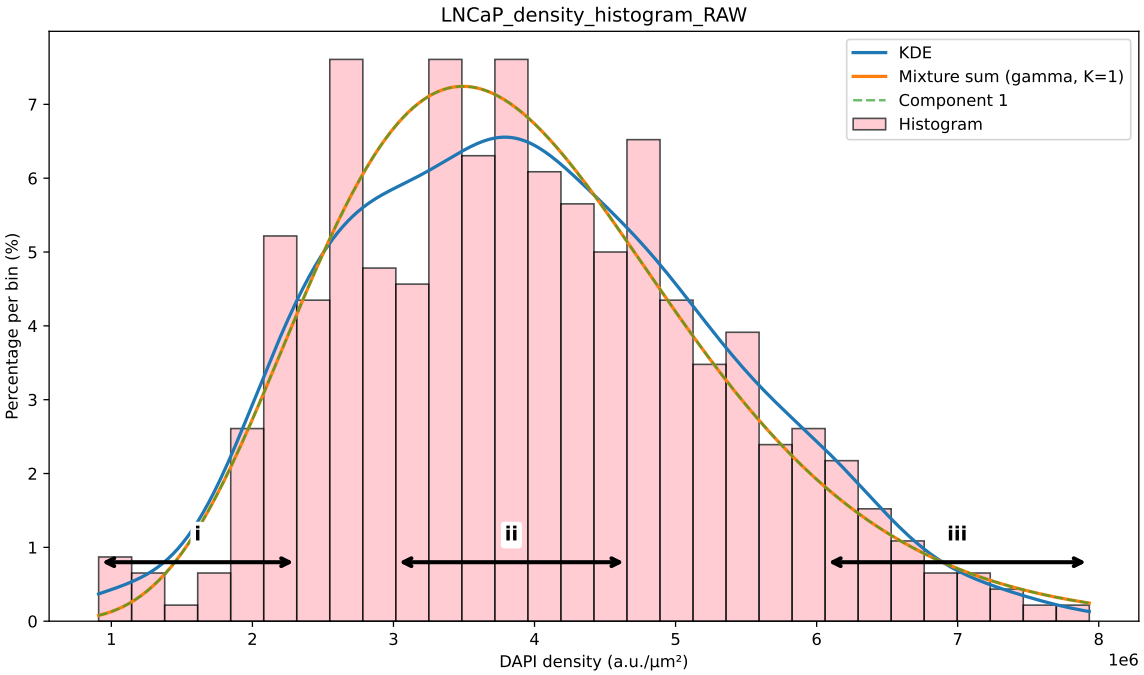


Figure 4.33: Histogram showing the DAPI density in permeabilised LNCaP cells stained with DAPI and anti-fibrillarin (DAPI+FIB). A KDE is overlaid on top of the histogram, with fitting applied to the KDE to determine the overall distribution of the data [258]. The best model (chosen by ICL [259]) was a single Gamma distribution with a $= 8.5 \pm 0.1$, loc $= 0$, and scale $= (4.66 \pm 0.01) \times 10^5$ with a peak at $(3.49 \pm 0.06) \times 10^6$ a.u. µm^{-2}. The KDE peak location was $(3.8 \pm 0.2) \times 10^6$ a.u. µm^{-2}. Uncertainties are calculated from the covariance matrix.

In LNCaP, the DAPI density distribution (Figure 4.33) was well described by a single Gamma distribution, with a mode of $(3.8 \pm 0.2) \times 10^6$ a.u. µm^{-2}. Relative to the permeabilised DAPI-only reference (Figure 4.9), this represents a modest increase in the typical density scale ($\sim 12\%$ relative to the DAPI-only mode of $(3.4 \pm 0.1) \times 10^6$ a.u. µm^{-2}). The most noticeable change, however, was in tail behaviour; $\sim 7.5\%$ of LNCaP nuclei exceeded 6×10^6 a.u. µm^{-2}, with a 90th percentile of 5.81×10^6a.u. µm^{-2}, a marked increase compared to the permeabilised DAPI-only dataset, where the $>6 \times 10^6$ a.u. µm^{-2} fraction was $\sim 1\%$ and the 90th percentile was 4.7×10^6 a.u. µm^{-2}. This is directionally consistent with the increase in DAPI density observed following FIB staining in Figure 4.27, and indicates that the DAPI+FIB dataset contains a substantially enriched high density tail compared with the DAPI-only permeabilised reference, and therefore that density comparisons across conditions are highly sensitive to preparation dependent intensity scaling (as already indicated by the bright/dark ratio behaviour in Figure 4.32).

In PNT2-C2, the DAPI density histogram required a three-component Gaussian mixture to properly

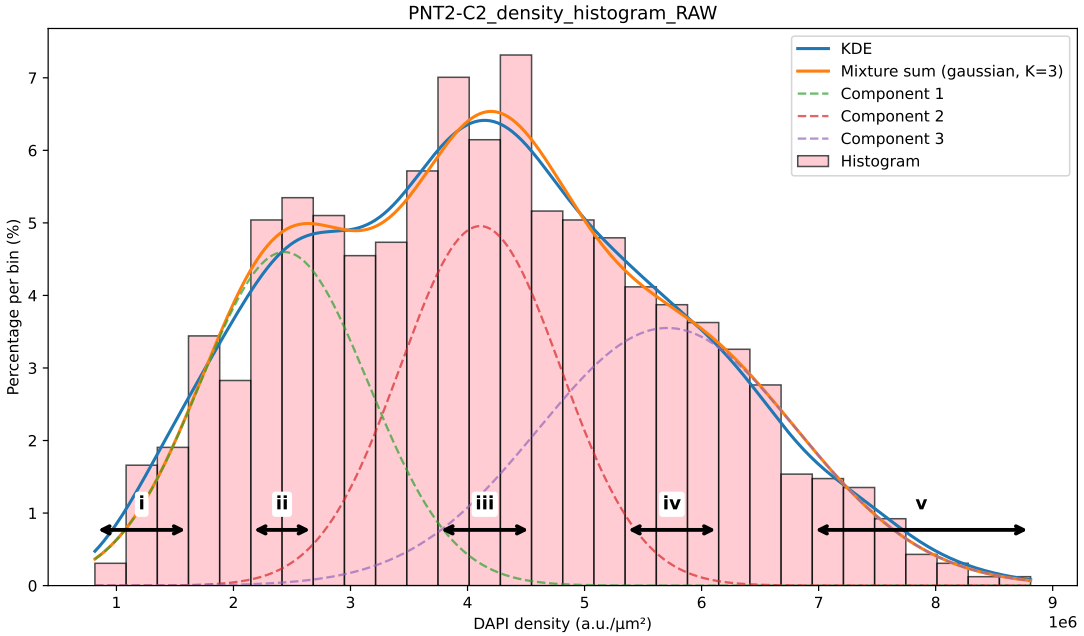


Figure 4.34: Histogram showing the DAPI density in permeabilised PNT2-C2 cells stained with DAPI and anti-fibrillarin (DAPI+FIB). A KDE is overlaid on top of the histogram, with fitting applied to the KDE to determine the overall distribution of the data [258]. The best model (chosen by ICL [259]) was a three component Gaussian distribution with $\mu_1 = (2.43 \pm 0.01) \times 10^6$, $\sigma_1 = (7.19 \pm 0.09) \times 10^5$ $\mu_2 = (4.108 \pm 0.009) \times 10^6$, $\sigma_2 = (6.87 \pm 0.09) \times 10^5$ $\mu_3 = (5.71 \pm 0.01) \times 10^6$, and $\sigma_3 = (1.10 \pm 0.01) \times 10^6$. Peaks were located at $(2.43 \pm 0.01) \times 10^6$ a.u. µm^{-2}, $(4.108 \pm 0.009) \times 10^6$ a.u. µm^{-2}, and $(5.71 \pm 0.01) \times 10^6$ a.u. µm^{-2} respectively. The KDE peak location was $(4.1 \pm 0.1) \times 10^6$ a.u. µm^{-2}. Uncertainties are calculated from the covariance matrix.

characterise the distribution, with fitted peaks at $(2.43 \pm 0.01) \times 10^6$ a.u. µm^{-2}, $(4.108 \pm 0.009) \times 10^6$ a.u. µm^{-2}, and $(5.71 \pm 0.01) \times 10^6$ a.u. µm^{-2}, and a KDE mode at $(4.1 \pm 0.1) \times 10^6$ a.u. µm^{-2} (Figure 4.34). Component weights (i.e. the proportion of data contained within each peak) were approximately equal (0.311, 0.320, and 0.369 respectively from left to right), however the prominence of each peak within the summed fit is a result of both weight and width. The second peak (iii) was comparatively narrower (and therefore taller) than the third (iv), despite having a slightly lower weight, and combined with overlap between the tails of the other two components (ii and iv), resulted in the second peak most closely reflecting the KDE mode.

This represents a considerable change relative to the DAPI-only baseline, where PNT2-C2 density was unimodal, and best described by an inverse-Gaussian form with a substantially lower mode of $(2.1 \pm 0.1) \times 10^6$ a.u. µm^{-2} (Figure 4.10); the modal DAPI density approximately doubled following FIB staining. In parallel, the high-density fraction was strongly elevated, with $\sim$ 15% of nuclei exceeding 6×10^6 a.u. µm^{-2} (compared to $\sim$ 1% in DAPI-only), and an increased 90th percentile of 6.37×10^6a.u. µm^{-2} (compared to 4.23×10^6a.u. µm^{-2}). Importantly, this is not interpreted as an increase in tailedness, the mean and mode were within uncertainty of one another, and so this instead reflects the global scaling of PNT2-C2 nuclei towards higher DAPI density. The emergence of explicit multi-peak structure in PNT2-C2 under DAPI+FIB is consistent with either an increase in inter-nuclear heterogeneity in the effective density scale (e.g. multiple preparation/batch dependent intensity regimes), and/or a genuine mixture of biological subpopulations. The present analysis cannot distinguish between these possibilities, but interpreted in the context of the other results in this section, the more likely cause is some preparation related change/experimental variation, rather than a meaningful biological difference.

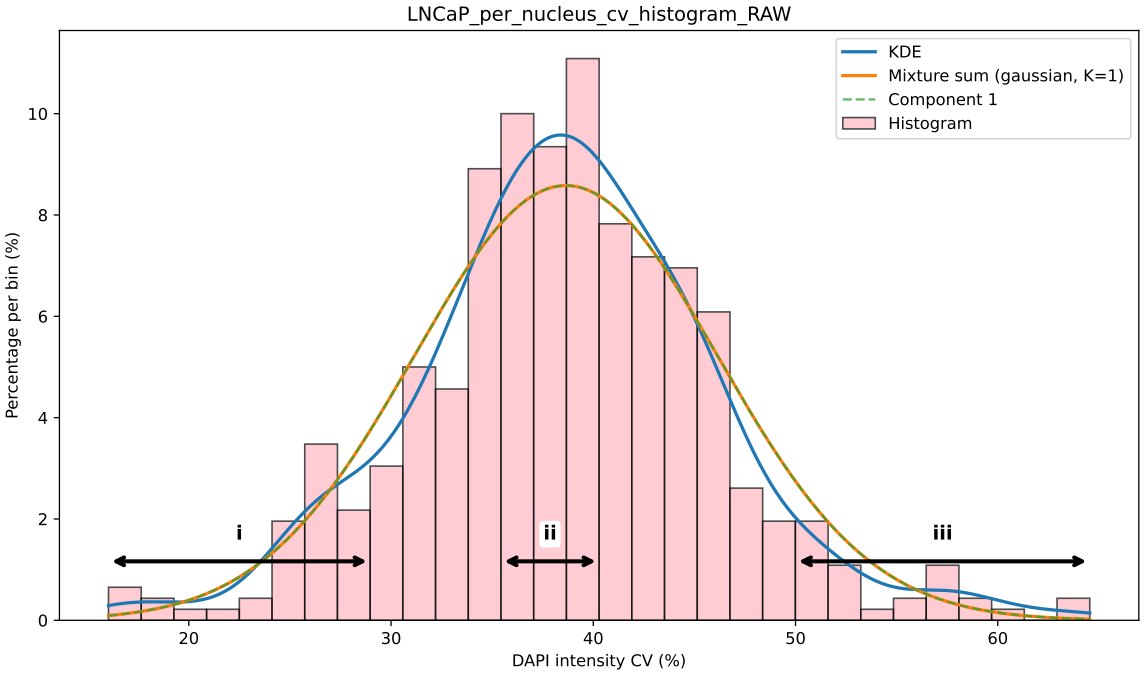


Figure 4.35: Histogram showing the CV of DAPI intensity in permeabilised LNCaP cells stained with DAPI and anti-fibrillarin (DAPI+FIB). A KDE is overlaid on top of the histogram, with fitting applied to the KDE to determine the overall distribution of the data [258]. The best model (chosen by ICL [259]) was a single Gaussian distribution with $\mu = 38.6 \pm 0.4$ and $\sigma = 7.5 \pm 0.2$, with a peak at $(38.6 \pm 0.4)\,\%$. The KDE peak location was $(38.4 \pm 0.7)\,\%$. Uncertainties are calculated from the covariance matrix.

For LNCaP, the CV of DAPI intensity (CV_{DAPI}) (Figure 4.35) was unimodal and well described by a single Gaussian, with a mode of $(38.4 \pm 0.7)\,\%$. This contrasts with the permeabilised DAPI-only reference (Figure 4.11), where CV_{DAPI} was bimodal and the KDE mode was representative of the lower-CV subpopulation ($(29.7 \pm 0.6)\,\%$). Quantitatively, the typical CV scale therefore increased by $\sim 30\%$ under FIB staining. Tail behaviour also shifted, while CV had a slight positive skew in the DAPI-only dataset due to the presence of the higher-CV subpopulation, the observed distribution for DAPI+FIB data was approximately symmetric about the mode. The high-CV fractions therefore more readily represent a shift in central tendency than tailedness, with $\sim 40\%$ of nuclei with $CV_{DAPI} > 40\%$, and a 90th percentile of 47.7%, compared with $\sim 30\%$ and 46.8% respectively in the DAPI-only permeabilised dataset. Taken together, these changes indicate that the DAPI+FIB LNCaP nuclei are, on average, more texture-rich (i.e. have a higher intranuclear contrast) than in the matched DAPI-only permeabilised reference, with less complex structure (justified by the loss of separable subpopulations), however, given the independent evidence for the lack of comparability of total absolute intensity across conditions, this is again interpreted primarily as experimental variation, rather than as indicating any biological changes.

In PNT2-C2, CV_{DAPI}, was best captured by a single skew-normal distribution, with KDE mode $(26.1 \pm 0.3)\,\%$ (Figure 4.36). This is close to the mode from the DAPI-only baseline of $(25.6 \pm 0.2)\,\%$ (Figure 4.12), implying that the typical PNT2-C2 nucleus retains a similar texture scale across the two permeabilised conditions. The overall shape of the distributions also appear broadly similar, with the exception of an enrichment of high-CV nuclei following FIB staining. While a very slight "bump" was observed in the 40 % to 50 % in Figure 4.12, this was much more prominent in Figure 4.36. This is reflected in the tail measures, with a substantially enlarged fraction of nuclei exceeding a CV of 40% ($\sim 15.5\%$ in DAPI+FIB vs. $\sim 5\%$ in DAPI-only), and a $sim20\%$ increase in 90th percentile (43.6% in

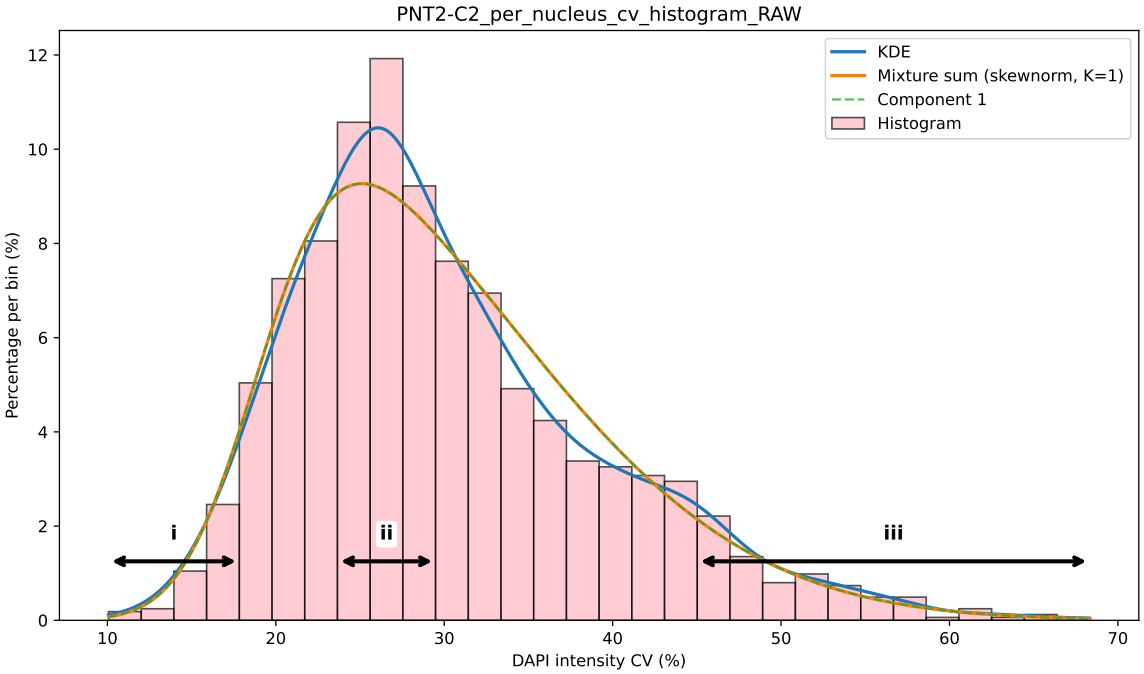


Figure 4.36: Histogram showing the CV of DAPI intensity in permeabilised PNT2-C2 cells stained with DAPI and anti-fibrillarin (DAPI+FIB). A KDE is overlaid on top of the histogram, with fitting applied to the KDE to determine the overall distribution of the data [258]. The best model (chosen by ICL [259]) was a single skew-normal distribution with a $= 3.9 \pm 0.3$, loc $= 19.0 \pm 0.3$, and scale $= 14.5 \pm 0.3$, with a peak at (25.1 ± 0.4) %. The KDE peak location was (26.1 ± 0.3) %. Uncertainties are calculated from the covariance matrix.

DAPI+FIB vs. 36.7% in DAPI-only). Thus, for CV_{DAPI} in PNT2-C2, the principal difference resulting from FIB staining was not a shift in central tendency, but an enrichment of highly heterogeneous nuclei.

Overall, the DAPI+FIB histograms show that relative to permeabilised DAPI-only staining, the DAPI intensity field is reshaped in a way that increases the prevalence of high-density nuclei in both cell lines, introduces clear subpopulation structure in PNT2-C2 DAPI density, and increases high CV_{DAPI} tail mass in the non-malignant model, whilst also shifting the LNCaP CV distribution towards a higher, unimodal texture regime. Given the independently observed bulk intensity scaling differences under DAPI+FIB (Figure 4.27) and the sensitivity of threshold-defined texture ratios (Figure 4.32), these distribution-level changes are interpreted as evidence that within this study, the DAPI channel following FIB staining was not directly comparable to that from cells stained with only DAPI, and its primary use here is therefore to provide contextual support for the nucleolar analyses, rather than as a standalone proxy for DNA content/chromatin state across conditions.

Now considering the nucleolar intensity histograms, the following provides a distribution-level analysis of several key measures derived from anti-fibrillarin (FIB) staining under the DAPI+FIB workflow. In contrast to bulk summaries in Figure 4.31, the histograms explicitly represent heterogeneity across individual nuclei, thus allowing for interrogation of skewness and meanmode separation, tail structure and rare-event prevalence, and evidence for/against the presence of any discrete subpopulations. As for the DAPI-based histograms, the KDE mode is used as the primary descriptor of the most probable state, while the parametric fits are treated as compact summaries of the dominant distributional form (summarising skewness, and, where required, mixture structure).

Both cell lines exhibit unimodal total nucleolar intensity, I_{nucleoli}, distributions with pronounced right

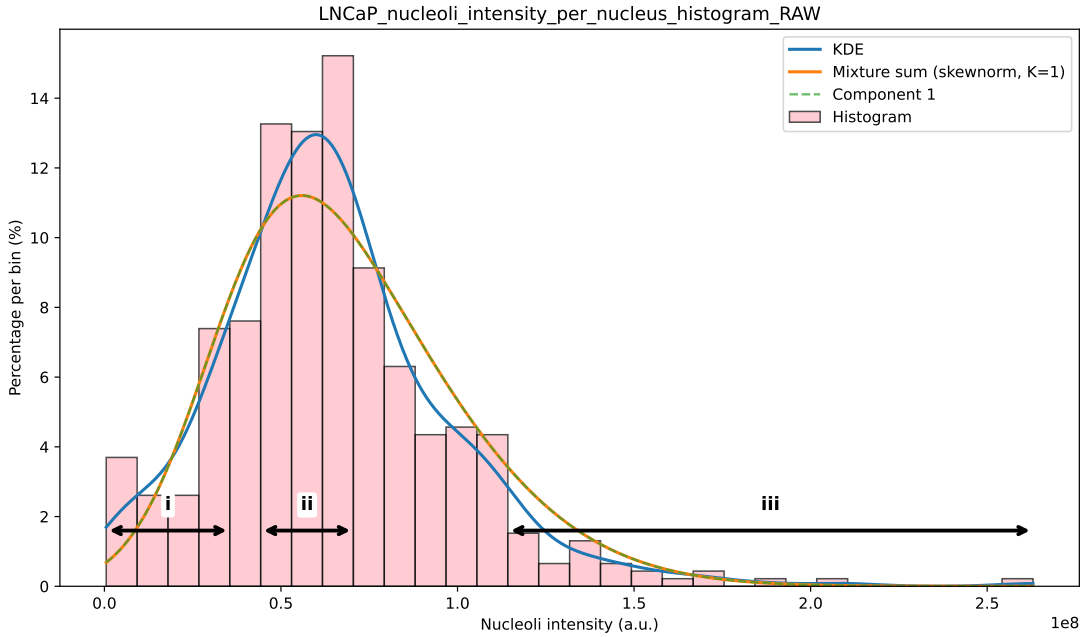


Figure 4.37: Histogram showing the total nucleoli intensity in permeabilised LNCaP cells stained with DAPI and anti-fibrillarin (DAPI+FIB). A KDE is overlaid on top of the histogram, with fitting applied to the KDE to determine the overall distribution of the data [258]. The best model (chosen by ICL [259]) was a single skew-normal distribution with a $= 2.5 \pm 0.2$, loc $= (3.1172 \pm 0.0002) \times 10^7$, and scale $= (4.9251 \pm 0.0002) \times 10^7$ with a peak at $(5.59 \pm 0.06) \times 10^7$ a.u.. The KDE peak location was $(6.0 \pm 0.2) \times 10^7$ a.u.. Uncertainties are calculated from the covariance matrix.

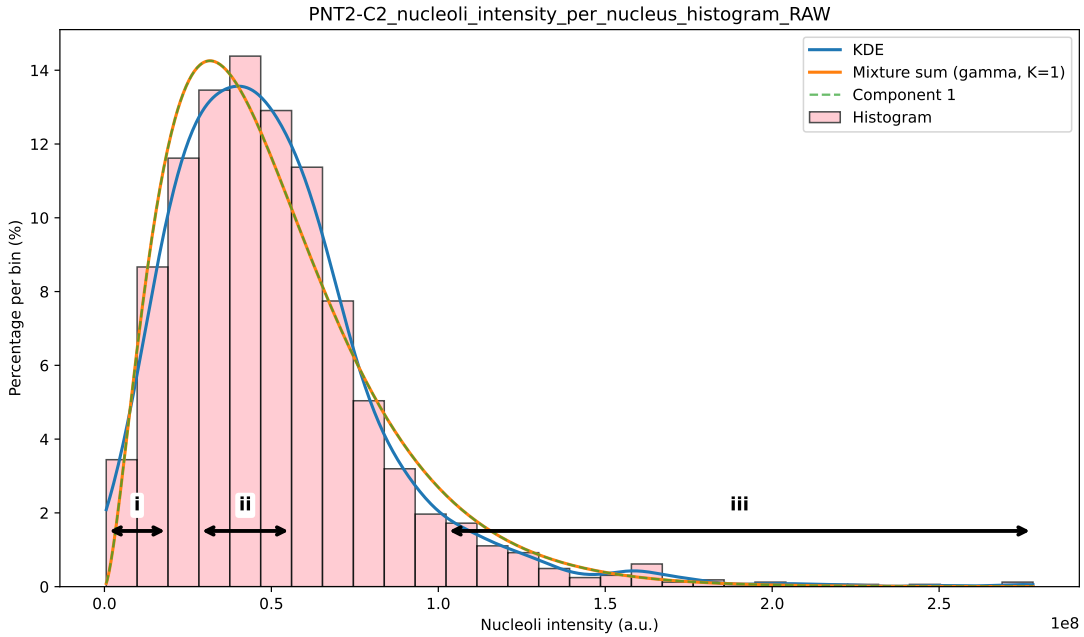


Figure 4.38: Histogram showing the total nucleoli intensity in permeabilised PNT2-C2 cells stained with DAPI and anti-fibrillarin (DAPI+FIB). A KDE is overlaid on top of the histogram, with fitting applied to the KDE to determine the overall distribution of the data [258]. The best model (chosen by ICL [259]) was a single Gamma distribution with a $= 2.65 \pm 0.04$, loc $= 0$, and scale $= (1.918\,73 \pm 0.000\,04) \times 10^7$, with a peak at $(3.16 \pm 0.07) \times 10^7$ a.u.. The KDE peak location was $(4.0 \pm 0.5) \times 10^7$ a.u.. Uncertainties are calculated from the covariance matrix.

tails (Figures 4.37 and 4.38), consistent with a continuum of nucleolar signal states rather than a predominantly bimodal or otherwise meaningfully separable process. Visually, the LNCaP histogram is centred at a slightly higher intensity, and shows a much less pronounced tail than in PNT2-C2. Comparing parametric forms, LNCaP was best described by a skew-normal distribution, with a KDE mode of $(6.0 \pm 0.2) \times 10^7$ a.u.. In contrast, PNT2-C2 was best described by a Gamma distribution with a KDE mode of $(4.0 \pm 0.5) \times 10^7$ a.u.. The use of a Gamma form in PNT2-C2 is consistent with a positive, strongly right-skewed distribution, whereas the skew-normal form in LNCaP indicates a distribution that is closer to a single central location, however still displays a notable positive skew. Practically, this difference forms a concise summary of the observations from a visual inspection; that PNT2-C2 appears more strongly skewed (with a heavier tail relative to the most probable value), than LNCaP.

As shown in Figure 4.31, LNCaP showed higher nucleolar signal than PNT2-C2 at the level of both mean and mode. Mean I_{nucleoli} was $(6.4 \pm 0.2) \times 10^7$ a.u. in LNCaP vs. $(5.07 \pm 0.08) \times 10^7$ a.u. in PNT2-C2 (a $\sim 30\%$ increase in LNCaP), while modal values show an even stronger separation of $\sim 50\%$ ($(6.0 \pm 0.2) \times 10^7$ a.u. vs. $(4.0 \pm 0.5) \times 10^7$ a.u.). In both cell lines the mean exceeded the mode, confirming positive skew, however, the mean-mode separation was substantially larger in PNT2-C2 ($\sim 26.5\%$) than in LNCaP ($\sim 7.5\%$). This implies that, relative to the typical nucleus, PNT2-C2 contains a proportionally more influential high-intensity tail; in other words, PNT2-C2 heterogeneity is more "tail-driven", even though the entire LNCaP distribution was shifted to higher intensities.

Tail summaries reinforce the conclusion that LNCaP carries a higher absolute burden of high-intensity nuclei, despite exhibiting less "tailedness" than PNT2-C2. Specifically, 13.3% of LNCaP nuclei had a nucleolar intensity greater than 1×10^8 a.u., with a 90th percentile of 1.07×10^8a.u., whereas only 6.87% of PNT2-C2 nuclei exceed 1×10^8 a.u., with a 90th percentile of 8.90×10^7a.u.. Normalising the 90th percentile to the mode reveals the larger relative tail extent in PNT2-C2, consistent with the larger mean-mode separation and higher inter-nuclear CV of I_{nucleoli} reported in Figure 4.31 previously ($(63 \pm 1)\,\%$ in PNT2-C2 vs. $(51 \pm 2)\,\%$ in LNCaP). The histograms indicate a biologically relevant set of observations; LNCaP exhibits a higher overall nucleolar signal state (evidenced by a greater mode and mean), while PNT2-C2 exhibits greater relative dispersion (expressed as greater tail leverage per unit typical signal).

Prominent nucleoli are a classical feature of prostatic adenocarcinoma, and are widely used diagnostically, reflecting (at minimum) altered nucleolar architecture and frequently increased nucleolar function [63, 72, 222]. Mechanistically, prostate neoplasia is strongly linked to activation of a nucleolar/ribosomal biogenesis programme, with MYC identified as a key driver, and importantly, fibrillarin is an MYC target required for proliferation and proper rRNA processing in multiple prostate cancer cell lines including LNCaP [17]. Within this framework, the global upwards shift of I_{nucleoli} in LNCaP provides quantitative support at the single nucleus level for a higher fibrillarin-enriched nucleolar state in the malignant model. Critically, this conclusion is distinct from and does not require concurrent nucleolar enlargement, therefore extending the findings beyond what is captured by morphology alone. In this dataset, nucleolar area was essentially identical between cell lines (Figure 4.18), yet I_{nucleoli} was substantially higher in LNCaP, implying an increase in mean nucleolar pixel intensity (analogous to nucleolar signal density since area was calculated based on pixel counts, i.e. a greater nucleolar signal per segmented nucleolar area), rather than an increase in segmented nucleolar footprint. This is a key point, and a novel finding relative to much of the prostate cancer nucleolar literature, (which

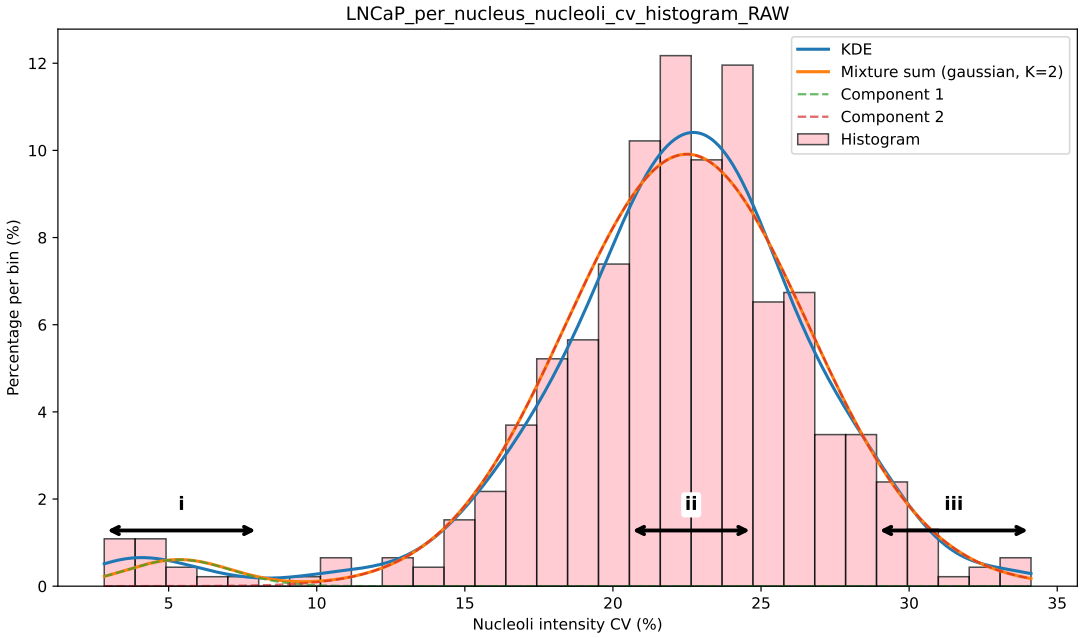
is dominated by tissue-level prominence scoring or morphometry such as nucleolar size/area, and therefore cannot distinguish "larger nucleoli" from "more fibrillarin dense nucleolar compartments") [140, 222, 430, 455].

A second novel insight is that the malignant model does not simply contain more extreme nucleoli, but shows a higher typical nucleolar intensity state, indicating that nucleolar upregulation in LNCaP is widespread across nuclei, and not confined to a small proliferative subfraction, whereas PNT2-C2 heterogeneity was more strongly tail driven. This type of distribution-level inference is largely absent from existing prostate cancer fluorescence and histology studies, which typically report mean expression or representative images without explicit quantification of tail fractions and mode structure [17, 139]

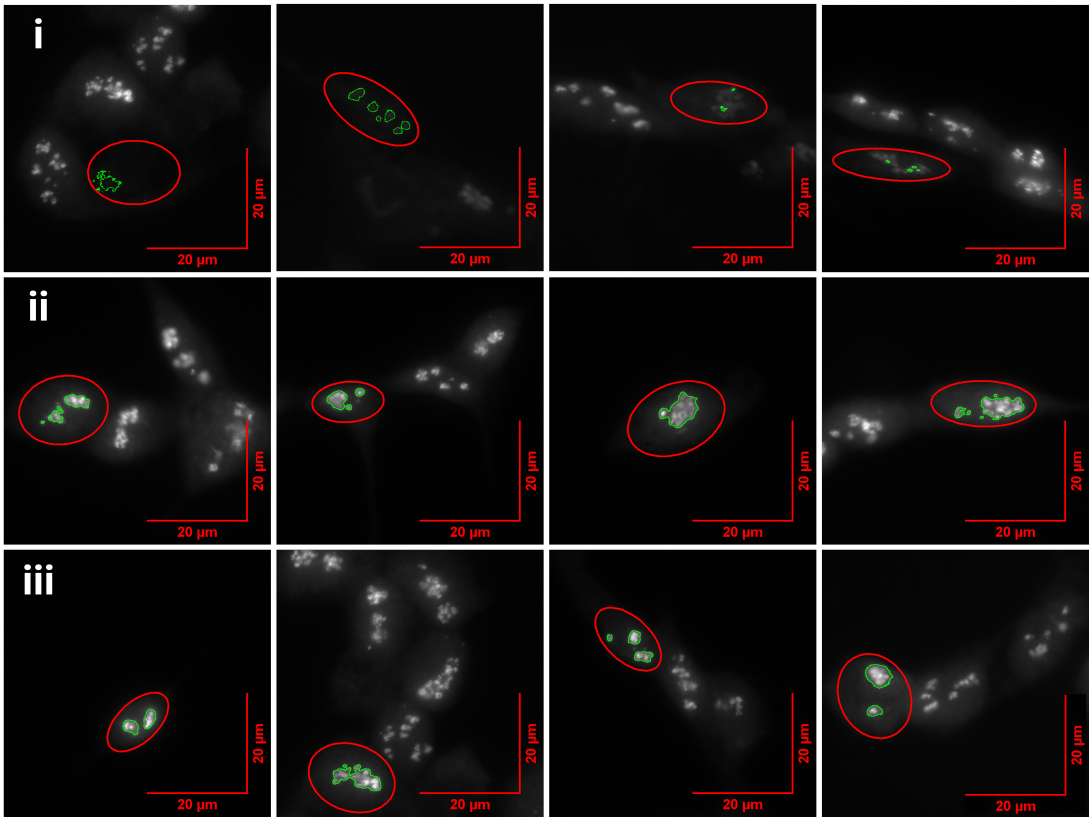
Moving to consider the intranucleolar texture, $CV_{nucleoli}$, histograms (Figures 4.39(a) and 4.40(a)), both had a broadly similar overall shape, with the notable exception of the visually evident low-CV bump in the far left tail of LNCaP (at $\sim 5\%$), which was not present in PNT2-C2, showing a smoother, unimodal form. This qualitative distinction is important, and suggests that LNCaP contains a small subset of nuclei with unusually uniform nucleolar pixel intensities (a discrete/separable low-CV subpopulation), while PNT2-C2 exhibited more continuous low-end dispersion rather than a distinct low-CV mode. In LNCaP, $CV_{nucleoli}$ required a two-component Gaussian mixture to properly characterise the distribution, with the first component containing $\sim 2.6\%$ of nuclei, centred at $(5.3 \pm 0.5)\,\%$. The second component contained the remaining $\sim 97.4\%$ of nuclei, and was centred at $(22.5 \pm 0.2)\,\%$. The KDE mode of $(22.7 \pm 0.4)\,\%$, closely mirrored the dominant component. In PNT2-C2, a single skew-normal distribution was sufficient, with a KDE mode at $(21.4 \pm 0.6)\,\%$. The negative skew parameter in PNT2-C2 is consistent with the visually heavier low-CV tail.

The mean nucleolar intensity CV was slightly higher in LNCaP than PNT2-C2 ($(22.0 \pm 0.2)\,\%$ vs. $(20.9 \pm 0.1)\,\%$), with LNCaP also showing a similarly higher KDE mode (22.7174 % vs. 21.3874 %). Tail descriptors indicate broadly similar high-end behaviour, but notable differences in low-end behaviour. LNCaP had 10.6% of nuclei with $CV_{nucleoli} > 27.5\%$, and a 90th percentile of 27.6%, while PNT2-C2 had 9.90% of nuclei with $CV_{nucleoli} > 27.5\%$, and a 90th percentile of27.5%. Considering now the low end, $CV_{nucleoli} < 12.5\%$, 6.38% of PNT2-C2 nuclei fell within this threshold, compared with 3.71% in LNCaP; correspondingly, the 10th percentile was lower in PNT2-C2 (13.9%) than in LNCaP (16.5%). This pattern supports a coherent summary; high $CV_{nucleoli}$ extremes were present at similar rates in both lines, but low $CV_{nucleoli}$ events were more prevalent in PNT2-C2, despite the small, practically homogeneous subpopulation in LNCaP.

Because $CV_{nucleoli}$ was computed over Otsu-segmented nucleolar pixels, extremely low CV values can arise either as a result of real biological outcomes (e.g. unusually homogeneous fibrillarin distribution within the segmented nucleolar compartment), or algorithmically (e.g. artificially small segmented regions leading to limited pixel sampling; i.e. segmentation that captures only a highly uniform subregion). The significant presence of segmentation failures observed in the representative images for the LNCaP subpopulation motivates the interpretation of the discrete low-CV subpopulation as a segmentation artefact, rather than a meaningful biological state, and the presence of this subpopulation in LNCaP and not PNT2-C2 further suggests that the lack of observed change in nucleolar area between cell lines is the result of a greater proportion of segmentation failures in LNCaP which specifically act to minimise the detected nucleolar region. One biologically plausible contributor to

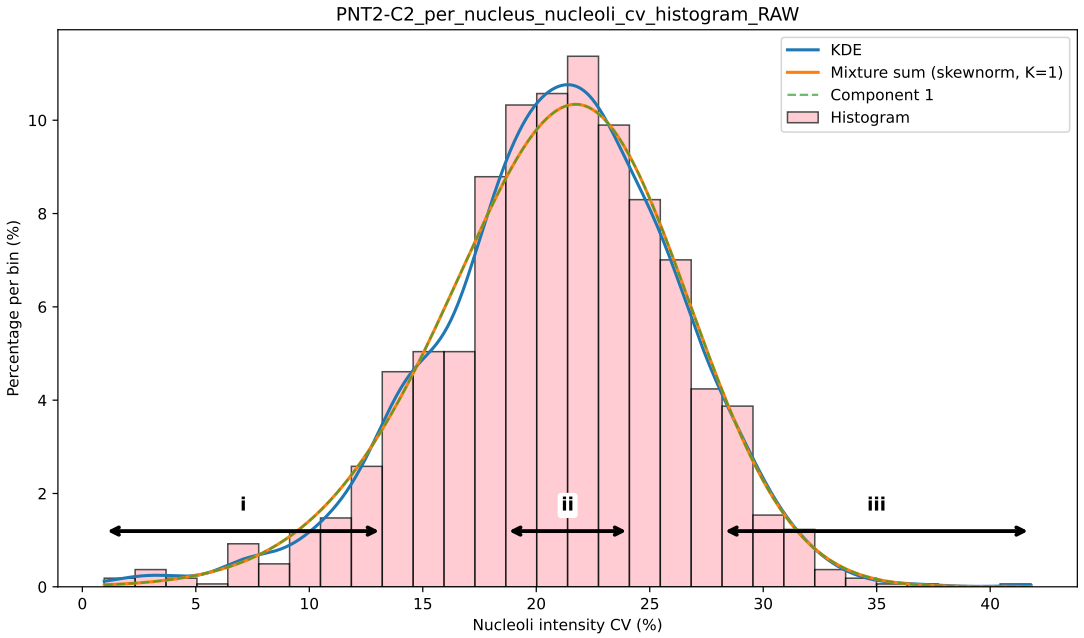


(a) Histogram showing the CV of nucleoli intensity in each LNCaP nucleus. A KDE is overlaid on top of the histogram, with fitting applied to the KDE to determine the overall distribution of the data [258]. The best model (chosen by ICL [259]) was a Gaussian distribution with two components, the first (containing $\sim 2.6\%$ of the data) had $\mu = 5.3 \pm 0.5$ and $\sigma = 1.8 \pm 0.4$, with a peak at $(5.3 \pm 0.5)\,\%$, and the second (containing $\sim 97.4\%$ of the data) had $\mu = 22.5 \pm 0.2$ and $\sigma = 4.1 \pm 0.1$, with a peak at $(22.5 \pm 0.2)\,\%$ The KDE peak location was $(22.7 \pm 0.4)\,\%$.

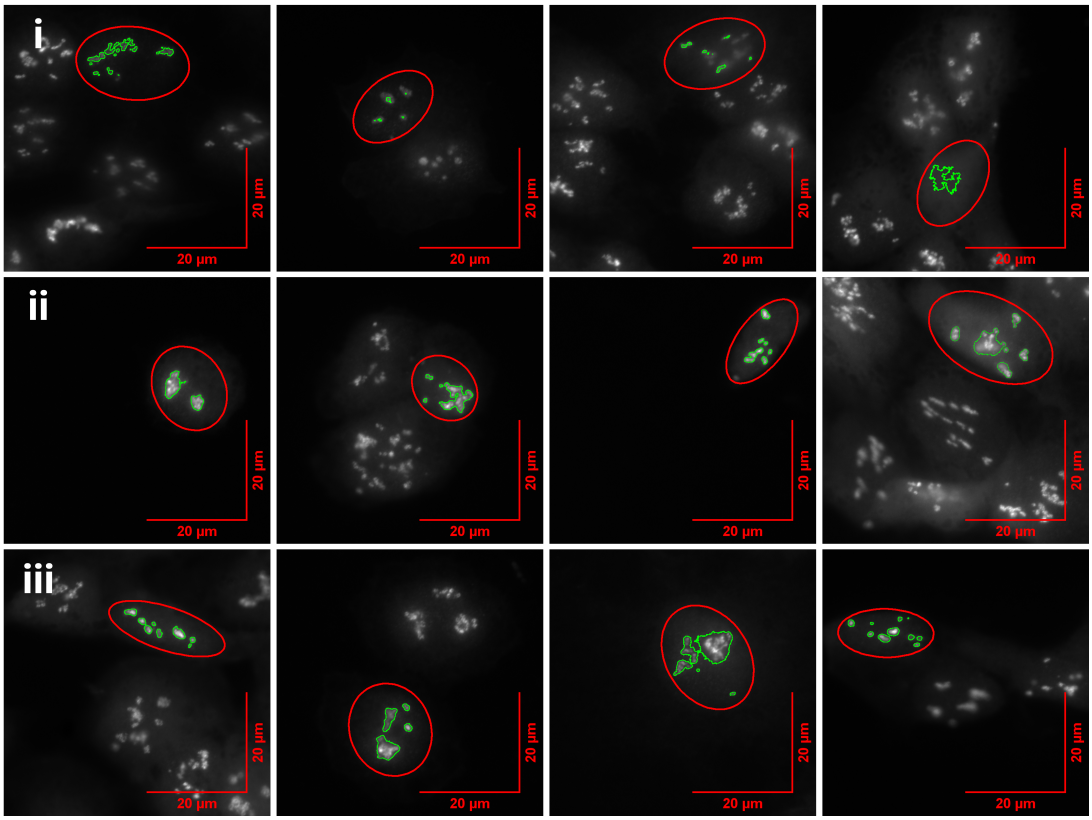


(b) Images of detected nuclei selected from the low/subpopulation (i), modal (ii), and high (iii) regions of the corresponding histogram. Red ellipses are drawn purely for illustrative purposes to show the approximate region which was detected as a nucleus by the segmentation code, with regions detected as nucleoli outlined in green. CV was calculated as 100 times the standard deviation of all nucleolar pixel intensities divided by the mean. A 20 µm scale bar is shown in red at the bottom right corner of each image.

Figure 4.39: Histogram showing the CV of nucleoli intensity in permeabilised LNCaP cells stained with DAPI and anti-fibrillarin (DAPI+FIB) ((a)), with some images of representative nuclei from the low, modal/first subpopulation, second subpopulation, and high regions of the histogram ((b)). The KDE is used as a basis for fitting, with the histogram being purely illustrative. The best fit was chosen using an Integrated Classification Likelihood (ICL) [259]. Uncertainties are calculated from the covariance matrix.



(a) Histogram showing the CV of nucleoli intensity in each PNT2-C2 nucleus. A KDE is overlaid on top of the histogram, with fitting applied to the KDE to determine the overall distribution of the data [258]. The best model (chosen by ICL [259]) was a single skew-normal distribution with a $= -1.5 \pm 0.2$, loc $= 25.6 \pm 0.3$, and scale $= 7.2 \pm 0.3$, with a peak at (21.7 ± 0.2) %. The KDE peak location was (21.4 ± 0.6) %.



(b) Images of detected nuclei selected from the low (i), modal (ii), and high (iii) regions of the corresponding histogram. Red ellipses are drawn purely for illustrative purposes to show the approximate region which was detected as a nucleus by the segmentation code, with regions detected as nucleoli outlined in green. CV was calculated as 100 times the standard deviation of all nucleolar pixel intensities divided by the mean. A 20 µm scale bar is shown in red at the bottom right corner of each image.

Figure 4.40: Histogram showing the CV of nucleoli intensity in permeabilised PNT2-C2 cells stained with DAPI and anti-fibrillarin (DAPI+FIB) ((a)), with some images of representative nuclei from the low, modal, and high regions of the histogram ((b)). The KDE is used as a basis for fitting, with the histogram being purely illustrative. The best fit was chosen using an Integrated Classification Likelihood (ICL) [259]. Uncertainties are calculated from the covariance matrix.

the extremely low nucleolar signal which contributes to the segmentation failure, however, is cell-state dependent reorganisation of nucleolar subcompartments. Fibrillarin localisation is known to change under nucleolar stress and during nucleolar reorganisation; such processes are routinely visualised by fibrillarin immunofluorescence [215, 220, 364, 452]. The present dataset does not directly investigate stress pathway, however, and so this assertion is presented simply as one potential biological route to the observed distribution. Irrespective of the cause/interpretation, this subpopulation would not be discoverable from bulk mean/mode values alone, again highlighting the importance of distribution-level analysis to uncover important information beyond summary statistics.

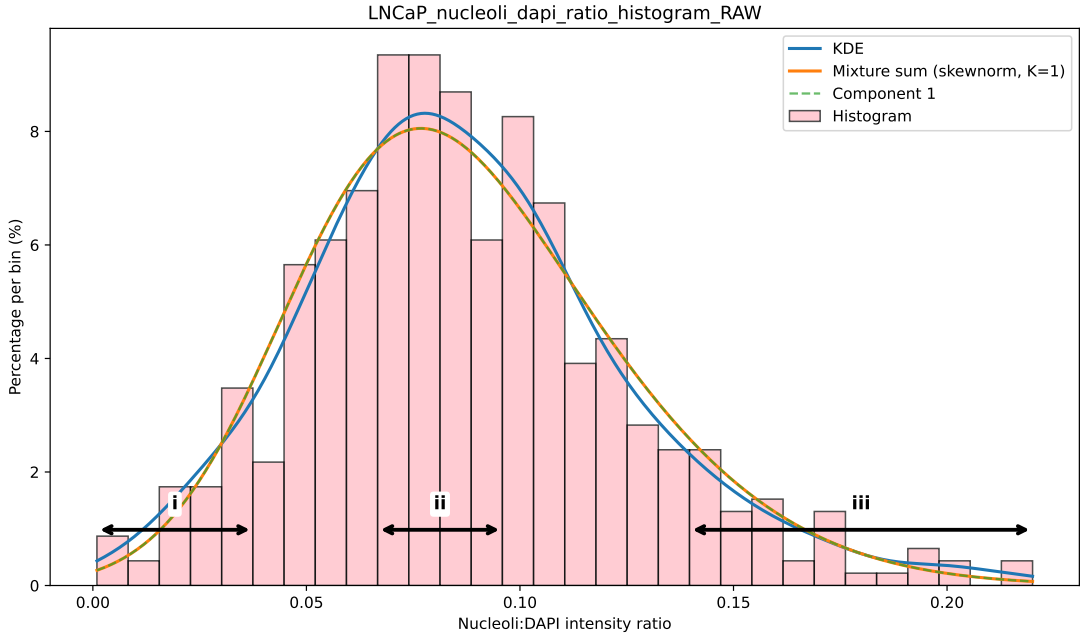


Figure 4.41: Histogram showing the ratio of nucleolar intensity to DAPI intensity in permeabilised LNCaP cells stained with DAPI and anti-fibrillarin (DAPI+FIB). A KDE is overlaid on top of the histogram, with fitting applied to the KDE to determine the overall distribution of the data [258]. The best model (chosen by ICL [259]) was a single skew-normal distribution with a $= 2.1 \pm 0.4$, loc $= 0.048 \pm 0.004$, and scale $= 0.054 \pm 0.003$ with a peak at 0.077 ± 0.003. The KDE peak location was 0.078 ± 0.004. Uncertainties are calculated from the covariance matrix.

The final metric considered via histogram-based analysis was the nucleolar:DAPI intensity ratio, $R_{N:D}$. Across both cell lines, $R_{N:D}$ was unimodal and right-skewed (Figures 4.41 and 4.42), consistent with a dominant baseline nucleolar fraction accompanied by a minority of nuclei in which nucleolar FIB intensity occupied a larger proportion of the overall nuclear fluorescence signal. Visually, the distributions appear similar to the absolute nucleolar intensity distributions (Figures 4.37 and 4.38), however the right tails were attenuated somewhat following normalisation to total DAPI intensity, a trend especially noticeable in LNCaP. Comparing the two cell lines, the LNCaP nucleolar ratio distribution was centred at higher values, and appeared to have a greater spread of central values than PNT2-C2, which rose more sharply from near-zero, and decayed more gradually with a longer right tail relative to the mode. This qualitative difference is informative, and suggests that PNT2-C2 contains a subset of nuclei with a disproportionately elevated nucleolar fraction relative to the typical state, even though the overall nucleolar intensity ratio was lower than in LNCaP.

In LNCaP, $R_{N:D}$ was best described by a single skew-normal distribution with a KDE mode of 0.078 ± 0.004. In PNT2-C2, on the other hand, $R_{N:D}$ was best described by a single Gamma distribution with a KDE mode of 0.049 ± 0.002. Thus, the nucleolar intensity ratio not only shifts upward in LNCaP relative to PNT2-C2, but also changes its distributional form in a way that is

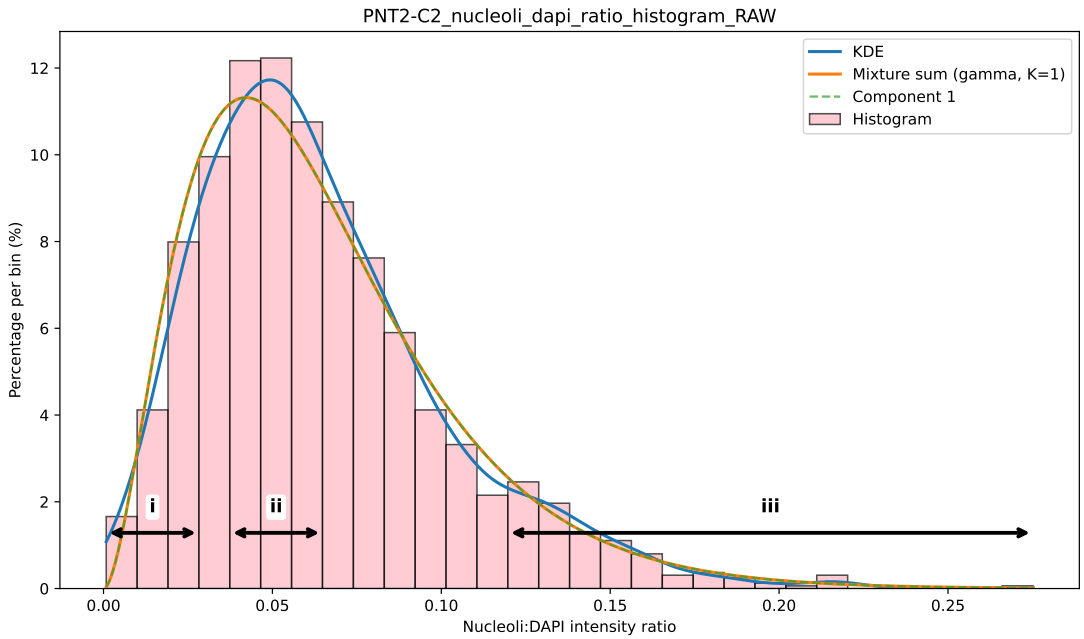


Figure 4.42: Histogram showing the ratio of nucleolar intensity to DAPI intensity in permeabilised PNT2-C2 cells stained with DAPI and anti-fibrillarin (DAPI+FIB). A KDE is overlaid on top of the histogram, with fitting applied to the KDE to determine the overall distribution of the data [258]. The best model (chosen by ICL [259]) was a single Gamma distribution with a $= 2.9 \pm 0.1$, loc $= 0$, and scale $= 0.0226 \pm 0.0008$, with a peak at 0.042 ± 0.001. The KDE peak location was 0.049 ± 0.002. Uncertainties are calculated from the covariance matrix.

consistent with reduced relative skewness (with skew-normal distributions typically being less tail-heavy than Gamma distributions) in the malignant model.

Making use of the bulk means discussed in Figure 4.31, LNCaP had a greater mean $R_{N:D}$ than PNT2-C2 (0.087 ± 0.002 vs. 0.0644 ± 0.0009, a $\sim 35\%$ relative increase). The KDE modes showed an even greater degree of separation, with modal values approximately 60% higher in LNCaP. Since $R_{N:D}$ explicitly couples nucleolar intensity to the integrated DAPI signal, it provides an informative intranuclear normalisation in the context of the DAPI+FIB DAPI results. As shown in Figure 4.27, PNT2-C2 had a higher mean total DAPI intensity than LNCaP under DAPI+FIB staining ($(8.8 \pm 0.1) \times 10^8$ a.u. vs. $(7.7 \pm 0.2) \times 10^8$ a.u.). This on its own would push $R_{N:D}$ downwards in PNT2-C2, however, the numerator I_{nucleoli} was also lower in PNT2-C2, and as such the reduction in $R_{N:D}$ reflects both the decreased nucleolar signal and increased DAPI signal in the non-malignant model. Consequently, the nucleolar intensity ratio strengthens (rather than merely restates) the conclusion that LNCaP nuclei exhibit a greater nucleolar burden relative to the nuclear DAPI signal, on a per-nucleus basis.

Tail descriptors again separate cell lines in a biologically interpretable direction. In LNCaP, 6.18% of nuclei had $R_{N:D} > 0.15$, and the 90th percentile was 0.137. In comparison, only 3.05% of PNT2-C2 nuclei exceeded 0.15, and the 90th percentile was 0.116. Stated explicitly, LNCaP not only experienced an upwards shift in the central tendency of the nucleolar intensity ratio compared to PNT2-C2, but also had approximately double the prevalence of very high nucleolar fractions under the defined fixed threshold. From a mechanistic standpoint, this type of tail enrichment is consistent with a subset of malignant nuclei in which the fibrillarin-enriched nucleolar substructure was particularly dominant relative to the nuclear DNA-associated fluorescence, plausibly reflecting enhanced ribosome biogenesis demand and/or nucleolar activation (driven by MYC or other oncogenes) [137, 214, 222, 456, 457].

Existing prostate cancer studies most commonly link nucleolar prominence to diagnosis and aggres-

siveness using morphological measures [63, 140, 430]. While these approaches are valuable, they do not decompose nucleolar biology into a protein-specific, compartment-enriched signal normalised to nuclear context at the single-cell level. In contrast, $R_{N:D}$ effectively operationalises a per-nucleus "nucleolar fraction", in a manner analogous to quantitative fluorescence protocols used to study nucleolar dynamics and stress responses in general cell biology, here applied to prostate cancer models and explicitly analysed via distributional shape and tails [139, 215]. The histograms presented here therefore contribute a distribution-level quantification of nucleolar enrichment; one that is largely absent from the wider prostate cancer literature.

The histogram analyses outline three key conclusions that are central to the contribution of the present study to the advancement of established knowledge in the field.

First, LNCaP displayed a globally elevated nucleolar signal state relative to the non-malignant comparator: both I_{nucleoli} and $R_{N:D}$ were higher at the level of mode (i.e. the most probable nucleus), as well as mean, indicating widespread upregulation rather than an outlier-driven effect. This is consistent with the established link between prostate malignancy and nucleolar activation, and with direct evidence that fibrillarin is MYC-regulated and functionally required for prostate cancer cell proliferation and rRNA processing [17, 63, 214, 222].

Second, the distributions reveal distinct heterogeneity patterns between models; PNT2-C2 showed larger mean-mode separations for both I_{nucleoli} and $R_{N:D}$, implying that the population averages were more strongly influenced by a minority of fringe events, whereas LNCaP shifted the typical state upwards and concurrently increased the absolute prevalence of very high-intensity nuclei. This resolves an apparent tension between "LNCaP has higher nucleolar intensity" and "PNT2-C2 has greater dispersion"; both are true, but refer to different aspects of the distributions (i.e. global shifts vs relative tail leverage).

Third, combining the histogram results with previous morphology context clarifies precisely what is changing. Nucleolar area remained essentially unchanged between cell lines (Figure 4.18), and total nuclear FIB intensity was also closely matched (Figure 4.31). Therefore, the elevated LNCaP I_{nucleoli} distribution is best explained as a re-partitioning of the nuclear FIB signal toward bright, nucleolus-assigned pixels (more simply an increased fibrillarin-enriched intensity, likely relating to the dense fibrillar centre), rather than a global increase in nuclear antibody signal or a simple expansion of nucleolar footprint. By taking an approximate ratio of mean values, the implied nucleolar fraction within the FIB channel $I_{\text{nucleoli}}/I_{\text{FIB}}$ was ~ 0.289 in LNCaP, and ~ 0.221 in PNT2-C2 (a $\sim 30\%$ relative increase), reinforcing that the key difference lies in specifically the nucleolar concentration of FIB signal; background nuclear fibrillarin staining effectively obscures real biological differences between cell lines. This compartment-specific redistribution is not directly recoverable from classical nucleolar prominence scoring or morphometry alone, and therefore constitutes substantive new knowledge added through the fluorescence based, distribution-level analysis in this study [18, 136, 139, 430].

4.3 Summary

This chapter tested two linked questions. First, it asked whether detergent permeabilisation materially alters DAPI-derived nuclear descriptors relative to the non-permeabilised baseline from Chapter 3. Second, it asked whether direct nucleolar measures derived from the nucleolar marker fibrillarin (FIB)

identify additional malignant-associated differences between LNCaP and the non-malignant comparator PNT2-C2 (Sections 4.1 and 4.2). Convergence was therefore re-evaluated for the permeabilised datasets rather than assumed from Chapter 3, and the cross-preparation synthesis figures were used to summarise the comparison of DAPI-derived metrics across non-permeabilised DAPI, permeabilised DAPI-only, and permeabilised DAPI+FIB preparations (Figures 4.16, 4.17 and 4.28 to 4.30). For the legacy DAPI-derived descriptors, all reported convergence points remained comfortably below the available nucleus counts in both permeabilised branches, confirming statistical sufficiency under the same %SEM $< 2.5\%$ criterion used previously Tables 4.1 and 4.2. However, the newly introduced nucleolar descriptors required substantially larger N, particularly I_{nucleoli} in LNCaP ($N_{\text{conv}} = 455$, with $N_{\max} = 460$) and $R_{N:NA}$ in PNT2-C2 ($N_{\text{conv}} = 1045$, with $N_{\max} = 1627$), and therefore small nucleolar effects should be interpreted more cautiously than the corresponding DAPI-derived morphology readouts (Table 4.2). Relative heterogeneity, assessed by γ, also changed in a descriptor-specific manner after permeabilisation: the strong LNCaP $<$ PNT2-C2 ordering in γ was largely erased for nuclear area, remained directionally present but weaker for aspect ratio, and reversed for DAPI density, indicating that detergent treatment had a pronounced effect on several of the pre-existing between-line heterogeneity contrasts (Tables 3.1 and 4.1).

The standout result is that preparation state materially affects several DAPI-derived nuclear measurements. Most clearly, nuclear area was not preparation-invariant Figure 4.16. In the non-permeabilised baseline, LNCaP nuclei were larger than their PNT2-C2 counterparts by approximately 45% in the mean and $\sim 30\%$ in the KDE mode; this ordering reversed after permeabilisation, with PNT2-C2 nuclear area exceeding LNCaP by approximately 20% at both mean and modal level under the DAPI-only control (Figures 3.8, 4.2 and 4.16). This reversal persisted under the full DAPI+FIB workflow, although the separation reduced to approximately 7% in the mean and 5% in the mode (Figures 4.15 and 4.16). Relative area dispersion showed the same preparation sensitivity: the greater area CV in LNCaP ($\sim 25\%$ relative increase compared to PNT2-C2) prior to permeabilisation was effectively erased in the permeabilised DAPI-only and DAPI+FIB conditions, with the two cell lines becoming indistinguishable outside of uncertainty in both permeabilised branches (Figure 4.16).

In contrast, nuclear aspect ratio was more robust Figure 4.17. LNCaP remained the more elongated model across all three preparations, exceeding PNT2-C2 by approximately 20%, 17%, and 40% in mean aspect ratio in the non-permeabilised, permeabilised DAPI-only, and DAPI+FIB conditions respectively, while the modal differences were approximately 10%, 5%, and 35% (Figure 4.17(a)). Likewise, aspect ratio CV also remained higher in LNCaP throughout, by approximately 60%, 45%, and 70% across the same three conditions (Figure 4.17(b)). As such, the DAPI-derived morphology results separate into preparation-sensitive and preparation-robust classes: apparent nuclear size was strongly preparation-sensitive, whereas increased elongation and shape heterogeneity in the malignant model persisted across workflows [16, 295, 296].

Absolute total DAPI intensity and DAPI density per nucleus were also strongly preparation-sensitive; in both cases the apparent between-line ordering was strongly attenuated or even reversed across conditions (Figures 4.28 and 4.30). For total I_{DAPI}, the non-permeabilised baseline showed an approximately 40% higher mean and $\sim 60\%$ higher mode in LNCaP, but after permeabilisation the mean and mode difference shrank to $\sim 10\%$ and $\sim 30\%$ respectively (Figures 3.14 and 4.7). Under DAPI+FIB, the mean ordering reversed such that PNT2-C2 exceeded LNCaP by approximately 15%, while the modes became indistinguishable within uncertainty (Figures 4.27 and 4.28). DAPI density behaved similarly: the

mean in the non-permeabilised baseline did not differ outside of uncertainty between cell lines, though the KDE mode was approximately 15% higher in LNCaP; after permeabilisation, both summaries shifted such that LNCaP had a $\sim$ 30% greater mean DAPI density, and $\sim$ 65% higher mode, and under DAPI+FIB the ordering then reversed in both the mean and mode, with PNT2-C2 becoming the denser model ($\sim$ 5% higher mean and $\sim$ 10% higher mode) (Figure 4.30(a)).

By contrast, the intranuclear DAPI texture metric $\mathrm{CV}_{\mathrm{DAPI}}$ preserved the same cell line ordering across all three conditions Figure 4.29(b). LNCaP exceeded PNT2-C2 by approximately 40%, 25%, and 30% in the mean $\mathrm{CV}_{\mathrm{DAPI}}$, and by approximately 60%, 15%, and 45% in the KDE mode, in the non-permeabilised, permeabilised DAPI-only, and DAPI+FIB preparations respectively (Figure 4.29). The main conclusion from within the DAPI channel is therefore not that all nuclear descriptors fail under detergent permeabilisation, but that absolute intensity and density scales are strongly preparation-dependent, whereas the intranuclear DAPI intensity heterogeneity in the malignant model remains directionally robust [8, 16].

Within that preparation-aware framework, the direct nucleolar measurements provide the central biological contribution of this chapter. Nucleolar morphology alone showed only limited separation between cell lines. Absolute nucleolar area was effectively unchanged within uncertainty, while the nucleoli:nuclei area ratio $R_{N:NA}$ showed only a modest elevation in LNCaP, of approximately 7% in the mean and 10% in the mode (Figure 4.18). In contrast, intensity-based nucleolar descriptors showed a clearer and more consistent malignancy-associated shift. Mean total nucleolar intensity per nucleus (I_{nucleoli}; Equation (2.21)) was approximately 25% higher in LNCaP than PNT2-C2, and the modal value was approximately 50% higher (Figures 4.31, 4.37 and 4.38). The nucleoli:DAPI ratio $R_{N:D}$ (Equation (2.24)) showed an even clearer separation, with the mean approximately 35% higher and the mode approximately 60% higher in LNCaP (Figures 4.31, 4.41 and 4.42).

Importantly, the total within-nucleus fibrillarin intensity I_{FIB} did not differ outside of uncertainty between cell lines, indicating that the principal contrast was not a global increase in nuclear fibrillarin signal, but rather a greater fraction of that signal being concentrated into nucleolar-assigned pixels in LNCaP. Expressed approximately as a fraction of mean signal, this corresponded to a $\sim$ 30% relative increase in nucleolar-assigned fibrillarin signal relative to the total FIB intensity in the malignant model. At the same time, several nucleolar measures remained more dispersed across nuclei in PNT2-C2, with heavier mean-mode separation and higher CV across nuclei, indicating that the non-malignant comparator contained a broader mixture of nucleolar states, even though its typical nucleolar signal state was lower (Figure 4.31). Interpreted cautiously, these results are consistent with a malignant nucleolar enrichment or compartment-specific concentration of fibrillarin, rather than with a simple gross enlargement of nucleolar footprint [17, 63, 136, 214, 219, 222, 429].

The issue of perturbation resulting from detergent permeabilisation does not disappear once fibrillarin is introduced, however. The nucleolar readouts presented here remain conditional on thresholding, mask quality, background nuclear FIB signal, and the possibility of permeabilisation-induced redistribution, uptake, extraction, or damage. Additionally, several histogram tails and low-intensity substructures were compatible with observed segmentation edge cases rather than unambiguously representing biological states, and the assignment of nucleolar pixels was susceptible to under-detection where FIB signal varied strongly within a given image (Figure 2.7 and Sections 4.2.2.1 and 4.2.2.2). Accordingly, the present data support a preparation-aware statement of stronger nucleolar enrichment in LNCaP

within this workflow, but do not by themselves prove larger total nucleolar volume, nor can they exclude the contribution of preparation-related artefacts.

Collectively, this chapter provides two clear, thesis-wide insights. First, preparation state materially affects DAPI-derived nuclear metrics, such that the apparent biological ordering observed prior to permeabilisation may be weakened, altered, or even reversed by workflow changes alone. Second, the malignant model showed a reproducible increase in nucleolar fibrillarin-associated intensity, both in absolute terms, and relative to nuclear DAPI signal, despite limited separation in nucleolar area. A direct practical implication of these findings is that future studies should explicitly treat permeabilisation time and conditions as quantitative variables to be calibrated jointly against preservation of DAPI-derived nuclear descriptors and adequate intranuclear marker access.

Chapter 5

Conclusion

5.1 Overview and final answer to the overarching thesis question

This thesis established and evaluated an automated fluorescence microscopy workflow for quantitative single-nucleus analysis in two prostate-derived cell lines: LNCaP as a malignant model, and PNT2-C2 as an immortalised non-malignant comparator. The workflow was applied under three matched preparation states: non-permeabilised DAPI staining, permeabilised DAPI-only staining, and permeabilised DAPI plus anti-fibrillarin (DAPI+FIB) immunofluorescence. Nuclear morphology, DAPI-derived intensity descriptors, and fibrillarin-derived nucleolar descriptors were then analysed at the level of the individual nucleus, with explicit attention given to full distributional structure, rather than to mean values alone.

This thesis addressed one overarching question through four linked research questions, with the overarching question being whether quantitative single-nucleus fluorescence microscopy could identify reproducible and biologically interpretable differences in nuclear and nucleolar state between the malignant and non-malignant models once sample-size sufficiency and preparation-dependent effects were considered explicitly. The four component questions each concerned sampling sufficiency, preparation dependence, comparative distributional behaviour, and nucleolar state via fibrillarin.

The overall answer is yes, but specifically under a preparation-aware interpretation. Under the non-permeabilised baseline, LNCaP nuclei were larger, more elongated, and more heterogeneous than PNT2-C2 nuclei, with the strongest separation observed in intranuclear DAPI texture. Detergent permeabilisation, however, materially altered several DAPI-derived measurements, and in some cases weakened or even reversed the apparent ordering between cell lines, though nuclear aspect ratio remained directionally robust across conditions. As such, this thesis does not support the broad assumption that all DAPI-derived nuclear readouts are preparation-invariant. Once that limitation is acknowledged directly, a clearer and more robust biological result emerges: the malignant model showed stronger fibrillarin-enriched nucleolar intensity, both in absolute terms and relative to nuclear DAPI signal, despite only limited separation in nucleolar area. This is the principal biological conclusion of the study.

A second major conclusion is methodological. This thesis showed that explicit statistical sufficiency testing is not an optional extra, but a necessary part of defensible single-cell or single-nucleus microscopy

analysis, especially when heterogeneity itself is part of the scientific question. In the present dataset, stable interpretation depended not only on what was measured, but also on whether enough nuclei (N) had been analysed for each descriptor.

5.2 Principal findings

5.2.1 Sample-size sufficiency and nucleus-level heterogeneity

A central strength of this study is that sample-size sufficiency was tested explicitly, rather than assumed. For the principal non-permeabilised nuclear descriptors, the standard error decreased with the expected approximate $N^{-1/2}$ behaviour, supporting the use of the convergence framework adopted here. Importantly, LNCaP generally required larger effective nucleus counts than PNT2-C2 to reach the same relative precision threshold. For example, the convergence requirement was approximately 60% higher for nuclear area, approximately 140% higher for aspect ratio, and approximately 30% higher for DAPI density in LNCaP than in PNT2-C2. Interpreted directly, the malignant model was not only distinct in terms of the average nuclear state, but also more heterogeneous at the level of individual nuclei.

This matters scientifically because heterogeneity was one of the key signals of interest, not merely background noise, and also has practical implications because it highlights that microscopy studies are at particular risk of unstable summary estimates in more heterogeneous cancer models. In the present work, convergence testing therefore served two roles simultaneously: it both justified that the analysed nucleus counts were sufficient to make the main conclusions with a reasonable degree of confidence, and provided a descriptor-specific readout of relative heterogeneity.

5.2.2 Baseline nuclear differences in the non-permeabilised condition

The non-permeabilised dataset provided the least perturbed fixed-cell nuclear baseline in the thesis, and therefore the clearest reference point for interpretation of the later permeabilised results. While still a fixed-cell preparation rather than a true measure of the native state, it is the condition most likely to preserve intrinsic nuclear morphology and baseline DAPI-derived nuclear characterisation without the additional perturbation introduced by detergent permeabilisation. Under this baseline condition, LNCaP showed a substantial increase in mean nuclear area relative to PNT2-C2 of approximately 45%, together with an increase in mean nuclear aspect ratio of approximately 20%. Thus, before any detergent permeabilisation was introduced, the malignant model already exhibited a quantitatively larger and less circular nuclear architecture. More broadly, this baseline was important not only as a comparator for the later permeabilised workflows, but as a standalone fixed-cell reference for assessing nuclear morphology and DAPI-based nuclear state in a manner least confounded by preparation-induced structural change.

The DAPI-derived texture measures strengthened that distinction further. The mean intranuclear DAPI intensity coefficient of variation was approximately 40% higher in LNCaP, while the modal value assessed via kernel density estimation (KDE) was approximately 60% higher. This was the strongest and most directionally consistent DAPI intensity-based separation observed in the study. In practical terms, both the average and the typical nucleus state in LNCaP were shifted towards greater internal DAPI heterogeneity relative to PNT2-C2. Interpreted cautiously, this is consistent

with greater irregularity in DNA-associated nuclear organisation in the malignant model, although DAPI texture should still be treated as an indirect optical readout rather than as a direct measurement of chromatin state.

The distribution-level analyses were also informative here, illustrating that the LNCaP populations were not adequately summarised by mean values alone. In particular, the DAPI texture distributions indicated that the malignant model combined an upward shift in the dominant nuclear state with additional internal structure that would have been obscured if only bulk-statistics were reported. More generally, the non-permeabilised baseline showed that the malignant and non-malignant models differed not only in average descriptor values, but also in dispersion, skew, and the relative weighting of distinct regions of the distribution. One of the broader lessons taken from this chapter is therefore that biologically relevant differences between these models are expressed at the level of full single-nucleus distributions, not solely through shifts in mean values.

5.2.3 Effect of permeabilisation on DAPI-derived descriptors

The most important thesis-wide methodological result is that preparation state materially affected several DAPI-derived nuclear measurements. Nuclear area provided the clearest and most intuitive example. In the non-permeabilised baseline, LNCaP nuclei were larger than PNT2-C2 nuclei. After permeabilisation, that ordering reversed, such that PNT2-C2 appeared larger under both the permeabilised DAPI-only workflow and the DAPI+FIB workflow. Absolute DAPI intensity and DAPI density were likewise strongly preparation-sensitive, with between-line differences being attenuated, altered, or reversed depending on the workflow. Importantly, these effects were not confined to means alone; the cross-condition comparisons and histogram/KDE analyses together showed that detergent permeabilisation could also shift the dominant nuclear state and alter tail structure, thereby changing how the underlying populations should be interpreted.

This point is important because it changes how the DAPI results should be interpreted. The data do not show that DAPI-derived nuclear measurements are intrinsically unreliable; they show that some DAPI-derived descriptors are highly conditional on the physical preparation pathway used to generate the image. In other words, the preparation is a critical part of the measurement model. A nuclear descriptor that appears biologically informative in one workflow cannot be assumed to carry the same interpretation in another workflow unless that assumption has been tested.

At the same time, not all descriptors were equally sensitive. Nuclear aspect ratio remained consistently higher in LNCaP across all three preparation states, and intranuclear DAPI texture also preserved the same overall direction of separation across all conditions, even where the KDE mode and broader distributional structure shifted. This distinction is scientifically useful. It suggests that detergent permeabilisation did not erase all malignant-associated nuclear structure, but instead affected descriptor classes differently. In this dataset, apparent nuclear size and some intensity-scale measures were strongly preparation-sensitive, whereas elongation and intranuclear texture were relatively more robust.

5.2.4 Direct nucleolar findings from fibrillarin staining

The nucleolar readouts provided the main added biological value beyond the DAPI-only analyses. Here, the most important result was that nucleolar morphology as assessed through the present workflow provided only weak separation between the two cell lines. Absolute nucleolar area was effectively

unchanged within uncertainty, and the nucleoli:nuclei area ratio showed only a modest increase in LNCaP.

By contrast, intensity-based nucleolar descriptors showed a much more pronounced malignancy-associated shift. Mean total nucleolar intensity per nucleus was approximately 25% higher in LNCaP, and the modal value was approximately 50% higher. The nucleoli:DAPI intensity ratio showed a stronger separation still, with an approximately 35% higher mean and approximately 60% higher KDE mode in the malignant model. These are not small effects, and indicate that both the average and the typical nucleolar signal state were shifted upward in LNCaP relative to PNT2-C2, despite nucleolar area only showing limited separation.

Crucially, the total within-nucleus fibrillarin signal outside of the nucleolus-assigned partition did not show the same clear between-line increase. The most defensible interpretation is therefore not that LNCaP simply contained more total nuclear fibrillarin everywhere, and not that it necessarily contained substantially larger nucleoli. Instead, the data from this study are more consistent with a greater concentration or enrichment of fibrillarin-associated signal within the segmented nucleolar compartment in the malignant model. Put more simply, the main nucleolar difference in this dataset was a stronger nucleolar signal, not clear gross nucleolar enlargement.

This distinction matters because it sharpens the biological interpretation. If nucleolar area had been observed to strongly increase on its own, the conclusion would have been limited to a larger nucleolar footprint. Instead, the present data suggest a more specific change in signal partitioning or compartmental concentration, more closely aligned with the idea of altered nucleolar activity than with a purely geometric change in nucleolar size. This conclusion must still remain cautious, however the direction of evidence is clear: within the present preparation-aware workflow, malignancy was expressed more strongly through nucleolar intensity enrichment than through nucleolar area expansion. Just as importantly, the clearest separation was seen at the level of the full distributions and their KDE peaks, not just through morphological averages in isolation.

5.3 Scientific meaning and contribution

Taken together, the results support three thesis-wide claims.

First, the malignant and non-malignant models were distinguishable at the level of single-nucleus descriptor distributions. This distinction was visible not only in central tendency, but also in heterogeneity, skew, tail behaviour, and the broader distributional structure as a whole. This is an important point because while cancer can be diagnosed at the single-cell level, cancer-associated differences are often discussed qualitatively or through a small number of mean values. In contrast, the present work demonstrates that the full distribution can contain additional information about the dominant nuclear state, the spread of the population, and the presence of internal substructure that would otherwise remain hidden.

Second, preparation state is itself a major experimental variable in fluorescence-based nuclear analysis; a finding which constitutes one of the most important scientific results of the thesis. The study therefore provides a practical warning for future work that extends beyond simple cross-workflow comparison. Where detergent permeabilisation is used, and DAPI-derived nuclear or related nucleolar descriptors are interpreted as informative of underlying cell state, the inclusion of a non-permeabilised or otherwise

minimally perturbed reference is strongly advised as part of the experimental design. Without such a reference, a permeabilisation-induced shift in the mean, KDE peak, or broader distributional form could easily be misread as a biological property of the cell rather than of the preparation itself. In that setting, interpretation should be limited carefully to only reflect the measured post-permeabilisation state unless the effect of permeabilisation has been quantified independently. This also highlights the value of including a minimally perturbed fixed-cell baseline wherever possible, since without it, even fixed-cell nuclear characterisation may become difficult to separate from preparation artefact.

Third, once preparation-induced shifts have been handled explicitly, the strongest malignancy-associated nucleolar signal in this dataset is not increased nucleolar size, but increased fibrillarin-enriched nucleolar intensity relative to nuclear context. This is the principal biological contribution of the thesis. It suggests that, within these controlled model systems, nucleolar reorganisation is expressed more clearly through signal enrichment than through simple enlargement of segmented nucleolar area.

More broadly, the thesis contributes methodologically by combining four elements that are not commonly integrated in a single workflow: automated batch image analysis, descriptor-level convergence testing, explicit comparison of matched preparation states, and distribution-focused single-nucleus interpretation of both nuclear and nucleolar measurements. In particular, the study does not reduce each descriptor to a single summary statistic, but treats the full single-nucleus distribution as an informative analytical object in its own right. The work therefore does not claim a specific clinical diagnostic advance; its contribution is instead to provide a reproducible and analytically explicit framework for quantitative nuclear and nucleolar characterisation in a controlled prostate cell line setting.

5.4 What the data do and do not show

The data support a clear within-study conclusion that LNCaP and PNT2-C2 differ in nuclear and nucleolar descriptor distributions under matched acquisition and preparation conditions. They also support the conclusion that while some of those differences are highly dependent on preparation state, others are more robust. Finally, they support a preparation-aware conclusion of stronger nucleolar fibrillarin enrichment in the malignant model.

However, the data are unable to show that these descriptors are clinically validated biomarkers for patient diagnosis, prognosis, or treatment selection. This is a controlled cell line study, not a tissue study, clinical cohort study, or even primary cell culture study. The data also do not prove the precise underlying mechanism responsible for the observed DAPI and fibrillarin shifts. For example, while altered DAPI texture is compatible with observed changes in nuclear organisation, it cannot by itself specify the exact chromatin, cell-cycle, or genomic process involved. Likewise, stronger fibrillarin-enriched nucleolar intensity is consistent with increased nucleolar activity, but does not directly measure ribosome biogenesis flux, transcription rate, or protein synthesis. More generally, while the distribution-level analyses identify shifts in dominant state, dispersion, and substructure, they cannot prove that each such feature maps uniquely onto any single biological mechanism.

The study therefore narrows the biologically plausible interpretation, rather than proving a single mechanistic explanation. This level of caution is appropriate to the data type, and to the scope of the thesis.

5.5 Limitations of the study

Several limitations should be stated explicitly.

The first is biological scope. Only one malignant line and one immortalised non-malignant comparator were studied, and while this pairing was suitable for a controlled proof-of-principle analysis, it necessarily limits generalisation. LNCaP cannot represent all prostate cancers, and PNT2-C2 should not be treated as a perfect surrogate for healthy prostate epithelium. The present conclusions therefore remain model-specific.

The second is that the study used fixed-cell, two-dimensional, widefield fluorescence imaging. This enabled controlled batch analysis, but also imposed interpretational limits. Widefield imaging integrates out-of-plane light, the measurements are static/endpoint rather than dynamic, and the segmented masks represent two-dimensional projections rather than full three-dimensional nuclear or nucleolar structure.

The third is that the intensity-derived variables are comparative rather than absolute. No external calibration was performed for DNA content or fibrillarin abundance. As such, DAPI intensity should not be interpreted as an exact measure of DNA mass, and fibrillarin intensity should not be interpreted as a direct measure of molecular copy number. These signals are meaningful within matched protocols, but not as absolute physical quantities, and therefore should not be directly compared across studies without careful qualification.

The fourth is that the descriptor analyses explicitly depended on the accuracy of the image segmentation pipeline. In the DAPI channel, imperfect nuclear segmentation could affect measured nuclear area, aspect ratio, and intensity-derived descriptors through boundary misassignment, local under-segmentation, over-segmentation, or failure in crowded or morphologically irregular fields. Because the DAPI-defined nuclear mask also provided the spatial framework within which nucleolar measurements were assigned, any inaccuracy at the nuclear segmentation stage could propagate directly into the downstream FIB-derived descriptors. The nucleolar analyses were additionally sensitive to successful thresholding and mask transfer within this framework. Background nuclear FIB signal, under-detection of dim nucleolar regions, imperfect partitioning of nucleolar vs. non-nucleolar signal, and local segmentation failures could all meaningfully affect the measured nucleolar descriptors. This is particularly relevant for the interpretation of distribution tails and very small subpopulations, where algorithmic edge cases, segmentation error, and true biological extremes may coexist. More generally, these limitations mean that apparent fine-scale distributional structure should be interpreted cautiously unless supported by consistent behaviour across related descriptors, rather than by isolated extremes alone. In a similar vein, some nucleolar descriptors required relatively large nucleus counts for stable estimation, meaning that smaller effects in those readouts should be interpreted more cautiously than the main DAPI-derived morphology findings.

The fifth is that the convergence analysis carried out in this study quantified within-dataset estimate stability, not the full biological reproducibility across independently prepared cohorts. The convergence results are therefore highly valuable, but do not replace repeated biological experiments across separate passages, preparations, operators, and imaging sessions.

5.6 Recommendations for future work

The immediate methodological priority is to treat permeabilisation as a tunable quantitative variable, rather than as a fixed background step. The present results strongly justify a more systematic optimisation study in which detergent type, concentration, and duration are varied deliberately, with nuclear preservation and intranuclear marker access assessed jointly. In practical terms, any future studies should give explicit design consideration to identifying the mildest preparation that still provides adequate marker access, while preserving the most interpretable DAPI-derived nuclear readouts.

A second priority is to strengthen the segmentation and validation pipeline at both the nuclear and nucleolar levels, since the fidelity of the DAPI-defined nuclear masks constrains not only the primary nuclear descriptors, but also the downstream assignment of FIB-derived nucleolar measurements. Future work should therefore go beyond incremental improvement of the nucleolar thresholding alone, and instead treat segmentation accuracy as a methodological priority. This should include improved nuclear mask generation in crowded or morphologically challenging fields, more robust handling of dim, heterogeneous, or partially resolved nucleolar signal, clearer propagation of uncertainty from the DAPI segmentation stage into downstream nucleolar descriptors, and more explicit quality-control measures with extensive targeted manual review of likely failure cases. It would also be valuable to incorporate orthogonal validation, both computationally against alternative segmentation approaches, and biologically against additional nucleolar markers or independent assays. In particular, any subsequent work should seek to test whether the observed increase in fibrillarin-enriched nucleolar signal in LNCaP is reproducible under a more robust segmentation framework, and whether it is accompanied by corresponding changes in ribosomal RNA synthesis, nucleolar subcompartment organisation, or other markers of nucleolar activity.

A third priority is external biological extension. The same analysis framework should be applied to additional prostate cancer models, additional non-malignant comparators, and repeated biological preparations, or more plainly, a sufficient number of independent biological replicates rather than technical replicates alone. Only then can the extent to which the present findings are specific to the LNCaP-PNT2-C2 pairing, or instead reflect a broader malignant-versus-non-malignant pattern, be determined robustly. Beyond this, extending the analysis to primary cell cultures would further strengthen biological relevance and improve the extent to which the observed nuclear and nucleolar patterns can be interpreted as reflecting prostate cancer-associated state beyond the present immortalised cell-line system.

A fourth priority is structural extension beyond simple 2D cell culture. Once the workflow is stable and well-established, it would be valuable to test whether the same preparation-aware nuclear and nucleolar patterns are retained in more complex systems such as spheroids, organoids, or tissue sections. This would provide a route from controlled model characterisation toward greater biological and clinical relevance, while still preserving the analytical discipline developed in this thesis.

Finally, future studies should continue to retain the full descriptor distributions wherever possible. The present work has shown that important information can be lost when measurements are immediately reduced to a single average value. Distribution-aware analysis should therefore be regarded not as an optional embellishment, but as a useful default whenever heterogeneity is scientifically relevant. More explicitly, future analyses should continue to report both conventional summary statistics and distribution-level descriptors such as the KDE mode, while maintaining appropriate caution in the

interpretation of sparse tails and apparent subpopulations.

5.7 Final statements

In conclusion, this thesis has shown that quantitative fluorescence microscopy can distinguish between malignant and non-malignant prostate cell models at the level of single-nucleus nuclear and nucleolar descriptor distributions, but that those conclusions must be interpreted through the lens of preparation state. The most important nuclear finding is that several DAPI-derived descriptors, especially nuclear area and intensity-scale measures, are strongly affected by detergent permeabilisation, and cannot be assumed to transfer unchanged between workflows, even where some features such as nuclear aspect ratio and DAPI texture retain directionally robust between-line separation. The most important nucleolar finding is that the malignant LNCaP model shows stronger fibrillarin-enriched nucleolar intensity, both absolutely and relative to nuclear DAPI signal, despite only limited separation in nucleolar area.

The scientific contribution of the thesis therefore lies in two linked advances. First, it provides a reproducible and convergence-tested framework for preparation-aware, single-nucleus fluorescence analysis. Second, it shows that, in this controlled prostate cell line setting, the malignant state is expressed more clearly through robust shifts in nuclear heterogeneity, dominant distributional state, and nucleolar signal enrichment than through simple nucleolar enlargement alone. As such, the thesis delivers both a substantive biological result and a methodological basis for more defensible future studies of nuclear and nucleolar organisation in prostate cancer models.

Bibliography

[1] J. S. Horoszewicz et al. *The LNCaP Cell Line–a New Model for Studies on Human Prostatic Carcinoma*. In: *Progress in Clinical and Biological Research* 37 (1980), pp. 115–132. ISSN: 0361-7742. PMID: 7384082.

[2] Julius S. Horoszewicz et al. *LNCaP Model of Human Prostatic Carcinoma*. In: *Cancer Research* 43.4 (1st Apr. 1983), pp. 1809–1818. ISSN: 0008-5472.

[3] P Berthon et al. *Functional Expression of SV40 in Normal Human Prostatic Epithelial and Fibroblastic Cells: Differentiation Pattern of Non-Tumorigenic Cell Lines*. In: *International Journal of Oncology* (1st Feb. 1995). ISSN: 1019-6439, 1791-2423. DOI: 10.3892/ijo.6.2.333.

[4] *Cellosaurus Cell Line PNT2 (CVCL_2164)*. Cellosaurus. 14th Aug. 25. URL: https://www.cellosaurus.org/CVCL_2164 (visited on 30/01/2026).

[5] Gurvinder Kaur and Jannette M. Dufour. *Cell Lines: Valuable Tools or Useless Artifacts*. In: *Spermatogenesis* 2.1 (1st Jan. 2012), pp. 1–5. ISSN: null. DOI: 10.4161/spmg.19885. PMID: 22553484.

[6] Takeshi Namekawa et al. *Application of Prostate Cancer Models for Preclinical Study: Advantages and Limitations of Cell Lines, Patient-Derived Xenografts, and Three-Dimensional Culture of Patient-Derived Cells*. In: *Cells* 8.1 (20th Jan. 2019), p. 74. ISSN: 2073-4409. DOI: 10.3390/cells8010074. PMID: 30669516.

[7] Zahra Ahmed et al. *How Clinically Relevant Are Prostate Cancer Cell Lines? A Comprehensive Characterisation and Multiomics Comparison*. 22nd Mar. 2024. DOI: 10.1101/2024.03.20.585982. URL: https://www.biorxiv.org/content/10.1101/2024.03.20.585982v2 (visited on 08/10/2025). Pre-published.

[8] Jan Kapuscinski. *DAPI: A DNA-Specific Fluorescent Probe*. In: *Biotechnic & Histochemistry* (1995). DOI: 10.3109/10520299509108199.

[9] Anabela Ferro et al. *Blue Intensity Matters for Cell Cycle Profiling in Fluorescence DAPI-stained Images*. In: *Laboratory Investigation* 97.5 (May 2017), pp. 615–625. ISSN: 1530-0307. DOI: 10.1038/labinvest.2017.13.

[10] Emily Honzel et al. *A Comparison of Confocal and Epifluorescence Microscopy for Quantification of RNAScope and Immunohistochemistry Fluorescent Images*. In: *Journal of Neuroscience Methods* 403 (1st Mar. 2024), p. 110050. ISSN: 0165-0270. DOI: 10.1016/j.jneumeth.2023.110050.

[11] Kyuseok Im et al. *An Introduction to Performing Immunofluorescence Staining*. In: *Methods in Molecular Biology* (2018). DOI: 10.1007/978-1-4939-8935-5_26.

[12] ShawnR. Yoshida, Barun K. Maity and Shasha Chong. *Visualizing Protein Localizations in Fixed Cells: Caveats and the Underlying Mechanisms*. In: *The Journal of Physical Chemistry. B* 127.19 (10th May 2023), pp. 4165–4173. ISSN: 1520-6106. DOI: 10.1021/acs.jpcb.3c01658. PMID: 37161904.

[13] Douglas Hanahan. *Hallmarks of Cancer: New Dimensions*. In: *Cancer Discovery* 12.1 (12th Jan. 2022), pp. 31–46. ISSN: 2159-8274. DOI: 10.1158/2159-8290.CD-21-1059.

[14] Karolina Sienkiewicz et al. *Genomic Analyses of the Metastasis-Derived Prostate Cancer Cell Lines LNCaP, VCaP, and PC3-AR*. In: *The Prostate* 82.4 (Mar. 2022), pp. 442–451. ISSN: 0270-4137. DOI: 10.1002/pros.24290. PMID: 34951700.

[15] Goutam Chakraborty et al. *Significance of BRCA2 and RB1 Co-Loss in Aggressive Prostate Cancer Progression*. In: *Clinical cancer research : an official journal of the American Association for Cancer Research* 26.8 (15th Apr. 2020), pp. 2047–2064. ISSN: 1078-0432. DOI: 10.1158/1078-0432.CCR-19-1570. PMID: 31796516.

[16] Ruoyu Cheng et al. *Influence of Fixation and Permeabilization on the Mass Density of Single Cells: A Surface Plasmon Resonance Imaging Study*. In: *Frontiers in Chemistry* 7 (23rd Aug. 2019), p. 588. ISSN: 2296-2646. DOI: 10.3389/fchem.2019.00588. PMID: 31508410.

[17] Cheryl M. Koh et al. *Alterations in Nucleolar Structure and Gene Expression Programs in Prostatic Neoplasia Are Driven by the MYC Oncogene*. In: *The American Journal of Pathology* 178.4 (Apr. 2011), pp. 1824–1834. ISSN: 0002-9440. DOI: 10.1016/j.ajpath.2010.12.040. PMID: 21435462.

[18] Sarina Saffarian et al. *Elevated NPM1 and FBL Expression Correlates with Prostate Cancer Aggressiveness and Progression*. In: *The Journal of Pathology* 267.1 (2025), pp. 56–68. ISSN: 1096-9896. DOI: 10.1002/path.6447.

[19] Donald S. Coffey. *Similarities of Prostate and Breast Cancer: Evolution, Diet, and Estrogens*. In: *Urology* 57.4 (1st Apr. 2001), pp. 31–38. ISSN: 0090-4295, 1527-9995. DOI: 10.1016/S0090-4295(00)00938-9.

[20] Susan Standring and Henry Gray, eds. *Gray's Anatomy: The Anatomical Basis of Clinical Practice*. Forty-second edition. Amsterdam: Elsevier, 2021. 1588 pp. ISBN: 978-0-7020-7705-0.

[21] Omesh Singh and Srinivasa Rao Bolla. *Anatomy, Abdomen and Pelvis, Prostate*. In: *StatPearls*. Treasure Island (FL): StatPearls Publishing, 17th July 2023. PMID: 31082031.

[22] Cleveland Clinic. *Prostate: Anatomy, Location, Function & Conditions*. Cleveland Clinic. 8th Sept. 2022. URL: https://my.clevelandclinic.org/health/body/23965-prostate (visited on 21/11/2025).

[23] Gowtham S. Mahadevan et al. *Anatomy of the Prostate Gland: Modalities and Techniques for Its Assessment*. In: *Journal of Gastrointestinal and Abdominal Radiology* 07.02 (May 2024), pp. 099–105. ISSN: 2581-9178. DOI: 10.1055/s-0044-1785197.

[24] Thomas A. Harding et al. *Optimising the Use of the Prostate- Specific Antigen Blood Test in Asymptomatic Men for Early Prostate Cancer Detection in Primary Care: Report from a UK Clinical Consensus*. In: *British Journal of General Practice* 74.745 (1st Aug. 2024), e534–e543. ISSN: 0960-1643, 1478-5242. DOI: 10.3399/BJGP.2023.0586. PMID: 39038964.

[25] Mikael Häggström. *English: Zones of the Prostate.* 29th Oct. 2019.

[26] Michael Ng, Stephen W. Leslie and Krishna M. Baradhi. *Benign Prostatic Hyperplasia*. In: *StatPearls*. Treasure Island (FL): StatPearls Publishing, 20th Oct. 2024. PMID: 32644346.

[27] Kim E. Barrett et al., eds. *Ganong's Review of Medical Physiology*. 26th ed. McGraw-Hill's AccessMedicine. New York, N.Y: McGraw-Hill Education LLC, 2019. 1 p. ISBN: 978-1-260-12240-4.

[28] John R. Srigley. *Benign Mimickers of Prostatic Adenocarcinoma*. In: *Modern Pathology* 17.3 (1st Mar. 2004), pp. 328–348. ISSN: 0893-3952, 1530-0285. DOI: 10.1038/modpathol.3800055.

[29] Mamello Sekhoacha et al. *Prostate Cancer Review: Genetics, Diagnosis, Treatment Options, and Alternative Approaches*. In: *Molecules* 27.17 (5th Sept. 2022), p. 5730. ISSN: 1420-3049. DOI: 10.3390/molecules27175730. PMID: 36080493.

[30] Farhan Ullah Khan et al. *Comprehensive Overview of Prostatitis*. In: *Biomedicine & Pharmacotherapy* (2017). DOI: 10.1016/j.biopha.2017.08.016.

[31] Joseph H. Yacoub and Aytekin Oto. *MR Imaging of Prostate Zonal Anatomy*. In: *Radiologic Clinics of North America* (2018). DOI: 10.1016/j.rcl.2017.10.003.

[32] Freddie Bray et al. *Global Cancer Statistics 2022: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries*. In: *CA: A Cancer Journal for Clinicians* 74.3 (2024), pp. 229–263. ISSN: 1542-4863. DOI: 10.3322/caac.21834.

[33] Prashanth Rawla. *Epidemiology of Prostate Cancer*. In: *World Journal of Oncology* 10.2 (Apr. 2019), pp. 63–89. ISSN: 1920-4531. DOI: 10.14740/wjon1191. PMID: 31068988.

[34] David G. Bostwick et al. *Human Prostate Cancer Risk Factors*. In: *Cancer* 101.S10 (15th Nov. 2004), pp. 2371–2490. ISSN: 0008-543X, 1097-0142. DOI: 10.1002/cncr.20408.

[35] Nicholas D. James et al. *The Lancet Commission on Prostate Cancer: Planning for the Surge in Cases*. In: *The Lancet* 403.10437 (27th Apr. 2024), pp. 1683–1722. ISSN: 0140-6736, 1474-547X. DOI: 10.1016/S0140-6736(24)00651-2. PMID: 38583453.

[36] MaryBeth B. Culp et al. *Recent Global Patterns in Prostate Cancer Incidence and Mortality Rates*. In: *European Urology* 77.1 (1st Jan. 2020), pp. 38–52. ISSN: 0302-2838, 1873-7560. DOI: 10.1016/j.eururo.2019.08.005. PMID: 31493960.

[37] Scott E. Eggener et al. *Low-Grade Prostate Cancer: Time to Stop Calling It Cancer*. In: *Journal of Clinical Oncology* 40.27 (20th Sept. 2022), pp. 3110–3114. ISSN: 0732-183X. DOI: 10.1200/JCO.22.00123.

[38] Minas Sakellakis, Laura Jacqueline Flores and Sumankalai Ramachandran. *Patterns of Indolence in Prostate Cancer (Review)*. In: *Experimental and Therapeutic Medicine* 23.5 (1st May 2022), pp. 1–10. ISSN: 1792-0981. DOI: 10.3892/etm.2022.11278.

[39] Peter C. Albertsen. *What Is the Risk Posed by Prostate Cancer?* In: *Journal of the National Cancer Institute. Monographs* 2012.45 (Sept. 2012), pp. 169–174. ISSN: 1052-6773. DOI: 10.1093/jncimonographs/lgs028. PMID: 23271769.

[40] J Ferlay et al. *Global Cancer Observatory: Cancer Today (Version 1.1). Lyon, France: International Agency for Research on Cancer*. 1st Feb. 2024. URL: https://gco.iarc.who.int/today/ (visited on 21/05/2025).

[41] Adalberto M. Filho et al. *The GLOBOCAN 2022 Cancer Estimates: Data Sources, Methods, and a Snapshot of the Cancer Burden Worldwide*. In: *International Journal of Cancer* 156.7 (Apr. 2025), pp. 1336–1346. ISSN: 0020-7136, 1097-0215. DOI: 10.1002/ijc.35278.

[42] Sarah Collen et al. *Why Early Detection for Prostate Cancer Is Vital*. Health Awareness. 16th Mar. 2021. URL: https://www.healthawareness.co.uk/mens-healthcare/why-early-detection-for-prostate-cancer-is-vital/ (visited on 21/05/2025).

[43] Le Wang et al. *Prostate Cancer Incidence and Mortality: Global Status and Temporal Trends in 89 Countries From 2000 to 2019*. In: *Frontiers in Public Health* 10 (16th Feb. 2022), p. 811044. ISSN: 2296-2565. DOI: 10.3389/fpubh.2022.811044. PMID: 35252092.

[44] Jacqui Wise. *PSA Screening Cuts Cancer Mortality by 13%, Study Shows, amid Debate about National Programme*. In: *BMJ* 391 (30th Oct. 2025), r2288. ISSN: 1756-1833. DOI: 10.1136/bmj.r2288.

[45] Monique J. Roobol et al. *European Study of Prostate Cancer Screening — 23-Year Follow-up*. In: *New England Journal of Medicine* 393.17 (30th Oct. 2025), pp. 1669–1680. ISSN: 0028-4793, 1533-4406. DOI: 10.1056/NEJMoa2503223.

[46] Emanuel Mejia et al. *Relationship between Universal Health Insurance Benefits and Prostate Cancer Mortality in Colombia*. In: *BMC Public Health* 24.1 (30th Sept. 2024), p. 2667. ISSN: 1471-2458. DOI: 10.1186/s12889-024-20117-4.

[47] Cancer Research UK. *Prostate Cancer Statistics*. Cancer Research UK. Feb. 2025. URL: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/prostate-cancer (visited on 22/07/2025).

[48] *Localized Prostate Cancer: Learn More – Low-risk Prostate Cancer: Active Surveillance or Treatment?* In: *InformedHealth.Org [Internet]*. Institute for Quality and Efficiency in Health Care (IQWiG), 16th Sept. 2022.

[49] Neal A. Patel et al. *Grade Group 1 Prostate Cancer Outcome by Biopsy Grade and Risk Group*. In: *JAMA Oncology* 11.9 (1st Sept. 2025), pp. 1083–1087. ISSN: 2374-2437. DOI: 10.1001/jamaoncol.2025.2304.

[50] US Preventive Services Task Force. *Screening for Prostate Cancer: US Preventive Services Task Force Recommendation Statement*. In: *JAMA* 319.18 (8th May 2018), pp. 1901–1913. ISSN: 0098-7484. DOI: 10.1001/jama.2018.3710.

[51] Jiajia Chen et al. *Translational Bioinformatics for Diagnostic and Prognostic Prediction of Prostate Cancer in the Next-Generation Sequencing Era*. In: *BioMed Research International* 2013.1 (2013), p. 901578. ISSN: 2314-6141. DOI: 10.1155/2013/901578.

[52] Lionne DF Venderbos and Monique J Roobol. *PSA-based Prostate Cancer Screening: The Role of Active Surveillance and Informed and Shared Decision Making*. In: *Asian Journal of Andrology* 13.2 (Mar. 2011), pp. 219–224. ISSN: 1008-682X. DOI: 10.1038/aja.2010.180. PMID: 21297655.

[53] Seidu Mumuni, Claire O'Donnell and Owen Doody. *The Risk Factors and Screening Uptake for Prostate Cancer: A Scoping Review*. In: *Healthcare* 11.20 (Jan. 2023), p. 2780. ISSN: 2227-9032. DOI: 10.3390/healthcare11202780.

[54] Jana McHugh et al. *Prostate Cancer Risk in Men of Differing Genetic Ancestry and Approaches to Disease Screening and Management in These Groups*. In: *British Journal of Cancer* 126.10 (June 2022), pp. 1366–1373. ISSN: 1532-1827. DOI: 10.1038/s41416-021-01669-3.

[55] Abdulghani A. Naeem et al. *Prostate Cell Lines*. In: *Open Access Journal of Oncology and Medicine* 5.2 (25th Feb. 2022), pp. 534–538. ISSN: 2638-5945. DOI: 10.32474/OAJOM.2022.05.000208.

[56] Sirin Saranyutanon et al. *Cellular and Molecular Progression of Prostate Cancer: Models for Basic and Preclinical Research*. In: *Cancers* 12.9 (17th Sept. 2020), p. 2651. DOI: 10.3390/cancers12092651. PMID: 32957478.

[57] Leanne Woods-Burnham et al. *The 22Rv1 Prostate Cancer Cell Line Carries Mixed Genetic Ancestry: Implications for Prostate Cancer Health Disparities Research Using Pre-clinical Models*. In: *The Prostate* 77.16 (Dec. 2017), pp. 1601–1608. ISSN: 0270-4137, 1097-0045. DOI: 10.1002/pros.23437.

[58] Henkel Valentine et al. *Expanding the Prostate Cancer Cell Line Repertoire with ACRJ-PC28, an AR-negative Neuroendocrine Cell Line Derived From an African-Caribbean Patient*. In: *Cancer Research Communications* 2.11 (7th Nov. 2022), pp. 1355–1371. ISSN: 2767-9764. DOI: 10.1158/2767-9764.CRC-22-0245. PMID: 36643868.

[59] Mike Harris. *UK NSC Opens Consultation on Draft Prostate Cancer Screening Recommendation*. UK National Screening Committee. 28th Nov. 2025. URL: https://nationalscreening.blog.gov.uk/2025/11/28/uk-nsc-opens-consultation-on-draft-prostate-cancer-screening-recommendation/ (visited on 08/03/2026).

[60] NHS. *PSA Test*. nhs.uk. 4 Sep 2024, 1:14 p.m. URL: https://www.nhs.uk/tests-and-treatments/psa-test/ (visited on 05/11/2025).

[61] Prostate Cancer UK. *PSA Blood Test*. Prostate Cancer UK. Jan. 2025. URL: http://prostatecanceruk.org/prostate-information-and-support/prostate-tests/psa-blood-test (visited on 05/11/2025).

[62] Donald F. Gleason. *Histologic Grading of Prostate Cancer: A Perspective*. In: *Human Pathology*. The Pathobiology of Prostate Cancer-Part 1 23.3 (1st Mar. 1992), pp. 273–279. ISSN: 0046-8177. DOI: 10.1016/0046-8177(92)90108-F.

[63] Peter A. Humphrey. *Histopathology of Prostate Cancer*. In: *Cold Spring Harbor Perspectives in Medicine* 7.10 (Oct. 2017), a030411. ISSN: 2157-1422. DOI: 10.1101/cshperspect.a030411. PMID: 28389514.

[64] D. F. Gleason. *Classification of Prostatic Carcinomas*. In: *Cancer Chemotherapy Reports* 50.3 (Mar. 1966), pp. 125–128. ISSN: 0069-0112. PMID: 5948714.

[65] Alireza Abdollahi et al. *Inter/Intra-Observer Reproducibility of Gleason Scoring in Prostate Adenocarcinoma in Iranian Pathologists*. In: *Urology Journal* 9.2 (24th May 2012), pp. 486–490. ISSN: 1735-546X. DOI: 10.22037/uj.v9i2.1476.

[66] Jennifer Gordetsky and Jonathan Epstein. *Grading of Prostatic Adenocarcinoma: Current State and Prognostic Implications*. In: *Diagnostic Pathology* 11.1 (9th Mar. 2016), p. 25. ISSN: 1746-1596. DOI: 10.1186/s13000-016-0478-2.

[67] Kryvenko and Epstein. *Changes in Prostate Cancer Grading: Including a New Patient-centric Grading System*. In: *The Prostate* (2015). DOI: 10.1002/pros.23142.

[68] Jonathan I. Epstein et al. *The 2005 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma*. In: *The American Journal of Surgical Pathology* 29.9 (Sept. 2005), p. 1228. ISSN: 0147-5185. DOI: 10.1097/01.pas.0000173646.99337.b1.

[69] Geert J L H van Leenders, Esther I Verhoef and Eva Hollemans. *Prostate Cancer Growth Patterns beyond the Gleason Score: Entering a New Era of Comprehensive Tumour Grading*. In: *Histopathology* 77.6 (2020), pp. 850–861. ISSN: 1365-2559. DOI: 10.1111/his.14214.

[70] John N. Eble, Weltgesundheitsorganisation and International Agency for Research on Cancer, eds. *Pathology and Genetics of Tumours of the Urinary System and Male Genital Organs: ... Editorial and Consensus Conference in Lyon, France, December 14 - 18, 2002*. Reprint. World Health Organization Classification of Tumours 6. Lyon: IARC Press, 2006. 8 pp. ISBN: 978-92-832-2415-0 978-92-832-2412-9.

[71] Ni Chen and Qiao Zhou. *The Evolving Gleason Grading System*. In: *Chinese Journal of Cancer Research* 28.1 (Feb. 2016), pp. 58–64. ISSN: 1000-9604. DOI: 10.3978/j.issn.1000-9604.2016.02.04. PMID: 27041927.

[72] Cristina Magi-Galluzzi. *Prostate Cancer: Diagnostic Criteria and Role of Immunohistochemistry*. In: *Modern Pathology* 31.1 (Jan. 2018), pp. 12–21. ISSN: 1530-0285. DOI: 10.1038/modpathol.2017.139.

[73] Hakan Aydin et al. *Number and Location of Nucleoli and Presence of Apoptotic Bodies in Diagnostically Challenging Cases of Prostate Adenocarcinoma on Needle Biopsy*. In: *Human Pathology* 36.11 (Nov. 2005), pp. 1172–1177. ISSN: 00468177. DOI: 10.1016/j.humpath.2005.09.001.

[74] P. Denoix. *Enquête permanente dans les centres anticancéreux*. In: *Bulletin. Institut National D'hygiene* 1 (1946), pp. 12–17. ISSN: 0366-0443. PMID: 20986738.

[75] James Brierley et al., eds. *TNM Classification of Malignant Tumours*. Ninth edition. Hoboken, NJ: Wiley, 2025. 1 p. ISBN: 978-1-394-21687-1 978-1-394-21686-4.

[76] Prostate Cancer UK. *Advanced Prostate Cancer*. Prostate Cancer UK. Apr. 2023. URL: http://prostatecanceruk.org/prostate-information-and-support/just-diagnosed/advanced-prostate-cancer (visited on 04/03/2026).

[77] Stefanie Gerstberger, Qingwen Jiang and Karuna Ganesh. *Metastasis*. In: *Cell* 186.8 (13th Apr. 2023), pp. 1564–1579. ISSN: 0092-8674. DOI: 10.1016/j.cell.2023.03.003.

[78] Noel W. Clarke, Claire A. Hart and Mick D. Brown. *Molecular Mechanisms of Metastasis in Prostate Cancer*. In: *Asian Journal of Andrology* 11.1 (Jan. 2009), pp. 57–67. ISSN: 1008-682X. DOI: 10.1038/aja.2008.29. PMID: 19050684.

[79] Arthur W. Lambert, Diwakar R. Pattabiraman and Robert A. Weinberg. *Emerging Biological Principles of Metastasis*. In: *Cell* 168.4 (9th Feb. 2017), pp. 670–691. ISSN: 0092-8674, 1097-4172. DOI: 10.1016/j.cell.2016.11.037. PMID: 28187288.

[80] Jung-Kang Jin, Farshid Dayyani and Gary E. Gallick. *Steps in Prostate Cancer Progression That Lead to Bone Metastasis*. In: *International journal of cancer. Journal international du cancer* 128.11 (1st June 2011), pp. 2545–2561. ISSN: 0020-7136. DOI: 10.1002/ijc.26024. PMID: 21365645.

[81] Jeanne M Dsouza et al. *Molecular Subtypes of Metastatic Prostate Cancer: From Pathophysiology to Diagnosis*. In: *Histopathology* 88.1 (Jan. 2026), pp. 24–39. ISSN: 0309-0167, 1365-2559. DOI: 10.1111/his.70033.

[82] Glenda G. Anderson and Lawrence M. Weiss. *Determining Tissue of Origin for Metastatic Cancers: Meta-analysis and Literature Review of Immunohistochemistry Performance*. In: *Applied Immunohistochemistry & Molecular Morphology* 18.1 (Jan. 2010), pp. 3–8. ISSN: 1541-2016. DOI: 10.1097/PAI.0b013e3181a75e6d.

[83] Alberto M. Marchevsky, Ruta Gupta and Bonnie Balzer. *Diagnosis of Metastatic Neoplasms: A Clinicopathologic and Morphologic Approach*. In: *Archives of Pathology & Laboratory Medicine* 134.2 (1st Feb. 2010), pp. 194–206. ISSN: 1543-2165, 0003-9985. DOI: 10.5858/134.2.194.

[84] Peter Brawn. *Histologic Features of Metastatic Prostate Cancer*. In: *Human Pathology*. The Pathobiology of Prostate Cancer-Part 1 23.3 (1st Mar. 1992), pp. 267–272. ISSN: 0046-8177. DOI: 10.1016/0046-8177(92)90107-E.

[85] National Institute for Health and Care Excellence (NICE). *NG12 Evidence Reviews for Diagnostic Accuracy of Prostate Specific Antigen (PSA) Thresholds for Referring People with Suspected Prostate Cancer*. National Institute for Health and Care Excellence (NICE), Dec. 2021.

[86] National Institute for Health and Care Excellence (NICE). *Prostate Cancer: Diagnosis and Management*. NICE Guideline NG131. London: National Institute for Health and Care Excellence, 15th Dec. 2021. ISBN: 978-1-4731-3375-4.

[87] NHS. *Tests and next Steps for Prostate Cancer*. nhs.uk. 31st July 2025. URL: https://www.nhs.uk/conditions/prostate-cancer/tests-and-next-steps/ (visited on 03/03/2026).

[88] Amelia Swift, Roberta Heale and Alison Twycross. *What Are Sensitivity and Specificity?* In: *Evidence-Based Nursing* 23.1 (1st Jan. 2020), pp. 2–4. ISSN: 1367-6539, 1468-9618. DOI: 10.1136/ebnurs-2019-103225. PMID: 31719126.

[89] Anthony K Akobeng. *Understanding Diagnostic Tests 1: Sensitivity, Specificity and Predictive Values*. In: *Acta Paediatrica* 96.3 (2007), pp. 338–341. ISSN: 1651-2227. DOI: 10.1111/j.1651-2227.2006.00180.x.

[90] A. El-Shater Bosaily et al. *PROMIS — Prostate MR Imaging Study: A Paired Validating Cohort Study Evaluating the Role of Multi-Parametric MRI in Men with Clinical Suspicion of Prostate Cancer*. In: *Contemporary Clinical Trials* 42 (May 2015), pp. 26–40. ISSN: 1551-7144. DOI: 10.1016/j.cct.2015.02.008. PMID: 25749312.

[91] Hashim U. Ahmed et al. *Diagnostic Accuracy of Multi-Parametric MRI and TRUS Biopsy in Prostate Cancer (PROMIS): A Paired Validating Confirmatory Study*. In: *The Lancet* 389.10071 (25th Feb. 2017), pp. 815–822. ISSN: 0140-6736, 1474-547X. DOI: 10.1016/S0140-6736(16)32401-1. PMID: 28110982.

[92] Joseph M. Norris et al. *What Type of Prostate Cancer Is Systematically Overlooked by Multiparametric Magnetic Resonance Imaging? An Analysis from the PROMIS Cohort*. In: *European Urology* 78.2 (Aug. 2020), pp. 163–170. ISSN: 0302-2838. DOI: 10.1016/j.eururo.2020.04.029. PMID: 32370911.

[93] Giovanni Tarantino et al. *Clinical Factors Affecting Prostate-Specific Antigen Levels in Prostate Cancer Patients Undergoing Radical Prostatectomy: A Retrospective Study*. In: *Future Science OA* 7.3 (1st Mar. 2021), FSO643. ISSN: null. DOI: 10.2144/fsoa-2020-0154. PMID: 33552540.

[94] Tasneem Khan et al. *Understanding of PSA Biology, Factors Affecting PSA Detection, Challenges, Various Biomarkers, Methods, and Future Perspective of Prostate Cancer Detection and Diagnosis*. In: *Advances in Cancer Biology - Metastasis* 5 (1st Oct. 2022), p. 100059. ISSN: 2667-3940. DOI: 10.1016/j.adcanc.2022.100059.

[95] Michael K. David and Stephen W. Leslie. *Prostate-Specific Antigen*. In: *StatPearls*. Treasure Island (FL): StatPearls Publishing, 2025. PMID: 32491427.

[96] Richard M. Martin et al. *Prostate-Specific Antigen Screening and 15-Year Prostate Cancer Mortality: A Secondary Analysis of the CAP Randomized Clinical Trial*. In: *JAMA* 331.17 (7th May 2024), pp. 1460–1470. ISSN: 0098-7484. DOI: 10.1001/jama.2024.4011.

[97] Hari T. Vigneswaran et al. *Stockholm3 Versus Prostate-specific Antigen in Prostate Cancer Screening: 9-Year Outcomes Demonstrating Improved Detection of Aggressive Cancers and Reduced Overdiagnosis from the STHLM3 Trial*. In: *European Urology* 89.1 (1st Jan. 2026), pp. 82–90. ISSN: 0302-2838. DOI: 10.1016/j.eururo.2025.10.001.

[98] Samuel W. D. Merriel et al. *Systematic Review and Meta-Analysis of the Diagnostic Accuracy of Prostate-Specific Antigen (PSA) for the Detection of Prostate Cancer in Symptomatic Patients*. In: *BMC Medicine* 20.1 (7th Feb. 2022), p. 54. ISSN: 1741-7015. DOI: 10.1186/s12916-021-02230-y.

[99] Jose Leal et al. *Estimating the Sensitivity of a Prostate Cancer Screening Programme for Different PSA Cut-off Levels: A UK Case Study*. In: *Cancer Epidemiology* 52 (1st Feb. 2018), pp. 99–105. ISSN: 1877-7821. DOI: 10.1016/j.canep.2017.12.002.

[100] Ian M. Thompson et al. *Prevalence of Prostate Cancer among Men with a Prostate-Specific Antigen Level <4.0 Ng per Milliliter*. In: *New England Journal of Medicine* 350.22 (27th May 2004), pp. 2239–2246. ISSN: 0028-4793. DOI: 10.1056/NEJMoa031918.

[101] Ian M. Thompson et al. *Operating Characteristics of Prostate-Specific Antigen in Men With an Initial PSA Level of 3.0 Ng/mL or Lower*. In: *JAMA* 294.1 (6th July 2005), pp. 66–70. ISSN: 0098-7484. DOI: 10.1001/jama.294.1.66.

[102] Leen Naji et al. *Digital Rectal Examination for Prostate Cancer Screening in Primary Care: A Systematic Review and Meta-Analysis*. In: *The Annals of Family Medicine* 16.2 (1st Mar. 2018), pp. 149–154. ISSN: 1544-1709, 1544-1717. DOI: 10.1370/afm.2205. PMID: 29531107.

[103] Daniel Jones et al. *The Diagnostic Test Accuracy of Rectal Examination for Prostate Cancer Diagnosis in Symptomatic Patients: A Systematic Review*. In: *BMC Family Practice* 19.1 (2nd June 2018), p. 79. ISSN: 1471-2296. DOI: 10.1186/s12875-018-0765-y.

[104] Niranjan J. Sathianathen et al. *Negative Predictive Value of Multiparametric Magnetic Resonance Imaging in the Detection of Clinically Significant Prostate Cancer in the Prostate Imaging Reporting and Data System Era: A Systematic Review and Meta-analysis*. In: *European Urology* 78.3 (1st Sept. 2020), pp. 402–414. ISSN: 0302-2838. DOI: 10.1016/j.eururo.2020.03.048.

[105] Paul C. Moldovan et al. *What Is the Negative Predictive Value of Multiparametric Magnetic Resonance Imaging in Excluding Prostate Cancer at Biopsy? A Systematic Review and Meta-analysis from the European Association of Urology Prostate Cancer Guidelines Panel*. In: *European Urology* (2017). DOI: 10.1016/j.eururo.2017.02.026.

[106] Maria Clara Fernandes et al. *The Role of MRI in Prostate Cancer: Current and Future Directions*. In: *Magnetic Resonance Materials in Physics, Biology and Medicine* 35.4 (1st Aug. 2022), pp. 503–521. ISSN: 1352-8661. DOI: 10.1007/s10334-022-01006-6.

[107] Hiten D. Patel et al. *Variability in Prostate Cancer Detection among Radiologists and Urologists Using MRI Fusion Biopsy*. In: *BJUI Compass* 5.2 (2024), pp. 304–312. ISSN: 2688-4526. DOI: 10.1002/bco2.294.

[108] Camille Berquin et al. *Transperineal vs. Transrectal Prostate Biopsies under Local Anesthesia: A Prospective Cohort Study on Patient Tolerability and Complication Rates*. In: *Urologic Oncology: Seminars and Original Investigations* 41.9 (1st Sept. 2023), 388.e17–388.e23. ISSN: 1078-1439. DOI: 10.1016/j.urolonc.2023.05.020.

[109] Emiel Ruijter et al. *Errors in Histological Grading by Prostatic Needle Biopsy Specimens: Frequency and Predisposing Factors*. In: *The Journal of Pathology* 192.2 (2000), pp. 229–233. ISSN: 1096-9896. DOI: 10.1002/1096-9896(2000)9999:9999<::AID-PATH703>3.0.CO;2-X.

[110] Manish A. Vira et al. *Impact of the Percentage of Positive Biopsy Cores on the Further Stratification of Primary Grade 3 and Grade 4 Gleason Score 7 Tumors in Radical Prostatectomy Patients*. In: *Urology* 66.5 (1st Nov. 2005), pp. 1015–1019. ISSN: 0090-4295, 1527-9995. DOI: 10.1016/j.urology.2005.05.051.

[111] Rui M. Bernardino et al. *Limitations of Prostate Biopsy in Detection of Cribriform and Intraductal Prostate Cancer*. In: *European Urology Focus* 10.1 (1st Jan. 2024), pp. 146–153. ISSN: 2405-4569. DOI: 10.1016/j.euf.2023.08.010.

[112] Okyaz Eminaga et al. *Critical Evaluation of Artificial Intelligence as a Digital Twin of Pathologists for Prostate Cancer Pathology*. In: *Scientific Reports* 14.1 (4th Mar. 2024), p. 5284. ISSN: 2045-2322. DOI: 10.1038/s41598-024-55228-w.

[113] Sakina Mohammed Mota et al. *Artificial Intelligence Improves the Ability of Physicians to Identify Prostate Cancer Extent*. In: *The Journal of Urology* (July 2024). DOI: 10.1097/JU.0000000000003960.

[114] C. Eloy et al. *Clinical Advantages in Providing Artificial Intelligence-Assisted Prostate Cancer Diagnosis: A Pilot Study*. In: *Pathology - Research and Practice* 271 (1st July 2025), p. 156007. ISSN: 0344-0338. DOI: 10.1016/j.prp.2025.156007.

[115] John Abel et al. *AI Powered Quantification of Nuclear Morphology in Cancers Enables Prediction of Genome Instability and Prognosis*. In: *npj Precision Oncology* 8.1 (19th June 2024), p. 134. ISSN: 2397-768X. DOI: 10.1038/s41698-024-00623-9.

[116] Zeyu Luo et al. *Study of AI Algorithms on mpMRI and PHI for the Diagnosis of Clinically Significant Prostate Cancer*. In: *Urologic Oncology: Seminars and Original Investigations* 43.9 (1st Sept. 2025), 527.e17–527.e24. ISSN: 1078-1439. DOI: 10.1016/j.urolonc.2025.05.007.

[117] Yibo Xu et al. *Feasibility Study of AI-assisted Multi-Parameter MRI Diagnosis of Prostate Cancer*. In: *Scientific Reports* 15.1 (27th Mar. 2025), p. 10530. ISSN: 2045-2322. DOI: 10.1038/s41598-024-84516-8.

[118] Francesco Giganti et al. *AI-powered Prostate Cancer Detection: A Multi-Centre, Multi-Scanner Validation Study*. In: *European Radiology* 35.8 (1st Aug. 2025), pp. 4915–4924. ISSN: 1432-1084. DOI: 10.1007/s00330-024-11323-0.

[119] Mieke Van Hemelrijck et al. *Reasons for Discontinuing Active Surveillance: Assessment of 21 Centres in 12 Countries in the Movember GAP3 Consortium*. In: *European Urology* 75.3 (1st Mar. 2019), pp. 523–531. ISSN: 0302-2838. DOI: 10.1016/j.eururo.2018.10.025.

[120] Zaure Dushimova et al. *Overdiagnosis and Overtreatment in Prostate Cancer*. In: *Diseases* 13.6 (6 June 2025), p. 167. ISSN: 2079-9721. DOI: 10.3390/diseases13060167.

[121] Renée C. A. Leenen et al. *Prostate Cancer Early Detection in the European Union and UK*. In: *European Urology* 87.3 (1st Mar. 2025), pp. 326–339. ISSN: 0302-2838. DOI: 10.1016/j.eururo.2024.07.019.

[122] Qintao Ge et al. *Molecular Classifications of Prostate Cancer: Basis for Individualized Risk Stratification and Precision Therapy*. In: *Annals of Medicine* 55.2 (11th Aug. 23), p. 2279235. ISSN: 0785-3890. DOI: 10.1080/07853890.2023.2279235. PMID: 37939258.

[123] Csilla Olah et al. *Enhancing Risk Stratification Models in Localized Prostate Cancer by Novel Validated Tissue Biomarkers*. In: *Prostate Cancer and Prostatic Diseases* 28.3 (Sept. 2025), pp. 773–781. ISSN: 1476-5608. DOI: 10.1038/s41391-024-00918-9.

[124] Tina Keshavarzian et al. *Prostate Cancer Cells Converge to an Inflammatory-like State upon Metastatic Dissemination*. In: *Nature Communications* 16.1 (23rd Dec. 2025), p. 11339. ISSN: 2041-1723. DOI: 10.1038/s41467-025-67856-5.

[125] Jianpeng Xue et al. *Identifying Metastatic Ability of Prostate Cancer Cell Lines Using Native Fluorescence Spectroscopy and Machine Learning Methods*. In: *Scientific Reports* 11.1 (26th Jan. 2021), p. 2282. ISSN: 2045-2322. DOI: 10.1038/s41598-021-81945-7.

[126] Alexandra Sorvina et al. *Lipid Profiles of Prostate Cancer Cells*. In: *Oncotarget* 9.85 (30th Oct. 2018), pp. 35541–35552. ISSN: 1949-2553. DOI: 10.18632/oncotarget.26222. PMID: 30473749.

[127] Daniel West, Susan Stepney and Y. Hancock. *Unsupervised Self-Organising Map Classification of Raman Spectra from Prostate Cell Lines Uncovers Substratified Prostate Cancer Disease States*. In: *Scientific Reports* 15.1 (4th Jan. 2025), p. 773. ISSN: 2045-2322. DOI: 10.1038/s41598-024-83708-6.

[128] Eva Compérat et al. *Diagnosis of Prostate Cancer in One Day: The Benefits of Cytology in Tumour Detection*. In: *Cytopathology* 32.2 (2021), pp. 211–216. ISSN: 1365-2303. DOI: 10.1111/cyt.12930.

[129] Hanbing Song et al. *Single-Cell Analysis of Human Primary Prostate Cancer Reveals the Heterogeneity of Tumor-Associated Epithelial Cell States*. In: *Nature Communications* 13.1 (10th Jan. 2022), p. 141. ISSN: 2041-1723. DOI: 10.1038/s41467-021-27322-4.

[130] Ana Rita Lima et al. *Discrimination between the Human Prostate Normal and Cancer Cell Exometabolome by GC-MS*. In: *Scientific Reports* 8.1 (3rd Apr. 2018), p. 5539. ISSN: 2045-2322. DOI: 10.1038/s41598-018-23847-9.

[131] Ishita Singh and Tanmay P. Lele. *Nuclear Morphological Abnormalities in Cancer – a Search for Unifying Mechanisms*. In: *Results and problems in cell differentiation* 70 (2022), pp. 443–467. ISSN: 0080-1844. DOI: 10.1007/978-3-031-06573-6_16. PMID: 36348118.

[132] Robert W Veltri, Christhunesa S Christudass and Sumit Isharwal. *Nuclear Morphometry, Nucleomics and Prostate Cancer Progression*. In: *Asian Journal of Andrology* 14.3 (May 2012), pp. 375–384. ISSN: 1008-682X. DOI: 10.1038/aja.2011.148. PMID: 22504875.

[133] Gozde Kır, Billur Cosan Sarbay and Eyup Gumus. *Significance and Outcome of Nuclear Anaplasia and Mitotic Index in Prostatic Adenocarcinomas*. In: *Urologic Oncology: Seminars and Original Investigations* (2016). DOI: 10.1016/j.urolonc.2016.04.017.

[134] Guido Di Donato, Renée Laufer-Amorim and Chiara Palmieri. *Nuclear Morphometry in Histological Specimens of Canine Prostate Cancer: Correlation with Histological Subtypes, Gleason Score, Methods of Collection and Survival Time*. In: *Research in Veterinary Science* 114 (1st Oct. 2017), pp. 212–217. ISSN: 0034-5288. DOI: 10.1016/j.rvsc.2017.05.007.

[135] H.G. Van Der Poel et al. *Intratumoral Nuclear Morphologic Heterogeneity in Prostate Cancer*. In: *Urology* 49.4 (Apr. 1997), pp. 652–657. ISSN: 00904295. DOI: 10.1016/S0090-4295(96)00557-2.

[136] Lorenzo Montanaro, Davide Treré and Massimo Derenzini. *Nucleolus, Ribosomes, and Cancer*. In: *The American Journal of Pathology* 173.2 (Aug. 2008), pp. 301–310. ISSN: 0002-9440. DOI: 10.2353/ajpath.2008.070752. PMID: 18583314.

[137] Ines Orsolic et al. *The Relationship between the Nucleolus and Cancer: Current Evidence and Emerging Paradigms*. In: *Seminars in Cancer Biology*. Systemic DNA Damage: Mechanisms, Effects and Mitigation Strategies 37–38 (1st June 2016), pp. 36–50. ISSN: 1044-579X. DOI: 10.1016/j.semcancer.2015.12.004.

[138] Marianna Penzo et al. *The Ribosome Biogenesis—Cancer Connection*. In: *Cells* 8.1 (15th Jan. 2019), p. 55. ISSN: 2073-4409. DOI: 10.3390/cells8010055. PMID: 30650663.

[139] Mohamed Kodiha, Piotr Bański and Ursula Stochaj. *Computer-Based Fluorescence Quantification: A Novel Approach to Study Nucleolar Biology*. In: *BMC Cell Biology* 12.1 (3rd June 2011), p. 25. ISSN: 1471-2121. DOI: 10.1186/1471-2121-12-25.

[140] Myron Tannenbaum, Peter Desanctis and Carl Olsson. *Prognostic Significance of Nucleolar Surface Area in Prostate Cancer*. In: *Urology* XIX.5 (May 1982). DOI: 10.1016/0090-4295(82)90618-5.

[141] Jaclyn E. Quin et al. *Targeting the Nucleolus for Cancer Intervention*. In: *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*. Role of the Nucleolus in Human Disease 1842.6 (1st June 2014), pp. 802–816. ISSN: 0925-4439. DOI: 10.1016/j.bbadis.2013.12.009.

[142] Daniel Westreich, Brian W. Pence and Abigail Norris Turner. *In Populo*. In: *Epidemiology (Cambridge, Mass.)* 21.1 (Jan. 2010), pp. 152–153. ISSN: 1044-3983. DOI: 10.1097/EDE.0b013e3181c1e92a. PMID: 20010219.

[143] Galfridus Anglicus. *The Promptorium Parvulorum: The First English-Latin Dictionary*. Ed. by Anthony Lawson Mayhew. Compiled circa A.D. 1440. Vol. 102. Early English Text Society. Extra Series. London: Published for the Early English Text Society by K. Paul, Trench, Trübner & Co.; and H. Frowde, 1908.

[144] Verena Sailer et al. *Experimental in Vitro, Ex Vivo and in Vivo Models in Prostate Cancer Research*. In: *Nature Reviews Urology* 20.3 (Mar. 2023), pp. 158–178. ISSN: 1759-4812, 1759-4820. DOI: 10.1038/s41585-022-00677-z.

[145] David Cunningham and Zongbing You. *In Vitro and in Vivo Model Systems Used in Prostate Cancer Research*. In: *Journal of Biological Methods* 2.1 (4th June 2015), p. 1. ISSN: TBA. DOI: 10.14440/jbm.2015.63.

[146] Wout Devlies et al. *Preclinical Models in Prostate Cancer: Resistance to AR Targeting Therapies in Prostate Cancer*. In: *Cancers* 13.4 (22nd Feb. 2021), p. 915. ISSN: 2072-6694. DOI: 10.3390/cancers13040915. PMID: 33671614.

[147] Jonathan M. Matthews et al. *OrganoID: A Versatile Deep Learning Platform for Tracking and Analysis of Single-Organoid Dynamics*. In: *PLOS Computational Biology* 18.11 (9th Nov. 2022), e1010584. ISSN: 1553-734X. DOI: 10.1371/journal.pcbi.1010584. PMID: 36350878.

[148] Robert Adamiecki et al. *In Vivo Models for Prostate Cancer Research*. In: *Cancers* 14.21 (Jan. 2022), p. 5321. ISSN: 2072-6694. DOI: 10.3390/cancers14215321.

[149] Debasish Basak et al. *Preclinical and Clinical Research Models of Prostate Cancer: A Brief Overview*. In: *Life* 12.10 (Oct. 2022), p. 1607. ISSN: 2075-1729. DOI: 10.3390/life12101607.

[150] Sylwia Hasterok, Anna Gustafsson and Anette Gjörloff Wingren. *Applications of Tumor Cells in an In Vitro 3D Environment*. In: *Applied Sciences* 13.18 (Jan. 2023), p. 10349. ISSN: 2076-3417. DOI: 10.3390/app131810349.

[151] S. H. Lang et al. *Prostate Epithelial Cell Lines Form Spheroids with Evidence of Glandular Differentiation in Three-Dimensional Matrigel Cultures*. In: *British Journal of Cancer* 85.4 (Aug. 2001), pp. 590–599. ISSN: 1532-1827. DOI: 10.1054/bjoc.2001.1967.

[152] William Buchser et al. *Assay Development Guidelines for Image-Based High Content Screening, High Content Analysis and High Content Imaging*. In: *Assay Guidance Manual*. Ed. by Sarine Markossian et al. Bethesda (MD): Eli Lilly & Company and the National Center for Advancing Translational Sciences, 2004. PMID: 23035272.

[153] Chandra Prakash Prasad et al. *Passage Number of Cancer Cell Lines: Importance, Intricacies, and Way-Forward*. In: *Biotechnology and Bioengineering* 120.8 (2023), pp. 2049–2055. ISSN: 1097-0290. DOI: 10.1002/bit.28496.

[154] K. F. Idrisova, H.-U. Simon and M. O. Gomzikova. *Role of Patient-Derived Models of Cancer in Translational Oncology*. In: *Cancers* 15.1 (26th Dec. 2022), p. 139. ISSN: 2072-6694. DOI: 10.3390/cancers15010139. PMID: 36612135.

[155] Cory Abate-Shen and Francisca Nunes de Almeida. *Establishment of the LNCaP Cell Line – the Dawn of an Era for Prostate Cancer Research*. In: *Cancer research* 82.9 (3rd May 2022), pp. 1689–1691. ISSN: 0008-5472. DOI: 10.1158/0008-5472.CAN-22-1065. PMID: 35502546.

[156] Steven Kregel et al. *Acquired Resistance to the Second-Generation Androgen Receptor Antagonist Enzalutamide in Castration-Resistant Prostate Cancer*. In: *Oncotarget* 7.18 (29th Mar. 2016), pp. 26259–26274. ISSN: 1949-2553. DOI: 10.18632/oncotarget.8456.

[157] Leire Moya et al. *Characterisation of Cell Lines Derived from Prostate Cancer Patients with Localised Disease*. In: *Prostate Cancer and Prostatic Diseases* 26.3 (Sept. 2023), pp. 614–624. ISSN: 1476-5608. DOI: 10.1038/s41391-023-00679-x.

[158] Lisa Mohr et al. *Generation of Prostate Cancer Cell Models of Resistance to the Anti-mitotic Agent Docetaxel*. In: *Journal of Visualized Experiments : JoVE* 127 (8th Sept. 2017), p. 56327. ISSN: 1940-087X. DOI: 10.3791/56327. PMID: 28930981.

[159] R.E. Sobel and M.D. Sadar. *CELL LINES USED IN PROSTATE CANCER RESEARCH: A COMPENDIUM OF OLD AND NEW LINES—PART 1*. In: *Journal of Urology* 173.2 (Feb. 2005), pp. 342–359. ISSN: 0022-5347, 1527-3792. DOI: 10.1097/01.ju.0000141580.30910.57.

[160] R.E. Sobel and M.D. Sadar. *CELL LINES USED IN PROSTATE CANCER RESEARCH: A COMPENDIUM OF OLD AND NEW LINES—PART 2*. In: *Journal of Urology* 173.2 (Feb. 2005), pp. 360–372. ISSN: 0022-5347, 1527-3792. DOI: 10.1097/01.ju.0000149989.01263.dc.

[161] Xinyu Wu et al. *Current Mouse and Cell Models in Prostate Cancer Research*. In: *Endocrine-related cancer* 20.4 (Aug. 2013), 10.1530/ERC-12–0285. ISSN: 1351-0088. DOI: 10.1530/ERC-12-0285. PMID: 23580590.

[162] M Rumsby et al. *Human Prostate Cell Lines from Normal and Tumourigenic Epithelia Differ in the Pattern and Control of Choline Lipid Headgroups Released into the Medium on Stimulation of Protein Kinase C*. In: *British Journal of Cancer* 104.4 (15th Feb. 2011), pp. 673–684. ISSN: 0007-0920. DOI: 10.1038/sj.bjc.6606077. PMID: 21266973.

[163] *LNCaP Clone FGC - CRL-1740*. ATCC. URL: https://www.atcc.org/products/crl-1740 (visited on 31/01/2026).

[164] Khalid Ahmed et al. *Comparison of Baseline Global Gene Expression Profiles of Prostate Cancer Cell Lines LNCaP and DU145*. In: *BMC Research Notes* 17.1 (31st Dec. 2024), p. 398. ISSN: 1756-0500. DOI: 10.1186/s13104-024-07050-w.

[165] Zenon Gibas et al. *A High-Resolution Study of Chromosome Changes in a Human Prostatic Carcinoma Cell Line (LNCaP)*. In: *Cancer Genetics and Cytogenetics* 11.4 (1st Apr. 1984), pp. 399–404. ISSN: 0165-4608. DOI: 10.1016/0165-4608(84)90020-7.

[166] O. Cussenot et al. *Immortalization of Human Adult Normal Prostatic Epithelial Cells by Liposomes Containing Large T-SV40 Gene*. In: *The Journal of Urology* 146.3 (1st Sept. 1991), pp. 881–886. ISSN: 0022-5347. DOI: 10.1016/S0022-5347(17)37953-3.

[167] Maria E Mycielska and Mustafa B A Djamgoz. *Citrate Transport in the Human Prostate Epithelial PNT2-C2 Cell Line: Electrophysiological Analyses*. In: *The Journal of Physiology* 559 (Pt 3 15th Sept. 2004), pp. 821–833. ISSN: 0022-3751. DOI: 10.1113/jphysiol.2004.067850. PMID: 15254154.

[168] M. Cameron et al. *Raman Spectroscopy Reveals Oxidative Stress-Induced Metabolic Vulnerabilities in Early-Stage AR-negative Prostate-Cancer versus Normal-Prostate Cell Lines*. In: *Scientific Reports* 14.1 (25th Oct. 2024), p. 25388. ISSN: 2045-2322. DOI: 10.1038/s41598-024-70338-1.

[169] Sara L. Cole and M. J. Tevethia. *Simian Virus 40 Large T Antigen and Two Independent T-Antigen Segments Sensitize Cells to Apoptosis Following Genotoxic Damage*. In: *Journal of Virology* 76.16 (Aug. 2002), pp. 8420–8432. ISSN: 0022-538X. DOI: 10.1128/JVI.76.16.8420-8432.2002. PMID: 12134045.

[170] Syed Hamid Ali and James A. DeCaprio. *Cellular Transformation by SV40 Large T Antigen: Interaction with Host Proteins*. In: *Seminars in Cancer Biology* 11.1 (1st Feb. 2001), pp. 15–22. ISSN: 1044-579X. DOI: 10.1006/scbi.2000.0342.

[171] Stella Corsetti et al. *Raman Spectroscopy for Accurately Characterizing Biomolecular Changes in Androgen-independent Prostate Cancer Cells*. In: *Journal of Biophotonics* 11.3 (Mar. 2018), e201700166. ISSN: 1864-063X. DOI: 10.1002/jbio.201700166. PMID: 28925566.

[172] Aaron M. Horning et al. *Single-Cell RNA-seq Reveals a Subpopulation of Prostate Cancer Cells with Enhanced Cell Cycle-Related Transcription and Attenuated Androgen Response*. In: *Cancer research* 78.4 (15th Feb. 2018), pp. 853–864. ISSN: 0008-5472. DOI: 10.1158/0008-5472.CAN-17-1924. PMID: 29233929.

[173] S. Taavitsainen et al. *Single-Cell ATAC and RNA Sequencing Reveal Pre-Existing and Persistent Cells Associated with Prostate Cancer Relapse*. In: *Nature Communications* 12.1 (6th Sept. 2021), p. 5307. ISSN: 2041-1723. DOI: 10.1038/s41467-021-25624-1.

[174] N. Poulin et al. *Risk Biomarker Assessment for Breast Cancer Progression: Replication Precision of Nuclear Morphometry*. In: *Analytical Cellular Pathology* 25.3 (Jan. 2003), pp. 129–138. ISSN: 2210-7177, 2210-7185. DOI: 10.1155/2003/262918.

[175] The Inkscape Project. *Inkscape 1.4*. Version 1.4. 13th Oct. 2024.

[176] Maciej Roman et al. *Lipid Droplets in Prostate Cancer Cells and Effect of Irradiation Studied by Raman Microspectroscopy*. In: *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* 1865.9 (1st Sept. 2020), p. 158753. ISSN: 1388-1981. DOI: 10.1016/j.bbalip.2020.158753.

[177] Shailender S. Chauhan et al. *PIM1 Drives Lipid Droplet Accumulation to Promote Proliferation and Survival in Prostate Cancer*. In: *Oncogene* 43.6 (2024), pp. 406–419. ISSN: 0950-9232. DOI: 10.1038/s41388-023-02914-0. PMID: 38097734.

[178] Sriram Venigalla, Chen Zhao and Hiroshi Miyamoto. *Histopathologic Features of Atypical Glands on Prostate Biopsy: Nucleolar Size Is a Predictor of Subsequent Detection of Prostatic Adenocarcinoma*. In: *The Prostate* 73.4 (Mar. 2013), pp. 376–381. ISSN: 1097-0045. DOI: 10.1002/pros.22577. PMID: 22926960.

[179] Geoffrey M. Cooper. *The Cell: A Molecular Approach*. Eighth edition. New York Oxford: Sinauer Associates, Oxford University Press, 2019. 783 pp. ISBN: 978-1-60535-707-2.

[180] David Rintoul, ed. *Student Companion for Molecular Cell Biology*. 3. ed. New York: Scientific American Books, 1995. 518 pp. ISBN: 978-0-7167-2672-2.

[181] Amanda L. Rynearson and Caroline R. Sussman. *Nuclear Structure, Organization, and Oncogenesis*. In: *Journal of Gastrointestinal Cancer* 42.2 (June 2011), pp. 112–117. ISSN: 1941-6628, 1941-6636. DOI: 10.1007/s12029-011-9253-5.

[182] Stephen W. Leslie, Taylor L. Soon-Sutton and William P. Skelton. *Prostate Cancer*. In: *StatPearls*. Treasure Island (FL): StatPearls Publishing, 2025. PMID: 29261872.

[183] Daniele Zink, Andrew H. Fischer and Jeffrey A. Nickerson. *Nuclear Structure in Cancer Cells*. In: *Nature Reviews Cancer* 4.9 (Sept. 2004), pp. 677–687. ISSN: 1474-175X, 1474-1768. DOI: 10.1038/nrc1430.

[184] Edgar G. Fischer. *Nuclear Morphology and the Biology of Cancer Cells*. In: *Acta Cytologica* 64.6 (2020), pp. 511–519. ISSN: 0001-5547, 1938-2650. DOI: 10.1159/000508780.

[185] R. W. Veltri et al. *Quantitative Nuclear Morphometry, Markovian Texture Descriptors, and DNA Content Captured on a CAS-200 Image Analysis System, Combined with PCNA and HER-2/Neu Immunohistochemistry for Prediction of Prostate Cancer Progression*. In: *Journal of Cellular Biochemistry. Supplement* 19 (1994), pp. 249–258. ISSN: 0733-1959. PMID: 7529856.

[186] P. H. Bartels et al. *Nuclear Chromatin Texture in Prostatic Lesions. II. PIN and Malignancy Associated Changes*. In: *Analytical and Quantitative Cytology and Histology* 20.5 (Oct. 1998), pp. 397–406. PMID: 9801758.

[187] Tarjei S. Hveem et al. *Chromatin Changes Predict Recurrence after Radical Prostatectomy*. In: *British Journal of Cancer* 114.11 (24th May 2016), pp. 1243–1250. ISSN: 1532-1827. DOI: 10.1038/bjc.2016.96. PMID: 27124335.

[188] Robert W. Veltri et al. *Nuclear Roundness Variance Predicts Prostate Cancer Progression, Metastasis, and Death: A Prospective Evaluation with up to 25 Years of Follow-up after Radical Prostatectomy*. In: *The Prostate* 70.12 (1st Sept. 2010), pp. 1333–1339. ISSN: 1097-0045. DOI: 10.1002/pros.21168. PMID: 20623633.

[189] Robert W. Veltri, Alan W. Partin and M. Craig Miller. *Quantitative Nuclear Grade (QNG): A New Image Analysis-Based Biomarker of Clinically Relevant Nuclear Structure Alterations*. In: *Journal of Cellular Biochemistry* 79.S35 (2000), pp. 151–157. ISSN: 1097-4644. DOI: 10.1002/1097-4644(2000)79:35+<151::AID-JCB1139>3.0.CO;2-7.

[190] Dmytro Fishman et al. *Practical Segmentation of Nuclei in Brightfield Cell Images with Neural Networks Trained on Fluorescently Labelled Samples*. In: *Journal of Microscopy* 284.1 (2021), pp. 12–24. ISSN: 1365-2818. DOI: 10.1111/jmi.13038.

[191] Robert Mandelkow et al. *Detection and Quantification of Nuclear Morphology Changes in Apoptotic Cells by Fluorescence Microscopy and Subsequent Analysis of Visualized Fluorescent Signals*. In: *Anticancer Research* 37.5 (1st May 2017), pp. 2239–2244. ISSN: 0250-7005, 1791-7530. PMID: 28476788.

[192] Andrea Rizzotto et al. *Reduction in Nuclear Size by DHRS7 in Prostate Cancer Cells and by Estradiol Propionate in DHRS7-Depleted Cells*. In: *Cells* 13.1 (Jan. 2024), p. 57. ISSN: 2073-4409. DOI: 10.3390/cells13010057.

[193] Vassilis Roukos et al. *Cell Cycle Staging of Individual Cells by Fluorescence Microscopy*. In: *Nature protocols* 10.2 (Feb. 2015), pp. 334–348. ISSN: 1754-2189. DOI: 10.1038/nprot.2015.016. PMID: 25633629.

[194] Bruce Alberts et al. *An Overview of the Cell Cycle*. In: *Molecular Biology of the Cell. 4th Edition*. Garland Science, 2002.

[195] Ugo Testa, Germana Castelli and Elvira Pelosi. *Cellular and Molecular Mechanisms Underlying Prostate Cancer Development: Therapeutic Implications*. In: *Medicines* 6.3 (30th July 2019), p. 82. ISSN: 2305-6320. DOI: 10.3390/medicines6030082. PMID: 31366128.

[196] Tatjana Vlajnic et al. *Exploring the Intratumoral Heterogeneity of DNA Ploidy in Prostate Cancer*. In: *Cancer Reports* 7.2 (26th Dec. 2023), e1953. ISSN: 2573-8348. DOI: 10.1002/cnr2.1953. PMID: 38148577.

[197] Emily S. Bell and Jan Lammerding. *Causes and Consequences of Nuclear Envelope Alterations in Tumour Progression*. In: *European Journal of Cell Biology*. Integrated Mechano-Chemical Signals in Invasion 95.11 (1st Nov. 2016), pp. 449–464. ISSN: 0171-9335. DOI: 10.1016/j.ejcb.2016.06.007.

[198] Toyoshi Fujimoto. *Nuclear Lipid Droplets – How Are They Different from Their Cytoplasmic Siblings?* In: *Journal of Cell Science* 135.5 (25th Feb. 2022), jcs259253. ISSN: 0021-9533. DOI: 10.1242/jcs.259253.

[199] Aino Siltari et al. *Role of Lipids and Lipid Metabolism in Prostate Cancer Progression and the Tumor's Immune Environment*. In: *Cancers* 14.17 (1st Sept. 2022), p. 4293. ISSN: 2072-6694. DOI: 10.3390/cancers14174293. PMID: 36077824.

[200] Farial A. Tanious et al. *DAPI (4',6-Diamidino-2-Phenylindole) Binds Differently to DNA and RNA: Minor-Groove Binding at AT Sites and Intercalation at AU Sites*. In: *Biochemistry* 31.12 (31st Mar. 1992), pp. 3103–3112. ISSN: 0006-2960. DOI: 10.1021/bi00127a010.

[201] Caneva, Rossetti and Savino. *DNA Minor Groove Accessibility in the Nucleosome Core Deduced from Specific Interactions with DAPI*. In: *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression* (1996). DOI: 10.1016/0167-4781(96)00015-2.

[202] Thermo Fisher Scientific. *DAPI Counterstaining Protocols*. 24th Apr. 2006. URL: https://www.thermofisher.com/uk/en/home/references/protocols/cell-and-tissue-analysis/dapi-protocol/basic-dapi-counterstaining-protocols.html (visited on 11/02/2026).

[203] Cecil J. Gomes et al. *Measuring DNA Content in Live Cells by Fluorescence Microscopy*. In: *Cell Division* 13.1 (4th Sept. 2018), p. 6. ISSN: 1747-1028. DOI: 10.1186/s13008-018-0039-z.

[204] Fabio Gasparri, Paolo Cappella and Arturo Galvani. *Multiparametric Cell Cycle Analysis by Automated Microscopy*. In: *Journal of Biomolecular Screening* 11.6 (1st Sept. 2006), pp. 586–598. ISSN: 1087-0571. DOI: 10.1177/1087057106289406.

[205] Ana Katrina Estandarte et al. *The Use of DAPI Fluorescence Lifetime Imaging for Investigating Chromatin Condensation in Human Chromosomes*. In: *Scientific Reports* 6.1 (16th Aug. 2016), p. 31417. ISSN: 2045-2322. DOI: 10.1038/srep31417.

[206] Giancarlo Mascetti, Sandro Carrara and Laura Vergani. *Relationship between Chromatin Compactness and Dye Uptake for in Situ Chromatin Stained with DAPI*. In: *Cytometry* 44.2 (2001), pp. 113–119. ISSN: 1097-0320. DOI: 10.1002/1097-0320(20010601)44:2<113::AID-CYTO1089>3.0.CO;2-A.

[207] Peder Rustøen Braadland and Alfonso Urbanucci. *Chromatin Reprogramming as an Adaptation Mechanism in Advanced Prostate Cancer*. In: *Endocrine-Related Cancer* 26.4 (1st Apr. 2019), R211–R235. ISSN: 1351-0088, 1479-6821. DOI: 10.1530/ERC-18-0579.

[208] Karthik Damodaran et al. *Nuclear Morphometrics and Chromatin Condensation Patterns as Disease Biomarkers Using a Mobile Microscope*. In: *PLoS ONE* 14.7 (17th July 2019), e0218757. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0218757. PMID: 31314779.

[209] Lucía Ramos-Alonso et al. *Mitotic Chromosome Condensation Resets Chromatin to Safeguard Transcriptional Homeostasis during Interphase*. In: *Proceedings of the National Academy of Sciences of the United States of America* 120.4 (19th Jan. 23), e2210593120. ISSN: 0027-8424. DOI: 10.1073/pnas.2210593120. PMID: 36656860.

[210] V. Nunes et al. *Chromosome Condensation Mechanically Primes the Nucleus for Mitosis*. 2nd Sept. 2025. DOI: 10.1101/2025.09.01.673438. URL: https://www.biorxiv.org/content/10.1101/2025.09.01.673438v1 (visited on 05/03/2026). Pre-published.

[211] Eduardo C. Filippi-Chiela et al. *Nuclear Morphometric Analysis (NMA): Screening of Senescence, Apoptosis and Nuclear Irregularities*. In: *PLOS ONE* 7.8 (8th Aug. 2012), e42522. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0042522.

[212] François-Michel Boisvert et al. *The Multifunctional Nucleolus*. In: *Nature Reviews Molecular Cell Biology* 8.7 (July 2007), pp. 574–585. ISSN: 1471-0072, 1471-0080. DOI: 10.1038/nrm2184.

[213] Yun Wah Lam and Laura Trinkle-Mulcahy. *New Insights into Nucleolar Structure and Function*. In: *F1000Prime Reports* 7 (2nd Apr. 2015), p. 48. ISSN: 2051-7599. DOI: 10.12703/P7-48. PMID: 26097721.

[214] Isabella N. Brown, M. Carmen Lafita-Navarro and Maralice Conacci-Sorrell. *Regulation of Nucleolar Activity by MYC*. In: *Cells* 11.3 (7th Feb. 2022), p. 574. ISSN: 2073-4409. DOI: 10.3390/cells11030574. PMID: 35159381.

[215] M.Carmen Lafita-Navarro and Maralice Conacci-Sorrell. *Nucleolar Stress: From Development to Cancer*. In: *Seminars in cell & developmental biology* 136 (28th Feb. 2023), pp. 64–74. ISSN: 1084-9521. DOI: 10.1016/j.semcdb.2022.04.001. PMID: 35410715.

[216] Andrew H. Fischer, Svetoslav Bardarov and Zhong Jiang. *Molecular Aspects of Diagnostic Nucleolar and Nuclear Envelope Changes in Prostate Cancer*. In: *Journal of Cellular Biochemistry* (17th Nov. 2003). DOI: 10.1002/jcb.10735.

[217] Peter H. Gann et al. *Development of a Nuclear Morphometric Signature for Prostate Cancer Risk in Negative Biopsies*. In: *PLOS ONE* 8.7 (26th July 2013), e69457. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0069457.

[218] Xue Zhang et al. *Advances in the Structure and Function of the Nucleolar Protein Fibrillarin*. In: *Frontiers in Cell and Developmental Biology* 12 (13th Nov. 2024), p. 1494631. ISSN: 2296-634X. DOI: 10.3389/fcell.2024.1494631.

[219] R. L. Ochs et al. *Fibrillarin: A New Protein of the Nucleolus Identified by Autoimmune Sera*. In: *Biology of the Cell* 54.2 (1985), pp. 123–133. ISSN: 1768-322X. DOI: 10.1111/j.1768-322X.1985.tb00387.x.

[220] Sabine Snaar et al. *Mutational Analysis of Fibrillarin and Its Mobility in Living Human Cells*. In: *The Journal of Cell Biology* 151.3 (30th Oct. 2000), pp. 653–662. ISSN: 0021-9525. DOI: 10.1083/jcb.151.3.653. PMID: 11062265.

[221] Carson J. Bryant et al. *A High-Throughput Assay for Directly Monitoring Nucleolar rRNA Biogenesis*. In: *Open Biology* 12.1 (26th Jan. 2022), p. 210305. DOI: 10.1098/rsob.210305.

[222] M. Uemura et al. *Overexpression of Ribosomal RNA in Prostate Cancer Is Common but Not Linked to rDNA Promoter Hypomethylation*. In: *Oncogene* 31.10 (Mar. 2012), pp. 1254–1263. ISSN: 1476-5594. DOI: 10.1038/onc.2011.319.

[223] Wenchao Jiang et al. *Rna Buffering Fluorogenic Probe for Nucleolar Morphology Stable Imaging And Nucleolar Stress-Generating Agents Screening*. In: *Advanced Science* 11.15 (2024), p. 2309743. ISSN: 2198-3844. DOI: 10.1002/advs.202309743.

[224] Enkhzaya Batnasan et al. *Platinum-Based Drugs Induce Phenotypic Alterations in Nucleoli and Cajal Bodies in Prostate Cancer Cells*. In: *Cancer Cell International* 24.1 (13th Jan. 2024), p. 29. ISSN: 1475-2867. DOI: 10.1186/s12935-023-03205-0.

[225] Ali Khodadadian et al. *Genomics and Transcriptomics: The Powerful Technologies in Precision Medicine*. In: *International Journal of General Medicine* 13 (17th Sept. 2020), pp. 627–640. ISSN: 1178-7074. DOI: 10.2147/IJGM.S249970. PMID: 32982380.

[226] Ana Ortega-Batista et al. *Single-Cell Sequencing: Genomic and Transcriptomic Approaches in Cancer Cell Biology*. In: *International Journal of Molecular Sciences* 26.5 (Jan. 2025), p. 2074. ISSN: 1422-0067. DOI: 10.3390/ijms26052074.

[227] David P. Nusinow et al. *Quantitative Proteomics of the Cancer Cell Line Encyclopedia*. In: *Cell* 180.2 (23rd Jan. 2020), 387–402.e16. ISSN: 0092-8674, 1097-4172. DOI: 10.1016/j.cell.2019.12.023. PMID: 31978347.

[228] Ahcene Taleb et al. *Raman Microscopy for the Chemometric Analysis of Tumor Cells*. In: *The Journal of Physical Chemistry B* 110.39 (1st Oct. 2006), pp. 19625–19631. ISSN: 1520-6106, 1520-5207. DOI: 10.1021/jp061981q.

[229] Reiner Salzer and Heinz W. Siesler. *Infrared and Raman Spectroscopic Imaging*. John Wiley & Sons, 7th Aug. 2014. 656 pp. ISBN: 978-3-527-67815-0.

[230] Sumanta Kar et al. *Label-Free Discrimination of Tumorigenesis Stages Using in Vitro Prostate Cancer Bone Metastasis Model by Raman Imaging*. In: *Scientific Reports* 12.1 (16th May 2022), p. 8050. ISSN: 2045-2322. DOI: 10.1038/s41598-022-11800-w.

[231] R. A. Rocha et al. *Biomolecular Phenotyping and Heterogeneity Assessment of Mesenchymal Stromal Cells Using Label-Free Raman Spectroscopy*. In: *Scientific Reports* 11 (23rd Feb. 2021), p. 4385. ISSN: 2045-2322. DOI: 10.1038/s41598-021-81991-1. PMID: 33623051.

[232] Chuanshan Xu, Siu Kan Law and Albert Wing Nang Leung. *Comparison of the Differences between Two-Photon Excitation, Upconversion, and Conventional Photodynamic Therapy on Cancers in In Vitro and In Vivo Studies*. In: *Pharmaceuticals* 17.6 (21st May 2024), p. 663. ISSN: 1424-8247. DOI: 10.3390/ph17060663. PMID: 38931331.

[233] Partha Pratim Mondal and Alberto Diaspro. *Fundamentals of Fluorescence Microscopy: Exploring Life with Light*. Dordrecht: Springer Netherlands, 2014. ISBN: 978-94-007-7544-2 978-94-007-7545-9. DOI: 10.1007/978-94-007-7545-9.

[234] Peter J. Verveer, ed. *Advanced Fluorescence Microscopy: Methods and Protocols*. Vol. 1251. Methods in Molecular Biology. New York, NY: Springer New York, 2015. ISBN: 978-1-4939-2079-2 978-1-4939-2080-8. DOI: 10.1007/978-1-4939-2080-8.

[235] T. Schimmang et al. *A Yeast Nucleolar Protein Related to Mammalian Fibrillarin Is Associated with Small Nucleolar RNA and Is Essential for Viability*. In: *The EMBO journal* 8.13 (20th Dec. 1989), pp. 4015–4024. ISSN: 0261-4189. DOI: 10.1002/j.1460-2075.1989.tb08584.x. PMID: 2686980.

[236] Katarzyna A. Solanko et al. *Fluorescent Sterols and Cholesteryl Esters as Probes for Intracellular Cholesterol Transport*. In: *Lipid insights* 8 (Suppl 1 9th June 2016), pp. 95–114. ISSN: 1178-6353. DOI: 10.4137/LPI.S31617. PMID: 27330304.

[237] Thomas A. Cole et al. *Use of Nile Red as a Rapid Measure of Lipid Content in Ciliates*. In: *European Journal of Protistology* 25.4 (29th June 1990), pp. 361–368. ISSN: 0932-4739. DOI: 10.1016/S0932-4739(11)80129-X.

[238] Chenghua Cui, Wei Shu and Peining Li. *Fluorescence In Situ Hybridization: Cell-Based Genetic Diagnostic and Research Applications*. In: *Frontiers in Cell and Developmental Biology* 4 (5th Sept. 2016). ISSN: 2296-634X. DOI: 10.3389/fcell.2016.00089.

[239] Figen Celep et al. *Detection of Chromosomal Aberrations in Prostate Cancer by Fluorescence In Situ Hybridization (FISH)*. In: *European Urology* 44.6 (1st Dec. 2003), pp. 666–671. ISSN: 0302-2838. DOI: 10.1016/S0302-2838(03)00414-7.

[240] Maisa Yoshimoto et al. *Three-Color FISH Analysis of TMPRSS2/ERG Fusions in Prostate Cancer Indicates That Genomic Microdeletion of Chromosome 21 Is Associated with Rearrangement*. In: *Neoplasia (New York, N.Y.)* 8.6 (June 2006), pp. 465–469. ISSN: 1522-8002. DOI: 10.1593/neo.06283. PMID: 16820092.

[241] Ulrich Kubitscheck, ed. *Fluorescence Microscopy: From Principles to Biological Applications*. 1st ed. Wiley, 8th May 2017. ISBN: 978-3-527-33837-5 978-3-527-68773-2. DOI: 10.1002/9783527687732.

[242] Sigrun M. Gustafsdottir et al. *Multiplex Cytological Profiling Assay to Measure Diverse Cellular States*. In: *PLOS ONE* 8.12 (2nd Dec. 2013), e80999. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0080999.

[243] Mark-Anthony Bray et al. *Cell Painting, a High-Content Image-Based Assay for Morphological Profiling Using Multiplexed Fluorescent Dyes*. In: *Nature Protocols* (2016). DOI: 10.1038/nprot.2016.105.

[244] clue.io Connectopedia. *What Are the Cell Painting Data Levels?* 22nd May 2018. URL: https://clue.io/connectopedia/cell_painting_data_levels (visited on 06/03/2026).

[245] Mark-Anthony Bray et al. *A Dataset of Images and Morphological Profiles of 30 000 Small-Molecule Treatments Using the Cell Painting Assay*. In: *GigaScience* 6.12 (1st Dec. 2017), giw014. ISSN: 2047-217X. DOI: 10.1093/gigascience/giw014.

[246] Hemaxi Narotamo et al. *A Machine Learning Approach for Single Cell Interphase Cell Cycle Staging*. In: *Scientific Reports* 11.1 (29th Sept. 2021), p. 19278. ISSN: 2045-2322. DOI: 10.1038/s41598-021-98489-5.

[247] Arkadiusz Gertych, Daniel L. Farkas and Jian Tajbakhsh. *Measuring Topology of Low-Intensity DNA Methylation Sites for High-Throughput Assessment of Epigenetic Drug-Induced Effects in Cancer Cells*. In: *Experimental Cell Research* (2010). DOI: 10.1016/j.yexcr.2010.08.013.

[248] Arkadiusz Gertych et al. *3-D DNA Methylation Phenotypes Correlate with Cytotoxicity Levels in Prostate and Liver Cancer Cell Models*. In: *BMC Pharmacology and Toxicology* 14.1 (11th Feb. 2013), p. 11. ISSN: 2050-6511. DOI: 10.1186/2050-6511-14-11.

[249] Priyanka Rana et al. *Estimation of Three-Dimensional Chromatin Morphology for Nuclear Classification and Characterisation*. In: *Scientific Reports* 11.1 (9th Feb. 2021), p. 3364. ISSN: 2045-2322. DOI: 10.1038/s41598-021-82985-9.

[250] Karl Olofsson et al. *Single Cell Organization and Cell Cycle Characterization of DNA Stained Multicellular Tumor Spheroids*. In: *Scientific Reports* 11.1 (23rd Aug. 2021), p. 17076. ISSN: 2045-2322. DOI: 10.1038/s41598-021-96288-6.

[251] Jin-Shu He et al. *High-Content Imaging Approaches to Quantitate Stress-Induced Changes in Nucleolar Morphology*. In: *ASSAY and Drug Development Technologies* (2018). DOI: 10.1089/adt.2018.861.

[252] Lisa M. Ogawa et al. *Increased Numbers of Nucleoli in a Genome-Wide RNAi Screen Reveal Proteins That Link the Cell Cycle to RNA Polymerase I Transcription*. In: *Molecular Biology of the Cell* 32.9 (19th Apr. 2021). Ed. by Jennifer Lippincott-Schwartz, pp. 956–973. ISSN: 1059-1524, 1939-4586. DOI: 10.1091/mbc.E20-10-0670.

[253] Tamara A. Potapova et al. *Distinct States of Nucleolar Stress Induced by Anti-Cancer Drugs*. In: *eLife* 12 (1st Nov. 2023). DOI: 10.7554/eLife.88799.2.

[254] James M. Jamison et al. *Nucleolar Changes and Fibrillarin Redistribution Following Apatone Treatment of Human Bladder Carcinoma Cells*. In: *Journal of Histochemistry and Cytochemistry* 58.7 (July 2010), pp. 635–651. ISSN: 0022-1554. DOI: 10.1369/jhc.2010.956284. PMID: 20385787.

[255] Emily C. Sutton and Victoria J. DeRose. *Early Nucleolar Responses Differentiate Mechanisms of Cell Death Induced by Oxaliplatin and Cisplatin*. In: *Journal of Biological Chemistry* 296 (1st Jan. 2021), p. 100633. ISSN: 0021-9258. DOI: 10.1016/j.jbc.2021.100633.

[256] Luc Snyers et al. *CX-5461 Causes Nucleolar Compaction, Alteration of Peri- and Intranucleolar Chromatin Arrangement, an Increase in Both Heterochromatin and DNA Damage Response*. In: *Scientific Reports* 12.1 (17th Aug. 2022), p. 13972. ISSN: 2045-2322. DOI: 10.1038/s41598-022-17923-4.

[257] Sille Blangstrup Geisler et al. *Ribosome Biogenesis Is Increased in Hepatocellular Carcinoma and Represents a Potential Therapeutic Target*. In: *NAR Cancer* 8.1 (1st Mar. 2026), zcaf058. ISSN: 2632-8674. DOI: 10.1093/narcan/zcaf058.

[258] Stanisław Węglarczyk. *Kernel Density Estimation and Its Application*. In: *ITM Web of Conferences* 23 (2018). Ed. by W. Zielinski et al., p. 00037. ISSN: 2271-2097. DOI: 10.1051/itmconf/20182300037.

[259] Christophe Biernacki, Gilles Celeux and Gérard Govaert. *Assessing a Mixture Model for Clustering with the Integrated Classification Likelihood*. RR-3521. INRIA, Oct. 1998.

[260] John Frederick William Herschel. *IV. 'Amorphota, No. I.— on a Case of Superficial Colour Presented by a Homogeneous Liquid Internally Colourless*. In: *Philosophical Transactions of the Royal Society of London* 135 (1st Jan. 1845), pp. 143–145. DOI: 10.1098/rstl.1845.0004.

[261] G. G. Stokes. *On the Change of Refrangibility of Light*. In: *Philosophical Transactions of the Royal Society of London* 142 (1852), pp. 463–562. ISSN: 0261-0523. JSTOR: 108550.

[262] Nicole Rusk. *The Fluorescence Microscope*. In: *Nature Cell Biology* 11.1 (Oct. 2009), S8–S9. ISSN: 1476-4679. DOI: 10.1038/ncb1941.

[263] Hellen C. Ishikawa-Ankerhold, Richard Ankerhold and Gregor P. C. Drummen. *Advanced Fluorescence Microscopy Techniques—FRAP, FLIP, FLAP, FRET and FLIM*. In: *Molecules* 17.4 (1st Apr. 2012), pp. 4047–4132. ISSN: 1420-3049. DOI: 10.3390/molecules17044047.

[264] Jeff W. Lichtman and José-Angel Conchello. *Fluorescence Microscopy*. In: *Nature Methods* 2.12 (Dec. 2005), pp. 910–919. ISSN: 1548-7105. DOI: 10.1038/nmeth817.

[265] Bernard Valeur and Mário Nuno Berberan-Santos. *Molecular Fluorescence: Principles and Applications*. 2nd ed. Weinheim: Wiley-VCH, 2012. 1 p. ISBN: 978-3-527-32837-6 978-3-527-65000-2 978-3-527-65003-3 978-1-299-44877-3 978-3-527-32846-8. DOI: 10.1002/9783527650002.

[266] Luis A. Bagatolli. *Fluorescence Spectroscopy: Basic Foundations and Methods*. In: *Analytical Techniques in the Pharmaceutical Sciences*. Ed. by Anette Müllertz, Yvonne Perrie and Thomas Rades. New York, NY: Springer, 2016, pp. 29–59. ISBN: 978-1-4939-4029-5. DOI: 10.1007/978-1-4939-4029-5_2.

[267] Michael Kasha. *Characterization of Electronic Transitions in Complex Molecules*. In: *Discussions of the Faraday Society* 9.0 (1st Jan. 1950), pp. 14–19. ISSN: 0366-9033. DOI: 10.1039/DF9500900014.

[268] Mortimer Abramowitz and Michael W. Davidson. *Molecular Expressions Microscopy Primer: Light and Color - Fluorescence - Overview of Excitation and Emission Fundamentals*. Tuesday, Sep 11, 2018 at 11:22 AM. URL: https://micro.magnet.fsu.edu/primer/lightandcolor/fluoroexcitation.html (visited on 29/08/2025).

[269] Hong Liu et al. *Organic Fluorophores with Large Stokes Shift for Bioimaging and Biosensing*. In: *ChemPhotoChem* 8.5 (2024), e202300277. ISSN: 2367-0932. DOI: 10.1002/cptc.202300277.

[270] Juan Carlos Stockert and Alfonso Blazquez-Castro. *Fluorescence Microscopy in Life Sciences*. BENTHAM SCIENCE PUBLISHERS, 14th Dec. 2017. ISBN: 978-1-68108-518-0. DOI: 10.2174/9781681085180117010 1.

[271] Joseph R. Lakowicz, ed. *Principles of Fluorescence Spectroscopy*. Boston, MA: Springer US, 2006. ISBN: 978-0-387-31278-1 978-0-387-46312-4. DOI: 10.1007/978-0-387-46312-4.

[272] Donna J. Webb and Claire M. Brown. *Epi-Fluorescence Microscopy*. In: *Cell Imaging Techniques: Methods and Protocols*. Ed. by Douglas J. Taatjes and Jürgen Roth. Totowa, NJ: Humana Press, 2013, pp. 29–59. ISBN: 978-1-62703-056-4. DOI: 10.1007/978-1-62703-056-4_2.

[273] Rayleigh. *XXXI. Investigations in Optics, with Special Reference to the Spectroscope*. In: *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 8.49 (1st Oct. 1879), pp. 261–274. ISSN: 1941-5982. DOI: 10.1080/14786447908639684.

[274] Rayleigh. *XV. On the Theory of Optical Images, with Special Reference to the Microscope*. In: *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 42.255 (1st Aug. 1896), pp. 167–195. ISSN: 1941-5982. DOI: 10.1080/14786449608620902.

[275] Nataliya Panchuk-Voloshina et al. *Alexa Dyes, a Series of New Fluorescent Dyes That Yield Exceptionally Bright, Photostable Conjugates*. In: *Journal of Histochemistry & Cytochemistry* 47.9 (1st Sept. 1999), pp. 1179–1188. ISSN: 0022-1554. DOI: 10.1177/002215549904700910.

[276] Thermo Fisher Scientific. *Alexa Fluor 488 Dye - UK*. URL: https://www.thermofisher.com/uk/en/home/life-science/cell-analysis/fluorophores/alexa-fluor-488.html (visited on 09/03/2026).

[277] S H Lang, M Stower and N J Maitland. *In Vitro Modelling of Epithelial and Stromal Interactions in Non-Malignant and Malignant Prostates*. In: *British Journal of Cancer* 82.4 (Feb. 2000), pp. 990–997. ISSN: 0007-0920, 1532-1827. DOI: 10.1054/bjoc.1999.1029.

[278] Fiona M. Frame et al. *Tumor Heterogeneity and Therapy Resistance - Implications for Future Treatments of Prostate Cancer*. In: *Journal of Cancer Metastasis and Treatment* 3.0 (6th Dec. 2017), pp. 302–314. ISSN: ISSN 2454-2857 (Online)
ISSN 2394-4722 (Print). DOI: 10.20517/2394-4722.2017.34.

[279] Engin Ulukaya et al. *Differential Cytotoxic Activity of a Novel Palladium-Based Compound on Prostate Cell Lines, Primary Prostate Epithelial Cells and Prostate Stem Cells*. In: *PLoS ONE* 8.5 (10th May 2013). Ed. by Kamyar Afarinkia, e64278. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0064278.

[280] Amanda R. Noble et al. *Phospholipases D1 and D2 Regulate Cell Cycling in Primary Prostate Cancer Cells and Are Differentially Associated with the Nuclear Matrix*. 20th May 2021. DOI: 10.1101/2021.05.19.444795. URL: https://www.biorxiv.org/content/10.1101/2021.05.19.444795v1 (visited on 22/11/2024). Pre-published.

[281] Bregje van Oorschot et al. *Decay of Y-H2AX Foci Correlates with Potentially Lethal Damage Repair in Prostate Cancer Cells*. In: *Oncology Reports* 29.6 (1st June 2013), pp. 2175–2180. ISSN: 1021-335X. DOI: 10.3892/or.2013.2364.

[282] Connor Hill et al. *NAB2-STAT6 Drives an EGR1-dependent Neuroendocrine Program in Solitary Fibrous Tumors*. In: *eLife* 13 (28th Aug. 2025). Ed. by Arduino A Mangoni and Eduardo L Franco, RP98072. ISSN: 2050-084X. DOI: 10.7554/eLife.98072.

[283] Jiali Cai et al. *USP37 Promotes Lung Cancer Cell Migration by Stabilizing Snail Protein via Deubiquitination*. In: *Frontiers in Genetics* 10 (10th Jan. 2020). ISSN: 1664-8021. DOI: 10.3389/fgene.2019.01324.

[284] Bin Lan et al. *CRISPR-Cas9 Screen Identifies DYRK1A as a Target for Radiotherapy Sensitization in Pancreatic Cancer*. In: *Cancers* 14.2 (Jan. 2022), p. 326. ISSN: 2072-6694. DOI: 10.3390/cancers14020326.

[285] David Whaley et al. *Cryopreservation: An Overview of Principles and Cell-Specific Considerations*. In: *Cell Transplantation* 30 (24th Mar. 2021), p. 0963689721999617. ISSN: 0963-6897. DOI: 10.1177/0963689721999617. PMID: 33757335.

[286] Laura Llobet et al. *Side Effects of Culture Media Antibiotics on Cell Differentiation*. In: *Tissue Engineering Part C: Methods* 21.11 (1st Nov. 2015), pp. 1143–1147. ISSN: 1937-3384. DOI: 10.1089/ten.tec.2015.0062.

[287] Leica Microsystems. *DM IL LED Technical Data*. 28th Jan. 2026.

[288] European Medicines Agency. *VICH GL34: Biologicals: Testing for the Detection of Mycoplasma Contamination*. Scientific Guideline EMA/CVMP/VICH/463/2002. European Medicines Agency, 13th Mar. 2013.

[289] Warren Strober. *Trypan Blue Exclusion Test of Cell Viability*. In: *Current Protocols in Immunology* 111.1 (2015), A3.B.1–A3.B.3. ISSN: 1934-368X. DOI: 10.1002/0471142735.ima03bs111.

[290] Thermo Fisher Scientific. *Useful Numbers for Cell Culture*. URL: https://www.thermofisher.com/uk/en/home/references/gibco-cell-culture-basics/cell-culture-protocols/cell-culture-useful-numbers.html (visited on 05/09/2025).

[291] Mengfei Chen et al. *Serum Starvation Induced Cell Cycle Synchronization Facilitates Human Somatic Cells Reprogramming*. In: *PLoS ONE* 7.4 (18th Apr. 2012), e28203. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0028203. PMID: 22529890.

[292] Jinjin Tong et al. *Serum Starvation and Thymidine Double Blocking Achieved Efficient Cell Cycle Synchronization and Altered the Expression of P27, P53, Bcl-2 in Canine Breast Cancer Cells*. In: *Research in Veterinary Science* 105 (1st Apr. 2016), pp. 10–14. ISSN: 0034-5288. DOI: 10.1016/j.rvsc.2016.01.008.

[293] Yu-Lin Chen, Syon Reddy and Aussie Suzuki. *Reversible and Effective Cell Cycle Synchronization Method for Studying Stage-Specific Processes*. In: *Life Science Alliance* 8.5 (4th Mar. 2025), e202403000. ISSN: 2575-1077. DOI: 10.26508/lsa.202403000. PMID: 40037894.

[294] Seong-Oh Kim et al. *Mechanical Properties of Paraformaldehyde-Treated Individual Cells Investigated by Atomic Force Microscopy and Scanning Ion Conductance Microscopy*. In: *Nano Convergence* 4 (20th Mar. 2017), p. 5. ISSN: 2196-5404. DOI: 10.1186/s40580-017-0099-9. PMID: 28386525.

[295] Melissa A. Melan and Greenfield Sluder. *Redistribution and Differential Extraction of Soluble Proteins in Permeabilized Cultured Cells Implications for Immunofluorescence Microscopy*. In: *Journal of Cell Science* 101.4 (1st Apr. 1992), pp. 731–743. ISSN: 0021-9533, 1477-9137. DOI: 10.1242/jcs.101.4.731.

[296] M. J. Hannah, U. Weiss and W. B. Huttner. *Differential Extraction of Proteins from Paraformaldehyde-Fixed Cells: Lessons from Synaptophysin and Other Membrane Proteins*. In: *Methods* 16.2 (1st Oct. 1998), pp. 170–181. ISSN: 1046-2023. DOI: 10.1006/meth.1998.0664.

[297] Kerry L. Inder et al. *Cavin-1/PTRF Alters Prostate Cancer Cell-derived Extracellular Vesicle Content and Internalization to Attenuate Extracellular Vesicle-mediated Osteoclastogenesis and Osteoblast Proliferation*. In: *Journal of Extracellular Vesicles* (2014). DOI: 10.3402/jev.v3.23784.

[298] K. A. Landers et al. *Identification of Claudin-4 as a Marker Highly Overexpressed in Both Primary and Metastatic Prostate Cancer*. In: *British Journal of Cancer* (2008). DOI: 10.1038/sj.bjc.6604486.

[299] Carl Zeiss Microscopy GmbH. *ZEISS ZEN 2 (Blue Edition) Microscopy Software (RRID:SCR_013672)*. Version 3.12.107.03000. 2025.

[300] Carl Zeiss Microscopy GmbH. *Carl Zeiss Limited - Inverted Microscopes - Axio Observer - Axio Observer 7*. URL: https://www.micro-shop.zeiss.com/en/uk/system/axio+observer-axio+observer+7-inverted+microscopes/10303/ (visited on 09/03/2026).

[301] Adam Zweifach. *Determining How Many Cells to Average for Statistical Testing of Microscopy Experiments*. In: *The Journal of Cell Biology* 223.8 (26th June 2024), e202401074. ISSN: 0021-9525. DOI: 10.1083/jcb.202401074. PMID: 38922298.

[302] Aabid Shariff et al. *Automated Image Analysis for High-Content Screening and Analysis*. In: *SLAS Discovery*. High-Content Screening, Imaging, & Data Analysis 15.7 (1st Aug. 2010), pp. 726–734. ISSN: 2472-5552. DOI: 10.1177/1087057110370894.

[303] David R. Stirling et al. *CellProfiler 4: Improvements in Speed, Utility and Usability*. In: *BMC Bioinformatics* 22.1 (10th Sept. 2021), p. 433. ISSN: 1471-2105. DOI: 10.1186/s12859-021-04344-9.

[304] Anne E. Carpenter et al. *CellProfiler: Image Analysis Software for Identifying and Quantifying Cell Phenotypes*. In: *BioMed Central Ltd* (Oct. 2006). ISSN: 1465-6906.

[305] Erick Moen et al. *Deep Learning for Cellular Image Analysis*. In: *Nature Methods* 16.12 (Dec. 2019), pp. 1233–1246. ISSN: 1548-7105. DOI: 10.1038/s41592-019-0403-1.

[306] Olaf Ronneberger, Philipp Fischer and Thomas Brox. *U-Net: Convolutional Networks for Biomedical Image Segmentation*. Version 1. 2015. DOI: 10.48550/ARXIV.1505.04597. URL: https://arxiv.org/abs/1505.04597 (visited on 09/03/2026). Pre-published.

[307] Carsen Stringer et al. *Cellpose: A Generalist Algorithm for Cellular Segmentation*. In: *Nature Methods* 18.1 (Jan. 2021), pp. 100–106. ISSN: 1548-7105. DOI: 10.1038/s41592-020-01018-x.

[308] Uwe Schmidt et al. *Cell Detection with Star-Convex Polygons*. In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*. Ed. by Alejandro F. Frangi et al. Cham: Springer International Publishing, 2018, pp. 265–273. ISBN: 978-3-030-00934-2. DOI: 10.1007/978-3-030-00934-2_30.

[309] Marius Pachitariu and Carsen Stringer. *Cellpose 2.0: How to Train Your Own Model*. In: *Nature Methods* 19.12 (Dec. 2022), pp. 1634–1641. ISSN: 1548-7105. DOI: 10.1038/s41592-022-01663-4.

[310] Csaba Molnar et al. *Accurate Morphology Preserving Segmentation of Overlapping Cells Based on Active Contours*. In: *Scientific Reports* 6.1 (26th Aug. 2016), p. 32412. ISSN: 2045-2322. DOI: 10.1038/srep32412.

[311] Guido Van Rossum. *Python 3.12.10*. Version 3.12.10. The Python Language Foundation, 4th Aug. 25.

[312] K. Stockman. *Fluorescence Microscopy Study of Prostate-Cancer Cells to Quantify Intracellular Molecular-Heterogeneity and Spatial Distribution of DNA and as a Complementary Method to Raman Spectroscopy*. Master's Theses. York: University of York, 5th Aug. 24.

[313] Johannes Schindelin et al. *Fiji: An Open-Source Platform for Biological-Image Analysis*. In: *Nature Methods* 9.7 (July 2012), pp. 676–682. ISSN: 1548-7105. DOI: 10.1038/nmeth.2019.

[314] Gabriel J. Selzer et al. *SciJava Ops: An Improved Algorithms Framework for Fiji and Beyond*. In: *Frontiers in Bioinformatics* 4 (27th Sept. 2024). ISSN: 2673-7647. DOI: 10.3389/fbinf.2024.1435733.

[315] Charles R. Harris et al. *Array Programming with NumPy*. In: *Nature* 585.7825 (Sept. 2020), pp. 357–362. ISSN: 1476-4687. DOI: 10.1038/s41586-020-2649-2.

[316] Pauli Virtanen et al. *SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python*. In: *Nature Methods* 17.3 (Mar. 2020), pp. 261–272. ISSN: 1548-7105. DOI: 10.1038/s41592-019-0686-2.

[317] Stéfan van der Walt et al. *Scikit-Image: Image Processing in Python*. In: *PeerJ* 2 (19th June 2014), e453. ISSN: 2167-8359. DOI: 10.7717/peerj.453.

[318] G Bradski. *The OpenCV Library*. In: *Dr Dobb's Journal of Software Tools* (2000).

[319] Christoph Gohlke. *Cgohlke/Tifffile: V2022.5.4*. Zenodo, 4th July 2022. DOI: 10.5281/zenodo.6795861.

[320] John D. Hunter. *Matplotlib: A 2D Graphics Environment*. In: *Computing in Science & Engineering* 9.3 (May 2007), pp. 90–95. ISSN: 1558-366X. DOI: 10.1109/MCSE.2007.55.

[321] S. R. Sternberg. *Biomedical Image Processing*. In: *Computer* 16.01 (1st Jan. 1983), pp. 22–34. ISSN: 0018-9162. DOI: 10.1109/MC.1983.1654163.

[322] R.A. Haddad and A.N. Akansu. *A Class of Fast Gaussian Binomial Filters for Speech and Image Processing*. In: *IEEE Transactions on Signal Processing* 39.3 (Mar. 1991), pp. 723–727. ISSN: 1941-0476. DOI: 10.1109/78.80892.

[323] Nobuyuki Otsu. *A Threshold Selection Method from Gray-Level Histograms*. In: *IEEE Transactions on Systems, Man, and Cybernetics* 9.1 (Jan. 1979), pp. 62–66. ISSN: 2168-2909. DOI: 10.1109/TSMC.1979.4310076.

[324] Petros Maragos. *PDEs for Morphological Scale Spaces and Eikonal Applications*. In: *Handbook of Image and Video Processing* (2005). DOI: 10.1016/B978-012119792-6/50098-X.

[325] Anton Kornilov, Ilia Safonov and Ivan Yakimchuk. *A Review of Watershed Implementations for Segmentation of Volumetric Images*. In: *Journal of Imaging* 8.5 (26th Apr. 2022), p. 127. ISSN: 2313-433X. DOI: 10.3390/jimaging8050127. PMID: 35621890.

[326] Affan Adly. *Affanadly/Ellipse-Fit-Uncertainty*. 25th May 2022.

[327] Eric O. Leibgot. *Uncertainties: A Python Package for Calculations with Uncertainties*. Version http://pythonhosted.org/uncertainties/. 21st Apr. 25.

[328] Juan C. Caicedo et al. *Evaluation of Deep Learning Strategies for Nucleus Segmentation in Fluorescence Images*. In: *Cytometry Part A* 95.9 (2019), pp. 952–965. ISSN: 1552-4930. DOI: 10.1002/cyto.a.23863.

[329] Ying Chun Li et al. *Elevated Levels of Cholesterol-Rich Lipid Rafts in Cancer Cells Are Correlated with Apoptosis Sensitivity Induced by Cholesterol-Depleting Agents*. In: *The American Journal of Pathology* 168.4 (Apr. 2006), pp. 1107–1118. ISSN: 0002-9440. DOI: 10.2353/ajpath.2006.050959. PMID: 16565487.

[330] Ariana Sandoval et al. *Altered Expression and Localization of Nuclear Envelope Proteins in a Prostate Cancer Cell System*. In: *Molecular Biology Reports* 51.1 (8th Aug. 2024), p. 898. ISSN: 1573-4978. DOI: 10.1007/s11033-024-09836-4.

[331] Lisa M. Butler et al. *Lipidomic Profiling of Clinical Prostate Cancer Reveals Targetable Alterations in Membrane Lipid Composition*. In: *Cancer Research* 81.19 (1st Oct. 2021), pp. 4981–4993. ISSN: 0008-5472. DOI: 10.1158/0008-5472.CAN-20-3863.

[332] Gary W. Oehlert. *A Note on the Delta Method*. In: *The American Statistician* 46.1 (1st Feb. 1992), pp. 27–29. ISSN: 0003-1305. DOI: 10.1080/00031305.1992.10475842.

[333] A. W. van der Vaart. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge: Cambridge University Press, 1998. ISBN: 978-0-521-78450-4. DOI: 10.1017/CBO9780511802256.

[334] A. C. Davison and D. V. Hinkley. *Bootstrap Methods and Their Application*. 1st ed. Cambridge University Press, 28th Oct. 1997. ISBN: 978-0-521-57391-7 978-0-521-57471-6 978-0-511-80284-3. DOI: 10.1017/CBO9780511802843.

[335] B. Efron. *Bootstrap Methods: Another Look at the Jackknife*. In: *The Annals of Statistics* (1979). DOI: 10.1214/aos/1176344552.

[336] Dimitris N. Politis, Joseph P. Romano and Michael Wolf. *On the Asymptotic Theory of Subsampling*. In: *DES - Working Papers. Statistics and Econometrics. WS* 6334 (June 1999).

[337] David W. Scott. *Multivariate Density Estimation: Theory, Practice, and Visualization*. Second edition. Hoboken, New Jersey: John Wiley & Sons, Inc, 2015. 350 pp. ISBN: 978-0-471-69755-8.

[338] Catherine Forbes et al. *Statistical Distributions*. 1st ed. Wiley, 29th Nov. 2010. ISBN: 978-0-470-39063-4 978-0-470-62724-2. DOI: 10.1002/9780470627242.

[339] Norman Lloyd Johnson, Adrienne W. Kemp and Samuel Kotz. *Univariate Discrete Distributions*. 3rd ed. Hoboken, N.J: Wiley, 2005. 646 pp. ISBN: 978-0-471-27246-5.

[340] Adelchi Azzalini and Antonella Capitanio. *The Skew-Normal and Related Families*. 1st ed. Cambridge University Press, 19th Dec. 2013. ISBN: 978-1-107-02927-9 978-1-139-24889-1. DOI: 10.1017/CBO9781139248891.

[341] Abdisalam Muse, Samuel Mwalili and Oscar Ngesa. *On the Log-Logistic Distribution and Its Generalizations: A Survey*. In: *International Journal of Statistics and Probability* 10.3 (20th Apr. 2021), p93. ISSN: 1927-7032. DOI: 10.5539/ijsp.v10n3p93.

[342] A. P. Dempster, N. M. Laird and D. B. Rubin. *Maximum Likelihood from Incomplete Data Via the EM Algorithm*. In: *Journal of the Royal Statistical Society: Series B (Methodological)* 39.1 (1977), pp. 1–22. ISSN: 2517-6161. DOI: 10.1111/j.2517-6161.1977.tb01600.x.

[343] Geoffrey J. McLachlan, Sharon X. Lee and Suren I. Rathnayake. *Finite Mixture Models*. In: *Annual Review of Statistics and Its Application* 6 (Volume 6, 2019 7th Mar. 2019), pp. 355–378. ISSN: 2326-8298, 2326-831X. DOI: 10.1146/annurev-statistics-031017-100325.

[344] Joseph E. Cavanaugh and Andrew A. Neath. *The Akaike Information Criterion: Background, Derivation, Properties, Application, Interpretation, and Refinements*. In: *WIREs Computational Statistics* 11.3 (2019), e1460. ISSN: 1939-0068. DOI: 10.1002/wics.1460.

[345] Marcus Duncan Cameron. *Biomolecular Stratification of Standard Prostate Cell Lines and Primary Prostate Cell Cultures Using Confocal Raman Spectroscopy*. University of York, Dec. 2021.

[346] Dong Kyu Lee, Junyong In and Sangseok Lee. *Standard Deviation and Standard Error of the Mean*. In: *Korean Journal of Anesthesiology* 68.3 (June 2015), pp. 220–223. ISSN: 2005-6419. DOI: 10.4097/kjae.2015.68.3.220. PMID: 26045923.

[347] George Casella and Roger L. Berger. *Statistical Inference*. 2. ed. Pacific Grove, Calif: Duxbury, 2002. 660 pp. ISBN: 978-0-534-24312-8.

[348] Abraham de Moivre. *A Method of approximating the Sum of the Terms of the Binomial a+b expanded into a Series, from whence are deduced some practical Rules to estimate the Degree of Assent which is to be given to Experiments*. In: *The Doctrine of Chances: A Method of Calculating the Probability of Events in Play*. Second Edition. London: H. Woodfall, 1738, pp. 235–243.

[349] Abraham de Moivre. *The Doctrine of Chances: A Method of Calculating the Probability of Events in Play*. Second Edition. H. Woodfall, 1738. 235-243.

[350] Sang Gyu Kwak and Jong Hae Kim. *Central Limit Theorem: The Cornerstone of Modern Statistics*. In: *Korean Journal of Anesthesiology* 70.2 (Apr. 2017), pp. 144–156. ISSN: 2005-6419. DOI: 10.4097/kjae.2017.70.2.144. PMID: 28367284.

[351] Philip R. Bevington and D. Keith Robinson. *Data Reduction and Error Analysis for the Physical Sciences*. 3. ed., [Nachdr.] Boston, Mass.: McGraw-Hill, 2003. 320 pp. ISBN: 978-0-07-247227-1 978-93-392-2120-1.

[352] William H. Press, ed. *Numerical Recipes: The Art of Scientific Computing*. 3. ed. Cambridge: Cambridge University Press, 2007. 1235 pp. ISBN: 978-0-521-88068-8.

[353] J Mandel and L F Nanni. *Measurement Evaluation*. NBS SP 700-2. Gaithersburg, MD: National Bureau of Standards, 1986, NBS SP 700–2. DOI: 10.6028/NBS.SP.700-2.

[354] Xiao Xiao et al. *On the Use of Log-Transformation vs. Nonlinear Regression for Analyzing Biological Power Laws*. In: *Ecology* 92.10 (2011), pp. 1887–1894. ISSN: 1939-9170. DOI: 10.1890/11-0538.1.

[355] Aaron Clauset, Cosma Rohilla Shalizi and M. E. J. Newman. *Power-Law Distributions in Empirical Data*. In: *SIAM Review* 51.4 (4th Nov. 2009), pp. 661–703. ISSN: 0036-1445, 1095-7200. DOI: 10.1137/070710111.

[356] Rene Andrae, Tim Schulze-Hartung and Peter Melchior. *Dos and Don'ts of Reduced Chi-Squared*. 16th Dec. 2010. DOI: 10.48550/arXiv.1012.3754. arXiv: 1012.3754 [astro-ph]. URL: http://arxiv.org/abs/1012.3754 (visited on 29/01/2026). Pre-published.

[357] Andrej-Nikolai Spiess and Natalie Neumeyer. *An Evaluation of R2 as an Inadequate Measure for Nonlinear Models in Pharmacological and Biochemical Research: A Monte Carlo Approach*. In: *BMC Pharmacology* 10 (7th June 2010), p. 6. ISSN: 1471-2210. DOI: 10.1186/1471-2210-10-6. PMID: 20529254.

[358] Laura Martin et al. *A Protocol to Quantify Chromatin Compaction with Confocal and Super-Resolution Microscopy in Cultured Cells*. In: *STAR Protocols* 2.4 (17th Dec. 2021), p. 100865. ISSN: 2666-1667. DOI: 10.1016/j.xpro.2021.100865.

[359] Jose D. Debes et al. *P300 Modulates Nuclear Morphology in Prostate Cancer*. In: *Cancer Research* 65.3 (10th Feb. 2005), pp. 708–712. ISSN: 0008-5472. DOI: 10.1158/0008-5472.708.65.3.

[360] Farhat Rashid and Anwar Ul Haque. *Frequencies of Different Nuclear Morphological Features in Prostate Adenocarcinoma*. In: *Annals of Diagnostic Pathology* 15.6 (Dec. 2011), pp. 414–421. ISSN: 10929134. DOI: 10.1016/j.anndiagpath.2011.06.002.

[361] Y. Pan et al. *Characterization of Chromosomal Abnormalities in Prostate Cancer Cell Lines by Spectral Karyotyping*. In: *Cytogenetic and Genome Research* (1999). DOI: 10.1159/000015432.

[362] Konrad H. Stopsack et al. *Aneuploidy Drives Lethal Progression in Prostate Cancer*. In: *Proceedings of the National Academy of Sciences of the United States of America* 116.23 (4th June 2019), pp. 11390–11395. ISSN: 0027-8424. DOI: 10.1073/pnas.1902645116. PMID: 31085648.

[363] Beth AA Weaver and Don W Cleveland. *Does Aneuploidy Cause Cancer?* In: *Current Opinion in Cell Biology*. Cell Division, Growth and Death / Cell Differentiation 18.6 (1st Dec. 2006), pp. 658–667. ISSN: 0955-0674. DOI: 10.1016/j.ceb.2006.10.002.

[364] Haitong Su et al. *Identification of Novel Markers That Demarcate the Nucleolus during Severe Stress and Chemotherapeutic Treatment*. In: *PLoS ONE* 8.11 (6th Nov. 2013), e80237. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0080237. PMID: 24223222.

[365] Jennifer Ferguson. *Characterisation and Investigation of the Antimicrobial Properties of Graphene Using Raman Spectroscopy*. University of York, Jan. 2020.

[366] Robert W. Veltri and Christhunesa S. Christudass. *Nuclear Morphometry, Epigenetic Changes, and Clinical Relevance in Prostate Cancer*. In: *Cancer Biology and the Nuclear Envelope* 773 (25th Nov. 2013), pp. 77–99. ISSN: 0065-2598. DOI: 10.1007/978-1-4899-8032-8_4. PMID: 24563344.

[367] Michelle S. Liberio et al. *Differential Effects of Tissue Culture Coating Substrates on Prostate Cancer Cell Adherence, Morphology and Behavior*. In: *PLOS ONE* 9.11 (6th Nov. 2014), e112122. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0112122.

[368] Lien Spans et al. *Comparative Genomic and Transcriptomic Analyses of LNCaP and C4-2B Prostate Cancer Cell Lines*. In: *PLOS ONE* 9.2 (28th Feb. 2014), e90002. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0090002.

[369] Malte Krönig et al. *Cell Type Specific Gene Expression Analysis of Prostate Needle Biopsies Resolves Tumor Tissue Heterogeneity*. In: *Oncotarget* 6.2 (1st Dec. 2014), pp. 1302–1314. ISSN: 1949-2553. DOI: 10.18632/oncotarget.2744. PMID: 25514598.

[370] Qiuhui Li et al. *Linking Prostate Cancer Cell AR Heterogeneity to Distinct Castration and Enzalutamide Responses*. In: *Nature Communications* 9.1 (6th Sept. 2018), p. 3600. ISSN: 2041-1723. DOI: 10.1038/s41467-018-06067-7.

[371] Joel S. Brown et al. *Updating the Definition of Cancer*. In: *Molecular Cancer Research* 21.11 (1st Nov. 2023), pp. 1142–1147. ISSN: 1541-7786. DOI: 10.1158/1541-7786.MCR-23-0411. PMID: 37409952.

[372] Douglas Hanahan and Robert A. Weinberg. *Hallmarks of Cancer: The Next Generation*. In: *Cell* 144.5 (4th Mar. 2011), pp. 646–674. ISSN: 0092-8674, 1097-4172. DOI: 10.1016/j.cell.2011.02.013. PMID: 21376230.

[373] Laurent Sansregret and Charles Swanton. *The Role of Aneuploidy in Cancer Evolution*. In: *Cold Spring Harbor Perspectives in Medicine* 7.1 (Jan. 2017), a028373. ISSN: 2157-1422. DOI: 10.1101/cshperspect.a028373. PMID: 28049655.

[374] Devon A. Lukow et al. *Chromosomal Instability Accelerates the Evolution of Resistance to Anti-Cancer Therapies*. In: *Developmental Cell* 56.17 (13th Sept. 2021), 2427–2439.e4. ISSN: 1534-5807. DOI: 10.1016/j.devcel.2021.07.009. PMID: 34352222.

[375] Arnab Bose et al. *Multiomic Assessments of LNCaP and Derived Cell Strains Reveal Determinants of Prostate Cancer Pathobiology*. In: *The Journal of Clinical Investigation* 135.22 (16th Sept. 2025), e194727. ISSN: 0021-9738. DOI: 10.1172/JCI194727. PMID: 40956611.

[376] Zbigniew Darzynkiewicz, H. Dorota Halicka and Hong Zhao. *Analysis of Cellular DNA Content by Flow and Laser Scanning Cytometry*. In: *Polyploidization and Cancer*. Ed. by Randy Y. C. Poon. New York, NY: Springer, 2010, pp. 137–147. ISBN: 978-1-4419-6199-0. DOI: 10.1007/978-1-4419-6199-0_9.

[377] Michelle K. Zeman and Karlene A. Cimprich. *Causes and Consequences of Replication Stress*. In: *Nature Cell Biology* 16.1 (Jan. 2014), pp. 2–9. ISSN: 1476-4679. DOI: 10.1038/ncb2897.

[378] Hélène Gaillard, Tatiana García-Muse and Andrés Aguilera. *Replication Stress and Cancer*. In: *Nature Reviews Cancer* 15.5 (May 2015), pp. 276–289. ISSN: 1474-1768. DOI: 10.1038/nrc3916.

[379] Tamara A. Potapova, Jin Zhu and Rong Li. *Aneuploidy and Chromosomal Instability: A Vicious Cycle Driving Cellular Evolution and Cancer Genome Chaos*. In: *Cancer and Metastasis Reviews* 32.3 (1st Dec. 2013), pp. 377–389. ISSN: 1573-7233. DOI: 10.1007/s10555-013-9436-6.

[380] Zeina S. Khan, Julianna M. Santos and Fazle Hussain. *Aggressive Prostate Cancer Cell Nuclei Have Reduced Stiffness*. In: *Biomicrofluidics* 12.1 (2nd Jan. 2018), p. 014102. ISSN: 1932-1058. DOI: 10.1063/1.5019728. PMID: 29333204.

[381] Caroline Uhler and G. V. Shivashankar. *Nuclear Mechanopathology and Cancer Diagnosis*. In: *Trends in Cancer* 4.4 (1st Apr. 2018), pp. 320–331. ISSN: 2405-8033. DOI: 10.1016/j.trecan.2018.02.009. PMID: 29606315.

[382] Mariana Reis-Sobreiro et al. *Emerin Deregulation Links Nuclear Shape Instability to Metastatic Potential*. In: *Cancer Research* 78.21 (1st Nov. 2018), pp. 6086–6097. ISSN: 0008-5472. DOI: 10.1158/0008-5472.CAN-18-0608.

[383] Eric T. Miller et al. *Chromosomal Instability in Untreated Primary Prostate Cancer as an Indicator of Metastatic Potential*. In: *BMC Cancer* 20 (7th May 2020), p. 398. ISSN: 1471-2407. DOI: 10.1186/s12885-020-06817-1. PMID: 32380981.

[384] Marta Popęda et al. *Emerin Mislocalization during Chromatin Bridge Resolution Can Drive Prostate Cancer Cell Invasiveness in a Collagen-Rich Microenvironment*. In: *Experimental & Molecular Medicine* 56.9 (Sept. 2024), pp. 2016–2032. ISSN: 2092-6413. DOI: 10.1038/s12276-024-01308-w.

[385] Uri Ben-David et al. *Genetic and Transcriptional Evolution Alters Cancer Cell Line Drug Response*. In: *Nature* 560.7718 (Aug. 2018), pp. 325–330. ISSN: 0028-0836. DOI: 10.1038/s41586-018-0409-3. PMID: 30089904.

[386] Peter Lapan et al. *Image-Based Assessment of Growth and Signaling Changes in Cancer Cells Mediated by Direct Cell-Cell Contact*. In: *PLoS ONE* 4.8 (31st Aug. 2009). Ed. by Andreas Bergmann, e6822. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0006822.

[387] Elizabeth R. Smith, Callinice D. Capo-chichi and Xiang-Xi Xu. *Defective Nuclear Lamina in Aneuploidy and Carcinogenesis*. In: *Frontiers in Oncology* 8 (20th Nov. 2018). ISSN: 2234-943X. DOI: 10.3389/fonc.2018.00529.

[388] Qing Hu et al. *Non-Homologous End Joining Shapes the Genomic Rearrangement Landscape of Chromothripsis from Mitotic Errors*. In: *Nature Communications* 15.1 (4th July 2024), p. 5611. ISSN: 2041-1723. DOI: 10.1038/s41467-024-49985-5.

[389] Rongbin Ge, Zongwei Wang and Liang Cheng. *Tumor Microenvironment Heterogeneity an Important Mediator of Prostate Cancer Progression and Therapeutic Resistance*. In: *NPJ Precision Oncology* 6 (4th May 2022), p. 31. ISSN: 2397-768X. DOI: 10.1038/s41698-022-00272-w. PMID: 35508696.

[390] Dean G. Tang. *Understanding and Targeting Prostate Cancer Cell Heterogeneity and Plasticity*. In: *Seminars in Cancer Biology*. Cancer Cell Heterogeneity and Plasticity: From Molecular Understanding to Therapeutic Targeting 82 (1st July 2022), pp. 68–93. ISSN: 1044-579X. DOI: 10.1016/j.semcancer.2021.11.001.

[391] Doğancan Özturan, Tunç Morova and Nathan A. Lack. *Androgen Receptor-Mediated Transcription in Prostate Cancer*. In: *Cells* 11.5 (5th Mar. 2022), p. 898. ISSN: 2073-4409. DOI: 10.3390/cells11050898. PMID: 35269520.

[392] Fahim Ahmad, Murali Krishna Cherukuri and Peter L. Choyke. *Metabolic Reprogramming in Prostate Cancer*. In: *British Journal of Cancer* 125.9 (Oct. 2021), pp. 1185–1196. ISSN: 1532-1827. DOI: 10.1038/s41416-021-01435-5.

[393] Ivana Grbesa et al. *Reshaping of the Androgen-Driven Chromatin Landscape in Normal Prostate Cells by Early Cancer Drivers and Effect on Therapeutic Sensitivity*. In: *Cell reports* 36.10 (7th Sept. 2021), p. 109625. ISSN: 2211-1247. DOI: 10.1016/j.celrep.2021.109625. PMID: 34496233.

[394] V. K. Vaishnavi Vedam, Karen Boaz and Srikant Natarajan. *Prognostic Efficacy of Nuclear Morphometry at Invasive Front of Oral Squamous Cell Carcinoma: An Image Analysis Microscopic Study*. In: *Analytical Cellular Pathology* 2014.1 (2014), p. 247853. ISSN: 2210-7185. DOI: 10.1155/2014/247853.

[395] Kamil Malshy et al. *Association between Nuclear Morphometry Parameters and Gleason Grade in Patients with Prostatic Cancer*. In: *Diagnostics* 12.6 (31st May 2022), p. 1356. ISSN: 2075-4418. DOI: 10.3390/diagnostics12061356. PMID: 35741165.

[396] Abdelbaset Buhmeida. *Quantitative Pathology: Historical Background, Clinical Research and Application of Nuclear Morphometry and DNA Image Cytometry*. In: *The Libyan Journal of Medicine* 1.2 (20th Sept. 2006), pp. 126–139. ISSN: 1993-2820. DOI: 10.4176/060911. PMID: 21526011.

[397] Helena Cantwell and Paul Nurse. *Unravelling Nuclear Size Control*. In: *Current Genetics* 65.6 (1st Dec. 2019), pp. 1281–1285. ISSN: 1432-0983. DOI: 10.1007/s00294-019-00999-3.

[398] Michael J. Moore, Joseph A. Sebastian and Michael C. Kolios. *Determination of Cell Nucleus-to-Cytoplasmic Ratio Using Imaging Flow Cytometry and a Combined Ultrasound and Photoacoustic Technique: A Comparison Study*. In: *Journal of Biomedical Optics* 24.10 (Oct. 2019), p. 106502. ISSN: 1083-3668. DOI: 10.1117/1.JBO.24.10.106502. PMID: 31625322.

[399] Predrag Jevtić et al. *Sizing and Shaping the Nucleus: Mechanisms and Significance*. In: *Current Opinion in Cell Biology*. Cell Nucleus 28 (1st June 2014), pp. 16–27. ISSN: 0955-0674. DOI: 10.1016/j.ceb.2014.01.003.

[400] Shruthi Balachandra, Sharanya Sarkar and Amanda A. Amodeo. *The Nuclear-to-Cytoplasmic Ratio: Coupling DNA Content to Cell Size, Cell Cycle, and Biosynthetic Capacity*. In: *Annual review of genetics* 56 (30th Nov. 2022), pp. 165–185. ISSN: 0066-4197. DOI: 10.1146/annurev-genet-080320-030537. PMID: 35977407.

[401] Jan Lammerding et al. *Lamins A and C but Not Lamin B1 Regulate Nuclear Mechanics **. In: *Journal of Biological Chemistry* 281.35 (1st Sept. 2006), pp. 25768–25780. ISSN: 0021-9258, 1083-351X. DOI: 10.1074/jbc.M513511200.

[402] Niina Dubik and Sabine Mai. *Lamin A/C: Function in Normal and Tumor Cells*. In: *Cancers* 12.12 (9th Dec. 2020), p. 3688. ISSN: 2072-6694. DOI: 10.3390/cancers12123688. PMID: 33316938.

[403] Griffin J. Salus, Marija Strojanova and Christopher M. Heaphy. *Lamin A/C Downregulation Alters Prostate Cancer Cellular Phenotypes*. In: *The FASEB Journal* 36.S1 (2022). ISSN: 1530-6860. DOI: 10.1096/fasebj.2022.36.S1.R2319.

[404] Chloe M. Funkhouser et al. *Mechanical Model of Blebbing in Nuclear Lamin Meshworks*. In: *Proceedings of the National Academy of Sciences* 110.9 (26th Feb. 2013), pp. 3248–3253. DOI: 10.1073/pnas.1300215110.

[405] Sanghee Lim, Ryan J. Quinton and Neil J. Ganem. *Nuclear Envelope Rupture Drives Genome Instability in Cancer*. In: *Molecular Biology of the Cell* 27.21 (Nov. 2016). Ed. by William Bement, pp. 3210–3213. ISSN: 1059-1524, 1939-4586. DOI: 10.1091/mbc.E16-02-0098.

[406] Zhixiao Liu et al. *Heterogeneous Responses to Mechanical Force of Prostate Cancer Cells Inducing Different Metastasis Patterns*. In: *Advanced Science* 7.15 (17th June 2020), p. 1903583. ISSN: 2198-3844. DOI: 10.1002/advs.201903583. PMID: 32775149.

[407] Deepraj Ghosh et al. *Spatial Heterogeneity in Cytoskeletal Mechanics Response to TGF-B1 and Hypoxia Mediates Partial Epithelial-to-Mesenchymal Transition in Epithelial Ovarian Cancer Cells*. In: *Cancers* 15.12 (Jan. 2023), p. 3186. ISSN: 2072-6694. DOI: 10.3390/cancers15123186.

[408] Joseph D. Schonhoft et al. *Morphology-Predicted Large-Scale Transition Number in Circulating Tumor Cells Identifies a Chromosomal Instability Biomarker Associated with Poor Outcome in Castration-Resistant Prostate Cancer*. In: *Cancer Research* 80.22 (13th Nov. 2020), pp. 4892–4903. ISSN: 0008-5472. DOI: 10.1158/0008-5472.CAN-20-1216.

[409] James L. Mohler et al. *Nuclear Roundness Factor Measurement for Assessment of Prognosis of Patients with Prostatic Carcinoma. II. Standardization of Methodology for Histologic Sections*. In: *Journal of Urology* 139.5 (May 1988), pp. 1085–1090. ISSN: 0022-5347, 1527-3792. DOI: 10.1016/S0022-5347(17)42792-3.

[410] Alan W. Partin et al. *A Comparison of Nuclear Morphometry and Gleason Grade as a Predictor of Prognosis in Stage A2 Prostate Cancer: A Critical Analysis*. In: *Journal of Urology* 142.5 (Nov. 1989), pp. 1254–1258. ISSN: 0022-5347, 1527-3792. DOI: 10.1016/S0022-5347(17)39049-3.

[411] Shyam B. Khatau et al. *A Perinuclear Actin Cap Regulates Nuclear Shape*. In: *Proceedings of the National Academy of Sciences* 106.45 (10th Nov. 2009), pp. 19017–19022. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.0908686106.

[412] DAVID B. Lovett et al. *Modulation of Nuclear Shape by Substrate Rigidity*. In: *Cellular and molecular bioengineering* 6.2 (1st June 2013), pp. 230–238. ISSN: 1865-5025. DOI: 10.1007/s12195-013-0270-2. PMID: 23914256.

[413] Sayed Reza Ramezani, Afsaneh Mojra and Mohammad Tafazzoli-Shadpour. *Mechanochemical Interactions in Cancer Cells: The Role of Substrate Stiffness in Cell Behavior and Drug Response*. In: *PLOS ONE* 21.1 (7th Jan. 2026), e0327874. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0327874.

[414] Yuan Li et al. *Moving Cell Boundaries Drive Nuclear Shaping during Cell Spreading*. In: *Biophysical Journal* 109.4 (18th Aug. 2015), pp. 670–686. ISSN: 0006-3495. DOI: 10.1016/j.bpj.2015.07.006.

[415] G. Carruba et al. *Morphometry of In Vitro Systems*. In: *Pathology - Research and Practice* (1989). DOI: 10.1016/S0344-0338(89)80222-5.

[416] James F. Gillooly, Andrew Hein and Rachel Damiani. *Nuclear DNA Content Varies with Cell Size across Human Cell Types*. In: *Cold Spring Harbor Perspectives in Biology* 7.7 (July 2015), a019091. ISSN: 1943-0264. DOI: 10.1101/cshperspect.a019091. PMID: 26134319.

[417] Ryosuke Imai et al. *Density Imaging of Heterochromatin in Live Cells Using Orientation-Independent-DIC Microscopy*. In: *Molecular Biology of the Cell* 28.23 (7th Nov. 2017), pp. 3349–3359. ISSN: 1059-1524. DOI: 10.1091/mbc.E17-06-0359. PMID: 28835378.

[418] Bunty Sharma et al. *Epigenetic Insights into Prostate Cancer: Exploring Histone Modifications and Their Therapeutic Implications*. In: *Frontiers in Oncology* 15 (28th Apr. 2025). ISSN: 2234-943X. DOI: 10.3389/fonc.2025.1570193.

[419] Jing Xia et al. *Chromatin Heterogeneity Modulates Nuclear Condensate Dynamics and Phase Behavior*. In: *Nature Communications* 16.1 (11th July 2025), p. 6406. ISSN: 2041-1723. DOI: 10.1038/s41467-025-60771-9.

[420] Brian T Helfand et al. *Chromosomal Regions Associated with Prostate Cancer Risk Localize to Lamin B-deficient Microdomains and Exhibit Reduced Gene Transcription*. In: *The Journal of Pathology* 226.5 (2012), pp. 735–745. ISSN: 1096-9896. DOI: 10.1002/path.3033.

[421] Tianwei Jia et al. *FOXA1 Regulates Ribosomal RNA Transcription in Prostate Cancer*. In: *The Prostate* 84.10 (2024), pp. 967–976. ISSN: 1097-0045. DOI: 10.1002/pros.24714.

[422] Maria Victoria Neguembor et al. *Transcription-Mediated Supercoiling Regulates Genome Folding and Loop Formation*. In: *Molecular Cell* 81.15 (5th Aug. 2021), 3065–3081.e12. ISSN: 1097-2765. DOI: 10.1016/j.molcel.2021.06.009. PMID: 34297911.

[423] M. S. Lin, D. E. Comings and O. S. Alfi. *Optical Studies of the Interaction of 4-6-Diamidino-2-Phenylindole with DNA and Metaphase Chromosomes*. In: *Chromosoma* 60.1 (1st Mar. 1977), pp. 15–25. ISSN: 1432-0886. DOI: 10.1007/BF00330407.

[424] Shuping Li et al. *Angiogenin Mediates Androgen-Stimulated Prostate Cancer Growth and Enables Castration Resistance*. In: *Molecular Cancer Research* 11.10 (1st Oct. 2013), pp. 1203–1214. ISSN: 1541-7786. DOI: 10.1158/1541-7786.MCR-13-0072.

[425] Darko Stefanovski et al. *Prostate Cancer Diagnosis Using Epigenetic Biomarkers, 3D High-Content Imaging and Probabilistic Cell-by-Cell Classifiers*. In: *Oncotarget* 8.34 (5th July 2017), pp. 57278–57301. ISSN: 1949-2553. DOI: 10.18632/oncotarget.18985.

[426] Trond Jørgensen et al. *Nuclear texture analysis: A new prognostic tool in metastatic prostate cancer*. In: *Cytometry* 24.3 (1996), pp. 277–283. ISSN: 1097-0320. DOI: 10.1002/(SICI)1097-0320(19960701)24:3<277::AID-CYTO11>3.0.CO;2-N.

[427] Mahmoud Mohamed et al. *Changes in Chromatin Phenotype Predict the Response to Hormonal Deprivation Therapy in Patients with Prostate Cancer*. In: *BJU international* 103.3 (Feb. 2009), pp. 391–398. ISSN: 1464-410X. DOI: 10.1111/j.1464-410X.2008.08063.x. PMID: 19021609.

[428] Michael C. Haffner et al. *Genomic and Phenotypic Heterogeneity in Prostate Cancer*. In: *Nature Reviews Urology* 18.2 (Feb. 2021), pp. 79–92. ISSN: 1759-4820. DOI: 10.1038/s41585-020-00400-w.

[429] Shannon E. Weeks, Brandon J. Metge and Rajeev S. Samant. *The Nucleolus: A Central Response Hub for the Stressors That Drive Cancer Progression*. In: *Cellular and Molecular Life Sciences* 76.22 (1st Nov. 2019), pp. 4511–4524. ISSN: 1420-9071. DOI: 10.1007/s00018-019-03231-0.

[430] M. Derenzini and D. Trerè. *Silver-stained Nucleolar Organizer Regions (AgNOR)*. In: *Pathologica* 93.2 (Apr. 2001), pp. 99–105. ISSN: 0031-2983. PMID: 11428299.

[431] Hogyoung Kim et al. *Estradiol-ERB2 Signaling Axis Confers Growth and Migration of CRPC Cells through TMPRSS2-ETV5 Gene Fusion*. In: *Oncotarget* 8.38 (17th Aug. 2016), pp. 62820–62833. ISSN: 1949-2553. DOI: 10.18632/oncotarget.11355.

[432] Leonard C. Shank et al. *Activation of the DNA-dependent Protein Kinase Stimulates Nuclear Export of the Androgen Receptor* in Vitro*. In: *Journal of Biological Chemistry* 283.16 (18th Apr. 2008), pp. 10568–10580. ISSN: 0021-9258. DOI: 10.1074/jbc.M800810200.

[433] Fabien Mongelard et al. *Quantitative Assessment of the Alteration of Chromatin during the Course of FISH Procedures*. In: *Cytometry* 36.2 (1999), pp. 96–101. ISSN: 1097-0320. DOI: 10.1002/(SICI)1097-0320(19990601)36:2<96::AID-CYTO2>3.0.CO;2-X.

[434] Anne R. Shim et al. *Formamide Denaturation of Double-Stranded DNA for Fluorescence in Situ Hybridization (FISH) Distorts Nanoscale Chromatin Structure*. In: *PLOS ONE* 19.5 (28th May 2024), e0301000. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0301000. PMID: 38805476.

[435] Yuan Qin et al. *The Combination of Paraformaldehyde and Glutaraldehyde Is a Potential Fixative for Mitochondria*. In: *Biomolecules* 11.5 (9th May 2021). ISSN: 2218-273X. DOI: 10.3390/biom11050711.

[436] UK Health Security Agency. *Cell Line Profile: PNT2*. Culture Collections. Dec. 2025. URL: https://www.culturecollections.org.uk/products/cell-cultures/ecacc-cell-line-profiles/pnt2/ (visited on 13/02/2026).

[437] Thermo Fisher Scientific. *Preparing Fixed Cells for Labeling*. URL: https://www.thermofisher.com/uk/en/home/life-science/cell-analysis/cell-analysis-learning-center/molecular-probes-school-of-fluorescence/imaging-basics/sample-considerations/preparing-fixed-cells-imaging.html (visited on 12/02/2026).

[438] Johannes Fels, Sergei N. Orlov and Ryszard Grygorczyk. *The Hydrogel Nature of Mammalian Cytoplasm Contributes to Osmosensing and Extracellular pH Sensing*. In: *Biophysical Journal* 96.10 (May 2009), pp. 4276–4285. ISSN: 00063495. DOI: 10.1016/j.bpj.2009.02.038.

[439] Catherine G. Chu et al. *Lamin B Loss in Nuclear Blebs Is Rupture Dependent Whereas Increased DNA Damage Is Rupture Independent*. In: *Journal of Cell Science* 138.21 (13th Nov. 2025), jcs263945. ISSN: 0021-9533. DOI: 10.1242/jcs.263945. PMID: 41047936.

[440] Ricardo Piña et al. *Ten Approaches That Improve Immunostaining: A Review of the Latest Advances for the Optimization of Immunofluorescence*. In: *International Journal of Molecular Sciences* 23.3 (25th Jan. 2022). ISSN: 1422-0067. DOI: 10.3390/ijms23031426.

[441] Bruno Mattei et al. *Membrane Permeabilization Induced by Triton X-100: The Role of Membrane Phase State and Edge Tension*. In: *Chemistry and Physics of Lipids* 202 (1st Jan. 2017), pp. 28–37. ISSN: 0009-3084. DOI: 10.1016/j.chemphyslip.2016.11.009.

[442] Emily M. Hatch and Martin W. Hetzer. *Nuclear Envelope Rupture Is Induced by Actin-Based Nucleus Confinement*. In: *Journal of Cell Biology* 215.1 (3rd Oct. 2016), pp. 27–36. ISSN: 0021-9525. DOI: 10.1083/jcb.201603053.

[443] Tanmay P. Lele, Richard B. Dickinson and Gregg G. Gundersen. *Mechanical Principles of Nuclear Shaping and Positioning*. In: *Journal of Cell Biology* 217.10 (7th Sept. 2018), pp. 3330–3342. ISSN: 0021-9525. DOI: 10.1083/jcb.201804052.

[444] Kirby and Lammerding. *Emerging Views of the Nucleus as a Cellular Mechanosensor*. In: *Nature Cell Biology* (2018). DOI: 10.1038/s41556-018-0038-y.

[445] Marie Versaevel et al. *Super-Resolution Microscopy Reveals LINC Complex Recruitment at Nuclear Indentation Sites*. In: *Scientific Reports* 4.1 (8th Dec. 2014), p. 7362. ISSN: 2045-2322. DOI: 10.1038/srep07362.

[446] Helenius and Simons. *Solubilization of Membranes by Detergents*. In: *Biochimica et Biophysica Acta (BBA) - Reviews on Biomembranes* (1975). DOI: 10.1016/0304-4157(75)90016-7.

[447] J. Erenpreisa and N. Zaleskaya. *Effect of Triton X-100 on Cytochemical and Ultrastructural Pattern of Chromatin*. In: *Acta Morphologica Hungarica* 31.4 (1983), pp. 387–393. ISSN: 0236-5391. PMID: 6421099.

[448] Natalya N. Degtyareva et al. *Hydration Changes Accompanying the Binding of Minor Groove Ligands with DNA*. In: *Biophysical Journal* 92.3 (1st Feb. 2007), pp. 959–965. ISSN: 0006-3495. DOI: 10.1529/biophysj.106.097451. PMID: 17114230.

[449] Yanming Wang et al. *Histone Hypercitrullination Mediates Chromatin Decondensation and Neutrophil Extracellular Trap Formation*. In: *The Journal of Cell Biology* 184.2 (26th Jan. 2009), pp. 205–213. ISSN: 0021-9525. DOI: 10.1083/jcb.200806072. PMID: 19153223.

[450] Maha Shahid, Kanwal Iftikhar and Shabana Usman Simjee. *Impact of Triton X-100 on Notch 1 Surface Receptor Immunofluorescence: A Cautionary Study*. In: *Discoveries* 13.1 (31st Mar. 2025), e206. ISSN: 2359-7232. DOI: 10.15190/d.2025.5. PMID: 40351502.

[451] André Huisman et al. *Discrimination between Benign and Malignant Prostate Tissue Using Chromatin Texture Analysis in 3-D by Confocal Laser Scanning Microscopy*. In: *The Prostate* 67.3 (2007), pp. 248–254. ISSN: 1097-0045. DOI: 10.1002/pros.20507.

[452] Minjeong Jo et al. *Reduced Dynamicity and Increased High-Order Protein Assemblies in Dense Fibrillar Component of the Nucleolus under Cellular Senescence*. In: *Redox Biology* 75 (1st Sept. 2024), p. 103279. ISSN: 2213-2317. DOI: 10.1016/j.redox.2024.103279.

[453] Marzia Tagliaferro et al. *Nucleolar Localization of the ErbB3 Receptor as a New Target in Glioblastoma*. In: *BMC Molecular and Cell Biology* 23.1 (7th Mar. 2022), p. 13. ISSN: 2661-8850. DOI: 10.1186/s12860-022-00411-y.

[454] Caiwei Jia et al. *Tumor Diagnosis Based on Nucleolus Labeling*. In: *Sensors & Diagnostics* 3.11 (7th Nov. 2024), pp. 1807–1821. ISSN: 2635-0998. DOI: 10.1039/D4SD00238E.

[455] Gregorio et al. *Nucleolar Organizer Regions in Malignant Melanoma and Melanocytic Nevi*. In: *The American Journal of Dermatopathology* (1991). DOI: 10.1097/00000372-199108000-00002.

[456] Lijuan Jiao et al. *Ribosome Biogenesis in Disease: New Players and Therapeutic Targets*. In: *Signal Transduction and Targeted Therapy* 8.1 (9th Jan. 2023), p. 15. ISSN: 2059-3635. DOI: 10.1038/s41392-022-01285-4.

[457] Squatrito et al. *EBP1 Is a Nucleolar Growth-Regulating Protein That Is Part of Pre-Ribosomal Ribonucleoprotein Complexes*. In: *Oncogene* (2004). DOI: 10.1038/sj.onc.1207579.

Appendix A

Methods

A.1 Exact implementation details for Stage 1 DAPI segmentation and FIB nucleolar assignment

This appendix records the exact parameter-level implementation underlying the Stage 1 image-processing steps summarised in Figure 2.4, the effects of which are shown concretely in Figures 2.5 and 2.6. The purpose of moving these details here is simply, to attempt to keep the main Methods focused on the operational logic of the pipeline, while still preserving full reproducibility.

Background correction and image branches

Uneven low-frequency background was reduced by headless ImageJ/Fiji rolling-ball subtraction with a rolling radius of 250 pixels [313, 321]. Native 16-bit images were retained as such, and where an input image was 8-bit, intensities were promoted to 16-bit before further processing. In practice, all images analysed in this thesis were native 16-bit. After background subtraction, the script retained two related image branches: a background-subtracted measurement image used for quantitative DAPI extraction, and a separate segmentation image created by min-max normalisation for mask generation only. This split is illustrated by the transition from Figure 2.5(b) to Figure 2.5(c).

Foreground detection

The segmentation image was smoothed using a Gaussian blur with a 5×5 kernel and $\sigma = 2$ pixels [322]. Multi-Otsu thresholding was then performed on an 8-bit representation of the segmentation image to keep runtime reasonable, partitioning the histogram into three classes [317, 323]. If T_1 denotes the first multi-Otsu threshold, the working foreground threshold was defined as

$$T_{\mathrm{fg}} = 1.2\,T_1, \tag{A.1}$$

and pixels satisfying $D_{\mathrm{seg}}(x, y) > T_{\mathrm{fg}}$ were assigned to the provisional nuclear foreground mask, where D_{seg} denotes the pre-processed DAPI image used only for segmentation. The progression from pre-processed input to smoothed image to provisional foreground mask is shown in Figures 2.5(c) to 2.5(e).

Separation of touching nuclei

The Euclidean distance transform was computed on the binary foreground mask [324]. Local maxima in this distance map were then used as candidate nucleus markers, subject to two suppression constraints intended to limit over-segmentation: a minimum spacing of 25 pixels between peaks, and a disk-shaped suppression footprint of radius 70 pixels. Watershed segmentation was then applied to the negative distance map, constrained to the foreground mask [325]. These steps are illustrated in Figures 2.5(f) and 2.5(g).

Post-segmentation filtering

Two post-segmentation filtering rules were applied. First, an area filter removed labelled objects with projected area < 9500 pixels or > 150000 pixels. Using the microscope calibration $s = 0.07\,\mu\mathrm{m\,px}^{-1}$, these cut-offs correspond approximately to $47\,\mu\mathrm{m}^2$ and $735\,\mu\mathrm{m}^2$, respectively. Second, an over-brightness filter was used to exclude saturated or otherwise aberrant objects. For each image, a high-intensity threshold was defined as

$$T_{\mathrm{sat}} = 0.95\, D_{\max}, \tag{A.2}$$

where $D_{\max}$ is the maximum DAPI intensity in that image. Any segmented object for which more than 10% of pixels exceeded T_{sat} was removed. Retained nuclei were then relabelled consecutively, and the numbers of detected and removed objects were recorded for auditability. The effect of this filtering step is illustrated by the transition from Figure 2.5(g) to Figure 2.5(h).

Global FIB nucleolar-candidate threshold

Global intranuclear FIB thresholding followed the procedure defined in Equations (2.16) to (2.18). In implementation terms, the pooled intranuclear FIB intensities for each image were histogrammed using 256 bins, and three-class multi-Otsu thresholding was applied to that pooled distribution. If multi-Otsu thresholding succeeded, the working global threshold T_{global} was computed according to Equation (2.18). If multi-Otsu thresholding failed for a given image, the fallback threshold was defined as the 90.5th percentile of the pooled intranuclear FIB pixel distribution. The corresponding global candidate-assignment stage is illustrated in Figure 2.6(c).

Per-nucleus guardrail reassignment

A nucleus-level guardrail step was then applied to reduce obvious global-threshold misassignment in unusually dim or unusually bright nuclei. For retained nucleus j, let

$$f_j = \frac{1}{|\mathcal{F}_j|} \sum_{p \in \mathcal{F}_j} \mathbb{I}(p > T_{\mathrm{global}}) \tag{A.3}$$

denote the fraction of intranuclear pixels exceeding the global threshold, where $\mathbb{I}(\cdot)$ is the indicator function and $\mathcal{F}_j$ is defined in Equation (2.16). If $|\mathcal{F}_j| \geq 2048$ pixels and $f_j < 0.002$ or $f_j > 0.35$, the threshold for that nucleus was replaced by the nucleus-specific 90.5th percentile of $\mathcal{F}_j$. Otherwise, the global threshold T_{global} was retained. The final per-nucleus nucleolar assignment therefore corresponds to the union of pixels satisfying the threshold ultimately selected for each nucleus individually, as shown in Figure 2.6(d).

Appendix B

Convergence

B.1 Non-permeabilised data

B.1.1 Summary table

Cell line	Metric	γ	*pow*	NRMSE	χ^2_ν	N_{conv}
LNCaP ($N_{\max} = 657$)	Area	0.0114 ± 0.0005	-0.498 ± 0.006	0.0972	0.001 01	365
	Aspect ratio error	1500 ± 200	-0.48 ± 0.02	0.362	0.004 42	—
	Aspect ratio	1.94 ± 0.08	-0.498 ± 0.006	0.104	9.61×10^{-4}	170
	Bright ratio	3.24 ± 0.07	-0.499 ± 0.004	0.0686	6.16×10^{-4}	460
	Dark ratio	7.4 ± 0.7	-0.49 ± 0.01	0.284	0.003 48	—
	Density	$(7.1 \pm 0.2) \times 10^{-7}$	-0.499 ± 0.004	0.0512	5.02×10^{-4}	190
	Intensity	$(2.6 \pm 0.1) \times 10^{-9}$	-0.497 ± 0.007	0.119	0.001 21	405
PNT2-C2 ($N_{\max} = 2378$)	Area	0.0215 ± 0.0003	-0.499 ± 0.002	0.113	5.50×10^{-4}	225
	Aspect ratio error	4300 ± 500	-0.45 ± 0.02	0.444	0.002 93	—
	Aspect ratio	3.72 ± 0.04	-0.500 ± 0.001	0.0838	3.94×10^{-4}	70
	Bright ratio	3.18 ± 0.02	-0.4999 ± 0.0006	0.0345	1.76×10^{-4}	660
	Dark ratio	15.0 ± 0.7	-0.493 ± 0.006	0.460	0.003 04	—
	Density	$(8.36 \pm 0.07) \times 10^{-7}$	-0.500 ± 0.001	0.0604	2.82×10^{-4}	145
	Intensity	$(3.82 \pm 0.08) \times 10^{-9}$	-0.498 ± 0.003	0.180	8.95×10^{-4}	400

Table B.1: Summary of convergence analysis for non-permeabilised LNCaP and PNT2-C2 cells stained with DAPI. Power law fit parameters and goodness of fit statistics refer to weighted fits performed to the mean SEM calculated through repeated random subsampling as a function of N, where γ is the rate of convergence (defined as $\gamma = \frac{1}{A}$), and *pow* is the fitted exponent (Equations (2.38) and (2.40)). The convergence point, N_{conv}, is defined by the earliest sampled N for which the percentage standard error remains below 2.5% for all subsequent samples (reported as a discrete value without uncertainty; Equation (2.39)). NRMSE refers to normalised root mean square error, normalised to terminal SEM, and χ^2_ν refers to reduced chi squared Equations (2.46) and (2.48). $N_{\max}$ is the total number of nuclei available in each cell line.

B.2 DAPI-only permeabilised data

B.2.1 Summary table

Cell line	Metric	γ	*pow*	NRMSE	χ^2_ν	N_{conv}
LNCaP ($N_{\max} = 944$)	Area	0.0134 ± 0.0006	-0.496 ± 0.007	0.142	0.001 15	340
	Aspect ratio error	3000 ± 300	-0.48 ± 0.01	0.363	0.003 59	—
	Aspect ratio	2.0 ± 0.1	-0.494 ± 0.008	0.127	0.0011	195
	Bright ratio	3.40 ± 0.03	-0.500 ± 0.001	0.0290	2.33×10^{-4}	—
	Dark ratio	7.6 ± 0.5	-0.493 ± 0.009	0.257	0.002 56	—
	Density	$(1.05 \pm 0.02) \times 10^{-6}$	-0.499 ± 0.003	0.0586	4.97×10^{-4}	140
	Intensity	$(3.4 \pm 0.2) \times 10^{-9}$	-0.495 ± 0.008	0.166	0.001 39	470
PNT2-C2 ($N_{\max} = 1438$)	Area	0.0115 ± 0.0002	-0.499 ± 0.003	0.0961	5.90×10^{-4}	310
	Aspect ratio error	3200 ± 600	-0.43 ± 0.03	0.324	0.002 11	—
	Aspect ratio	3.3 ± 0.1	-0.497 ± 0.005	0.138	9.46×10^{-4}	100
	Bright ratio	3.96 ± 0.07	-0.499 ± 0.002	0.136	6.67×10^{-4}	—
	Dark ratio	3.26 ± 0.04	-0.500 ± 0.002	0.0836	4.25×10^{-4}	—
	Density	$(9.1 \pm 0.2) \times 10^{-7}$	-0.499 ± 0.003	0.0898	5.49×10^{-4}	300
	Intensity	$(3.2 \pm 0.1) \times 10^{-9}$	-0.497 ± 0.005	0.149	9.92×10^{-4}	595

Table B.2: Summary of convergence analysis for permeabilised cells stained with only DAPI. Power law fit parameters and goodness of fit statistics refer to weighted fits performed to the mean SEM calculated through repeated random subsampling as a function of N, where γ is the rate of convergence (defined as $\gamma = \frac{1}{A}$), and *pow* is the fitted exponent. The convergence point, N_{conv}, is defined by the earliest sampled N for which the percentage standard error remains below 2.5% for all subsequent samples (reported as a discrete value without uncertainty). NRMSE refers to normalised root mean square error, normalised to terminal SEM, and χ^2_ν refers to reduced chi squared. $N_{\max}$ is the total number of nuclei available in each cell line.

B.3 DAPI+FIB permeabilised data

B.3.1 Summary table

Cell line	Metric	γ	*pow*	NRMSE	χ^2_ν	N_{conv}
LNCaP ($N_{\max} = 460$)	Area	0.0115 ± 0.0008	-0.49 ± 0.01	0.119	0.001 46	345
	Aspect ratio error	1800 ± 600	-0.40 ± 0.05	0.218	0.002 28	—
	Aspect ratio	1.5 ± 0.1	-0.49 ± 0.01	0.121	0.001 55	235
	Bright pixel ratio	3.26 ± 0.07	-0.499 ± 0.004	0.0428	5.38×10^{-4}	—
	DAPI:FIB ratio	1.0 ± 0.2	-0.47 ± 0.03	0.104	0.001 06	240
	Dark pixel ratio	7.9 ± 0.9	-0.48 ± 0.02	0.232	0.003 65	—
	DAPI density	$(7.7 \pm 0.2) \times 10^{-7}$	-0.499 ± 0.005	0.0498	6.29×10^{-4}	190
	Fibrillarin intensity	$(1.06 \pm 0.09) \times 10^{-8}$	-0.49 ± 0.01	0.133	0.001 79	355
	DAPI intensity	$(2.8 \pm 0.3) \times 10^{-9}$	-0.49 ± 0.02	0.150	0.001 97	420
	Nucleoli area	0.089 ± 0.005	-0.50 ± 0.01	0.102	0.001 29	—
	Nucleoli intensity	$(2.7 \pm 0.2) \times 10^{-8}$	-0.49 ± 0.01	0.0976	0.001 22	455
	Nucleoli-nucleus area ratio	27 ± 1	-0.498 ± 0.007	0.0870	0.001 11	230
PNT2-C2 ($N_{\max} = 1627$)	Area	0.0107 ± 0.0002	-0.499 ± 0.003	0.119	7.06×10^{-4}	345
	Aspect ratio error	2900 ± 300	-0.48 ± 0.01	0.433	0.003 44	—
	Aspect ratio	3.44 ± 0.09	-0.498 ± 0.003	0.136	7.99×10^{-4}	95.0
	Bright pixel ratio	2.84 ± 0.02	-0.4999 ± 0.0008	0.0349	1.88×10^{-4}	840
	DAPI:FIB ratio	0.476 ± 0.008	-0.499 ± 0.002	0.104	5.80×10^{-4}	450
	Dark pixel ratio	5.0 ± 0.1	-0.498 ± 0.004	0.293	0.001 63	—
	DAPI density	$(6.34 \pm 0.06) \times 10^{-7}$	-0.500 ± 0.001	0.0535	2.96×10^{-4}	250
	Fibrillarin intensity	$(8.8 \pm 0.2) \times 10^{-9}$	-0.498 ± 0.003	0.137	8.27×10^{-4}	450
	DAPI intensity	$(1.97 \pm 0.06) \times 10^{-9}$	-0.498 ± 0.004	0.148	9.19×10^{-4}	615
	Nucleoli area	0.032 ± 0.002	-0.487 ± 0.009	0.394	0.003 30	—
	Nucleoli intensity	$(2.19 \pm 0.08) \times 10^{-8}$	-0.496 ± 0.005	0.178	0.001 18	795

	Nucleoli-nucleus area ratio	9.3 ± 0.5	-0.493 ± 0.007	0.352	0.002 78	1050

Table B.3: Summary of convergence analysis for permeabilised LNCaP and PNT2-C2 cells stained with DAPI and anti-fibrillarin (DAPI+FIB). Power law fit parameters and goodness of fit statistics refer to weighted fits performed to the mean SEM calculated through repeated random subsampling as a function of N, where γ is the rate of convergence (defined as $\gamma = \frac{1}{A}$), and *pow* is the fitted exponent. The convergence point, N_{conv}, is defined by the earliest sampled N for which the percentage standard error remains below 2.5% for all subsequent samples (reported as a discrete value without uncertainty). NRMSE refers to normalised root mean square error, normalised to terminal SEM, and χ^2_ν refers to reduced chi squared. N_{max} is the total number of nuclei available in each cell line.