Engineering a Single-Cell Platform to Study Estrogen Receptor-α Regulation of the Genome in Breast Cancer Microenvironments.

Jack Philip Stenning

Ph.D.

University of York

Biology

November 2025

Abstract

Approximately 70% of breast cancers are estrogen receptor (ER) positive, and endocrine therapies targeting ERα remain the cornerstone of treatment. Despite their efficacy, nearly one-third of patients relapse due to acquired resistance, often accompanied by metastasis and poor prognosis. Tumour hypoxia, a hallmark of the breast cancer microenvironment, exacerbates intra-tumour heterogeneity, reshaping transcriptional programmes and contributing to therapy resistance. Hypoxia-inducible factors (HIFs) alter chromatin accessibility and transcription factor (TF) activity, yet the mechanisms by which hypoxia influences ERα binding heterogeneity remain unresolved.

Traditional methods, such as chromatin immunoprecipitation followed by sequencing (ChIP-Seq), have defined ERα binding landscapes but cannot capture heterogeneity in single-cell resolution. Recent advances in single-cell calling card (scCC) assays, in which engineered TF-transposase fusion proteins deposit genomic tags at binding sites, make scCC an attractive method to study ERα binding and transcriptomic heterogeneity across tumours. Self-reporting transposons are recoverable from mRNA, allowing direct linkage of TF binding events to cell identity. When integrated with single-cell RNA sequencing, Calling Cards uniquely provide simultaneous readouts of TF binding and cell fate.

In this thesis, we present the engineering and validation of a novel HyPB-ERα transposase fusion protein that reproduces ERα binding patterns across the genome. Bulk and long-read Calling Card assays demonstrate strong concordance with gold-standard ChIP-Seq and ChIA-PET datasets, confirming specificity and biological relevance. Building on this, we establish a three-dimensional MCF7 spheroid model that recapitulates hypoxia gradients observed in vivo, enabling interrogation of ERα binding heterogeneity under physiologically relevant conditions.

Together, these advances culminate in the development of a scCC ERα platform, integrating TF binding, transcriptomic identity, and microenvironmental context, working towards the aim of enabling investigations into the heterogeneity of ERα binding and gene expression across cancer models. Future applications of this work will enable mechanistic insight into how hypoxia drives ERα heterogeneity and therapy resistance.

Table of Contents

Abstract 2
Table of Contents 3
List of Figures 9
List of Tables 13
Abbreviations 15
Acknowledgments 18
Declaration 20
1. Introduction 21
1.1. Overview of ERα+ breast cancer 21
1.1.1. Introduction to Breast Cancer 21
1.1.2. The Estrogen Receptor Isoforms and their relevance to breast cancer research 23
1.1.3. The structure and function of ERα in breast cancer 25
1.2. Treatment of breast cancer with a focus on ERα+ disease 30
1.2.1. Treatment of ERα+ cancer specifically and how it differs from other subtypes 30
1.2.2. Prognosis in ERα-positive breast cancer 33
1.2.3. Resistance to endocrine therapy 34
1.2.4. Resistance and heterogeneity 35
1.3. Hypoxia and microenvironment in breast cancer 36
1.3.1. Primary mechanisms of the cellular hypoxic response 36
1.3.2. Cell fate consequences of HIF-α expression 37
1.3.3. Hypoxia-driven heterogeneity's effect on breast tumorigenesis 38
1.4. Three-dimensional cell culture methods used to study breast cancer 39
1.5. Methods to study transcription factor binding across mammalian genomes 41
1.5.1. Establishment of the gold standard - Chromatin immunoprecipitation and next-generation sequencing (ChIP-Seq) 41
1.5.2. ChIP-Seq is the gold standard for transcription factor binding analysis 42
1.5.3. Limitations of ChIP-Seq 42
1.5.4. Optimisation of ChIP-Seq for low input samples 44
1.5.5. Optimisation of ChIP-Seq for single-cell analysis 44
1.5.6. Alternatives to ChIP-Seq that avoid immunoprecipitation 46
1.5.7. Self-reporting transposons generate 'Calling Cards' to identify transcription factor binding sites 49
1.5.8. First applications of calling cards in yeast 49

1.5.9. Application of Calling Cards to mammalian systems 50
1.5.10. Single-cell calling cards 51
1.6. Project introduction 52
1.6.1. Hypothesis 52
1.6.2. Rationale 52
1.6.3. Our research aim 55
1.6.4. Research Objectives 55
2. Materials and Methods 56
2.1. General molecular biology methods and construct optimisation 56
2.1.1. Culturing competent bacteria 56
2.1.2. Fusion protein cloning 56
2.1.3. Construct optimisation 59
2.2. MCF7 cell culture, transfection, and puromycin selection 60
2.2.1. Cell recovery and maintenance 60
2.2.2. Lipofectamine 3000 transfection 61
2.2.3. Puromycin dose-response curve for MCF7 cells 62
2.3. Protein expression and HyPB functional validation 63
2.3.1. Expression confirmation via Western Blotting 63
2.3.2. HyPB fusion partner functional validation with TdTomato and BrokenHeart 66
2.4. ERα Calling Card Sequencing 68
2.4.1. Bulk ERα Calling Cards 68
2.4.2. Long-Read ERα Calling Cards 73
2.5. Computational analysis 79
2.5.1. Bulk ERα Calling Card pipeline (Chapter 3) 79
2.5.2. Long-read ERα Calling Card pipeline (Chapter 4) 79
2.5.3. Downstream statistical analyses 80
2.6. Spheroid culture and hypoxia modelling 81
2.6.1. General workflow 81
2.6.2. Seeding density optimisation 81
2.6.3. Validating MCF7 spheroids as a hypoxia model 82
2.6.4. Treating MCF7 spheroids with Amiloride 84
2.6.5. Image data analysis 85
3. Bulk HyPB-ERα Calling Cards demonstrates that our novel fusion protein can replicate ERα binding across the MCF7 genome 87
3.1. Introduction 87
3.2. Results 89

3.2.1. The ERα coding sequence was cloned at the C-and N-terminus of the HyPB coding sequence to generate a fusion protein expression plasmid 89
3.2.2. Full-length HyPB-ERα fusion protein expression was confirmed via Western Blot following transfection into MCF7 and MDA-MB-231 cell lines. 93
3.2.3. The HyPB-ERα fusion proteins are functionally active transposases. 96
3.2.4. MCF7 cells transfected with C-terminal HyPB-ERα calling card reagents produce sufficient mRNA to generate sequencing libraries 99
3.2.5. HyPB-ERα calling card normalised insertions at called peaks display correlation 101
3.2.6. HyPB-ERα calling cards show a significant correlation with ERα ChIP-Seq peak distribution across chromosomes 108
3.2.7. HyPB-ERα Calling Cards are distributed at expected genomic features and distances from gene transcription start sites 114
3.2.8. HyPB-ERα bulk calling cards peaks display statistically significant overlap with ChIP-Seq peaks 118
3.2.9. ChIP-Seq signal is enriched at the centre of HyPB-ESR1 calling card peaks 122
3.2.10. Sensitivity and Specificity of calling card peaks using an adjusted p-value threshold is not a good indicator of ChIP-Seq overlap 126
3.2.11. BRD4 peaks overlap highly with ERα Calling Card and ChIP-Seq peaks ... 130
3.2.12. Calling Card peaks overlap with ERα ChIA-PET 3D interaction anchors more than ERα ChIP-Seq peaks 133
3.2.13. Significant HyPB-ERα Calling Card peaks are located at the genomic loci of putative ERα targets 137
3.2.14. HyPB-ERα fusion calling cards are enriched for the ERα binding partner motifs. 139
3.3. Discussion 142
3.3.1. Summary of results 142
3.3.2. Fusion protein design and cloning implications on function 142
3.3.3. Western blot analysis confirms full-length HyPB-ERα expression 145
3.3.4. BrokenHeart and TdTomato Functional validation confirms the function of the HyPB domain in both fusion proteins 146
3.3.5. HyPB-ERα bulk calling cards were optimised to achieve sufficient material for library preparation 146
3.3.6. Bulk Calling Card data analysis 147
4. Engineering Long-read HyPB-ERα Calling Cards using Oxford Nanopore Technology 157
4.1. Introduction 157
4.2. Results 160

4.2.1. Addition of a Kozak sequence by site-directed mutagenesis improves full-length fusion protein expression 160
4.2.2. Seeding density and puromycin concentration optimisation to improve RNA yield 163
4.2.3. Long-read ERα calling cards overlap with ChIA-PET anchors more than ChIP-Seq peaks 166
4.2.4. Signal analysis at long-read calling card insertions confirms the overlap patterns with ChIP-Seq and ChIA-PET 169
4.2.5. Untrimmed peaks are located at known ERα binding sites that overlap closely with ChIP-Seq peaks 171
4.3. Discussion 175
4.3.1. Advancing ERα Calling Card Technology through long-read sequencing... 175
4.3.2. Enhancing fusion protein expression improves viability and RNA yield 176
4.3.3. Cell passage investigation and puromycin dose optimisation 177
4.3.4. Limited insertions per cell constrained single-cell analysis, necessitating considering samples as bulk calling cards for evaluation 178
4.3.5. Long-read calling cards reproduce bulk overlap patterns: Overlap with ChIP-Seq peaks but stronger concordance with ChIA-PET anchors 179
4.3.6. ChIP-Seq signal at the centre of long-read calling card peaks reflects overlap between the datasets 180
4.3.7. Limited overlap between bulk and long-read calling cards highlights the need for reproducibility metrics 181
4.3.8. Long-read calling card peaks are located at biologically relevant ERα peaks and highlight confidence in long-read calling cards 182
5. Creating an MCF7 spheroid model for hypoxia 184
5.1. Introduction 184
5.2. Results 187
5.2.1. MCF7 cells seeded at 500 and 1000 cells/well generate spheroids 400 to 600 µm in diameter after 7 days 187
5.2.2. MCF7 spheroids grown in the same media conditions as 2D culture grow to the desired size range in 7 days 189
5.2.3. MCF7 spheroids demonstrate hypoxia development correlated with culture time and spheroid size 193
5.2.4. MCF7 spheroids show radial distribution of EF5 signal intensity across time in culture, with the highest signal in spheroid cores 194
5.2.5. High fluorescent intensity regions make up to three-quarters of MCF7 spheroids' diameter and show threefold higher intensity compared to controls by day 13 197
5.2.6. ERα protein expression is enriched at the spheroid surface and heterogeneous within hypoxic regions 201

5.2.7. Treating spheres with 100 μM of the EnAC inhibitor amiloride impairs spheroid growth and prevents hypoxic core development. 204
5.2.8. MCF7 spheroids grown in 100 μM amiloride exhibit irregular surface structure and internal branching, preventing proper spheroid formation and hypoxic core development 207
5.3. Discussion 210
5.3.1. Summary 210
5.3.2. We establish a reliable MCF7 spheroid hypoxia model as a platform for the future study of ERα in low-oxygen environments. 211
5.3.3. ERα fluorescent signal co-localises with EF5 in sequential cryosections, enabling future application with ERα calling cards..... 213
5.3.4. The application of our hypoxic MCF7 model to study hypoxia-driven ERα transcriptional targets reveals that high-dose amiloride treatment impairs spheroid formation and growth. 214
5.3.5. Future work..... 216
6. General discussion 218
6.1. Recap of thesis aims 218
6.2. Summary of main findings 220
6.3. ERα calling cards are the first tool that can be used for the simultaneous study of single-cell ERα binding and its cell fate consequences over time 221
6.4. Single-cell ERα calling cards applied to a hypoxic spheroid model can be used to explore hypoxia-driven ERα heterogeneity over time..... 223
6.5. Broader applications of ERα calling cards and the standardised MCF7 spheroid model 226
6.6. Limitations of this work 227
6.6.1. General limitations of the platforms used 227
6.6.2. Limitations specific to this thesis 228
6.7. Future directions..... 231
6.7.1. Development of multicellular tumour spheroids to more faithfully recapitulate the in vivo cell-to-cell interactions..... 231
6.7.2. Use fusion proteins with other transposase reporters to multiplex TF binding analysis 232
6.7.3. Adapting calling cards to record 3D interactions in the same way as ChIA-PET 232
6.7.4. Single-cell ERα calling cards coupled with 3D cell culture and spatial sequencing can revolutionise ERα binding heterogeneity research..... 233
7. Conclusion 233
8. Appendices 236
8.1. Chapter 3 results appendices 236

8.1.1. Fusion protein cloning primer list 237
8.1.2. Fusion protein cloning plan diagram 238
8.1.3. Full Western blots 239
8.1.4. Puromycin kill curve 241
8.1.5. High-sensitivity tape station traces 242
8.1.6. Bulk Calling Cards RNA quality 243
8.1.7. Calling Card cDNA concentrations 245
8.1.8. Replicate ROC analysis 246
8.1.9. Markdown of Calling Card Analysis Pipeline 249
8.1.10. Karyotype plots of ER calling cards and ChIP-Seq peaks 249
8.1.11. Overlap analysis hypergeometric test results 252
8.1.12. Promoter selection data 255
8.1.13. Heterogeneous TdTomato expression 256
8.2. Chapter 4 results appendices 256
8.2.1. Primer list 256
8.2.2. Single-cell calling card primer cocktail 258
8.2.3. RNA concentration and RIN for follow-up bulk ERα calling cards 259
8.2.4. Overlap analysis hypergeometric test results 264
8.2.5. qbed files with non-unique insertions show inflated overlap with ChIP-Seq and ChIA-PET peaks/anchors 266
8.2.6. Long-read calling cards are enriched for 268
8.2.7. Long read calling cards do not overlap with Bulk calling cards 269
8.3. Chapter 5 results 272
8.3.1. Amiloride inhibition of EnAC in MCF7 spheroids assay plate layout 272
8.3.2. Example images of spheroids segmented into bins by MeasureObjectIntensityDistribution in CellProfiler 272
8.3.3. CellProfiler ‘High signal’ regions in day 7 spheroids are artefacts caused by high signal contamination 274
8.3.4. ERα CellProfiler MeanFrac plots reveal that staining is not quantitatively significant 275
8.3.5. Early treatment with 100 μM amiloride affects spheroid formation from day 3 and 4 in culture 276
9. References 277

List of Figures

Figure 2.1: Ultra-low attachment 96-well plate layout for phenol Red-free DMEM spheroids 82
Figure 3.1: Restriction mapping with SalI, XhoI and HindIII on a 1% agarose gel shows successful insertion of the *ESR1* gene upstream and downstream of the *HyPB* gene in the vector plasmid 'pRM 1258 pENTR_myc-hypPBase'. 90
Figure 3.2: The predicted sequence between attB1 and attB2 gateway cloning sites in the N-and C-terminal fusion protein plasmids maps matched to Sanger-Sequencing of clones plasmids exactly, confirming successful expression plasmid generation. 92
Figure 3.3: Fusion protein expression plasmids can express the full-length protein, as shown by the presence of a band of the expected weight on an α-ER (N-terminal) and α-ER (C-terminal) Western blot. 95
Figure 3.4: Images of cells co-transfected with ERα fusion proteins and TdTomato or BrokenHeart SRTs show successful insertion of PiggyBac Transposons. 98
Figure 3.5: Summary of high-sensitivity TapeStation total RNA traces shows good recovery of RNA for C-terminal fusion and undirected HyPB bulk calling cards, whereas N-terminal fusion calling cards fail to produce sufficient quality RNA. 100
Figure 3.6: PCR amplified SRT libraries run on a 1% agarose gel show smearing of PCR products between the expected range of 500 – 5000 bp, in addition to the β-actin control. 101
Figure 3.7: HyPB-ERα calling card replicates 1 and 2 show moderate correlation at MACCs peaks. 104
Figure 3.8: Over 95% of calling card peaks called with MACCS overlap with insertions in every ESR1-HyPB replicate**.** 107
Figure 3.9: Calling card peaks are not equally distributed across chromosomes in the genome. 110
Figure 3.10: Calling card peaks large overrepresentation in chromosomes 17 and 20 compared to ChIP-Seq, but otherwise generally reflect endogenous ERα binding across the genome 111
Figure 3.11: ChIP-Seq peaks are not equally distributed across chromosomes in the genome. 112
Figure 3.12: The correlation of percentage change between expected and observed peaks located across the genome is very strong within comparisons within ChIP-Seq or calling card datasets, while the correlation between methods is moderate. 114
Figure 3.13: 83% of calling Card peaks are located within introns or other distal regions, similar to endogenous ERα binding patterns, with up to 10% reduced binding within promoter regions. 116
Figure 3.14: Around 80% of calling Card peaks are located more than 5kb from transcription start sites, with less binding within 1 kb than peaks observed in ChIP-Seq datasets. 118
Figure 3.15: Up to 70% of HyPB-ERα directed calling card peaks overlap with ER ChIP-Seq peaks from GSE109820, ENCFF365BIT and ENCFF063JMY datasets. 121
Figure 3.16: ERα ChIP-Seq signal is enriched at calling card peaks. Violin plots show mean ChIP-Seq signal intensity within 1 kb of high and low stringency calling card peaks compared to matched shuffled controls. 123
Figure 3.17: GSE109820 ChIP-Seq signal is enriched at the centre of Calling Card peaks. 124
Figure 3.18: ENCFF063JMY ChIP-Seq signal is enriched at the centre of Calling Card peaks. 125

Figure 3.19: Area Under Curve (AUC) of 0.52 and 0.58 for the high stringency peak set overlap between GSE109820 and ENCFF063JMY indicates that successful peak calling using MACCs adjusted p-value is slightly higher than random chance. 128
Figure 3.20: AUC of 0.6 and 0.58 for the low stringency peak set overlap between GSE109820 and ENCFF063JMY indicates that successful peak calling is slightly higher than random chance. 129
Figure 3.21: Calling card peaks overlap with BRD4 ChIP-Seq peaks more than ERα ChIP-Seq peaks by up to 17%. 132
Figure 3.22: The BRD4 ChIP-Seq signal is enriched at calling card peaks. Violin plots display the mean BRD4 ChIP-Seq signal intensity within 1 kb of both high and low stringency HyPB-ERα calling card peaks compared to matched shuffled controls. 133
Figure 3.23: Up to 81.2% of HyPB-ERα directed calling card peaks overlap with ERα ChIA-PET anchors compared to up to 74.1% overlap from ChIP-Seq datasets..... 136
Figure 3.24: ERα ChIA-PET signal is enriched at calling card peaks..... 137
Figure 3.25: Calling Card peaks are located at genomic loci of putative ER targets GAT3, GREB1 and FOXA1. 139
Figure 3.26: Motifs for known ER binding partners were discovered in the high stringency and low stringency peak sets with significant p-values. 141
Figure 4.1: Site-directed mutagenesis successfully mutated three base pairs preceding the HyPB-ERα fusion construct to create a Kozak Sequence. 161
Figure 4.2: The addition of the Kozak sequence visibly increases the intensity of the full-length fusion protein band and almost eliminates the second smaller band in C-terminal fusion protein-transfected lysates. 162
Figure 4.3: The highest co-transfection efficiency of calling card plasmids is achieved when seeding MCF7 cells at 0.45 x10^6 cells per well in a 6-well. 163
Figure 4.4: Both cell passage and puromycin concentration greatly affect RNA yield and total cells harvested. 165
Figure 4.5: Trimmed long-read HyPB-ERα directed calling card insertions show superior overlap (67%) with ChIA-PET anchors compared with ERα ChIP-Seq peaks, which show 3 fold fewer overlapping calling card peaks..... **Error! Bookmark not defined.**
Figure 4.6: 77.8% of untrimmed UMI-filtered long-read HyPB-ERα directed calling card peaks overlap with ChIA-PET anchors compared to 63.1% of non-filtered peaks, while overlap with ChIP-Seq peaks remains limited. 263
Figure 4.7: GSE109820 ChIP-Seq signal is not enriched at the centre of full long-read Calling Card insertions. 170
Figure 4.8: ENCFF063JMY ChIP-Seq signal is enriched at the centre of full long-read Calling Card insertions. ..... **Error! Bookmark not defined.**
Figure 4.9: Untrimmed long-read calling card peaks are located at genomic loci of putative ER targets GREB1, KCNJ3 and VMP1. 172
Figure 4.10: Some full long-read calling card peaks are located at ER binding sites and show a limited number of insertions in individual cells, in favour of many peaks from many different cells. 174
Figure 5.1: MCF7 cells seeded at 500 and 1000 cells/well grow within the ~400-600 µm range where hypoxia is reported to develop..... 189
Figure 5.2: MCF7 Spheroids initially form cell aggregates that develop into defined spheroids after 7 days. 191
Figure 5.3: Culturing MCF7 spheroids in DMEM with or without phenol red is sufficient for growing spheroids >500 µm in diameter after 7 days. 192

Figure 5.4: MCF7 spheroids show EF5 fluorescent signal once they reach ~500 µm in diameter, with signal intensity and distribution increasing as spheroids grow larger. 194
Figure 5.5: Radial distribution of EF5 signal intensity in MCF7 spheroids shows a negative correlation with distance from the centre of the spheroids from day 9 onwards 196
Figure 5.6: Gradients of EF5 intensity slopes in hypoxic spheroids, day 9 onwards, show a significant negative correlation with bin size compared to vehicle-treated spheres..... 197
Figure 5.7: The area and diameter of high signal regions of spheres progressively increase between days 7 and 13, with high signal regions spanning large areas of spheroids at day 13..... 200
Figure 5.8: Signal intensity within high signal regions increases from day 7 to 13, with some spheres showing three times the mean intensity on day 13 compared to day 7 201
Figure 5.9: ERα fluorescent signal is mainly localised to the nuclei of cells at the surface of spheroids, with sporadic levels of signal within the hypoxic region of MCF7 spheroids 203
Figure 5.10: Treatment of MCF7 spheroids with 100 µM amiloride on day 9 inhibits growth and reduces spheroid diameter compared to other conditions. 205
Figure 5.11: Early treatment of MCF7 spheroids with amiloride impairs spheroid formation at 100 µM..... 207
Figure 5.12: Early treatment with 100 µM amiloride disrupts spheroid surface morphology and promotes cell shedding. 209
Figure 5.13: Cryosectioning reveals irregular internal organisation and absence of hypoxic cores in 100 µM amiloride-treated spheroids..... 210
Figure 8.1: A diagram of N-and C-terminal cloning of ERα using NEBuilder HiFi Assembly. 233
Figure 8.2: Fusion protein expression plasmids can express the full-length protein as shown by the presence of a band of the expected weight on an anti-ERα (N-terminal) Western blot. 239
Figure 8.3: Fusion protein expression plasmids can express the full-length protein as shown by the presence of a band of the expected weight on an anti-ERα (C-terminal) Western blot. 240
Figure 8.4: All concentrations of puromycin from 2 µg/mL to 10 µg/mL killed all MCF7 cells after one week; therefore, 2 µg/mL was selected to use in calling card selections. 241
Figure 8.5: High-sensitivity TapeStation total RNA traces show good recovery of RNA for C-terminal fusion and WT HyPB bulk calling cards, whereas N-terminal fusion calling cards fail to produce sufficient quality RNA. 242
Figure 8.6: Comparative analysis of HyPB-ERα calling card replicates at MACCs peaks. 249
Figure 8.7: HyPB-ERα calling card peaks are found across all chromosomes. 250
Figure 8.8: ERα ChIP-Seq peaks are found in similar locations in the genome as calling card peaks. 251
Figure 8.9: The hCMV promoter drives the highest levels of GFP expression in MCF7 and T47D cells over time. 255
Figure 8.10: Images of cells co-transfected with C-terminal ERα fusion proteins and TdTomato SRT show heterogenous insertion of PiggyBac Transposons. 256
Figure 8.11: Without the removal of non-unique full long-read calling card insertions, overlap with ChIP-Seq peaks and ChIA-PET anchors 266
Figure 8.12: Without the removal of non-unique trimmed long-read calling card insertions, overlap with ChIP-Seq peaks and ChIA-PET anchors..... 267
Figure 8.13: Longread HyPB-ERα calling card peaks show significant enrichment of ERα ChIP-Seq, BRD4 ChIP-Seq, and ERα ChIA-PET signal..... 269

Figure 8.14: Trimmed long read HyPB-ERα directed calling card insertions only show one insertion that overlaps with bulk ERα calling card peaks. 270
Figure 8.15: Full long read HyPB-ERα directed calling card insertions show less than 1.3 % overlap with bulk ERα calling card peals. 271
Figure 8.16: The ULA 96-well plate layout for the MCF7 spheroid amiloride treatment assay. 272
Figure 8.17: Representative example images of spheroids processed using CellProfiler's MeasureObjectIntensityDistribution function for EF5 and ERα immunostaining 273
Figure 8.18: Only 5 out of the 22 MCF7 spheroids harvested on day 7 show a high signal region, but each of them is caused by debris with high fluorescence in the AF488 channel. 274
Figure 8.19: Radial distribution of ER staining in MCF7 spheroids shows surface-biased intensity but nonsignificant slopes due to high noise. 275
Figure 8.20: Treating MCF7 spheroids with a range of amiloride on day 2 in culture impairs spheroid formation above 10 μM. 276

List of Tables

Table 2.1: General PCR components 58
Table 2.2: Fusion protein cloning PCR programme 58
Table 2.3: Site-directed mutagenesis PCR programme 60
Table 2.4: A table of Transfection conditions for MCF7 cells, subsequently used for Western Blotting, Kozak-mutated plasmids 62
Table 2.5: A table of Transfection conditions used to generate transfection reagents for the optimisation of puromycin treatment in cells transfected with the N-terminal fusion protein plasmid and puromycin resistance SRT. 63
Table 2.6: Plasmids used to transfect MCF7 cells in preparation for subsequent Western Blot 63
Table 2.7: A list of primary and secondary antibodies (Ab), their concentrations, buffers used, and protocols followed after iBlot2 transfer. 65
Table 2.8: Plasmids used to generate transfection reagents for functional validation of fusion proteins using the BrokenHeart and TdTomato SRTs 67
Table 2.9: A table of Transfection conditions used to generate transfection reagents for seeding optimisation using the BrokenHeart SRT 67
Table 2.10: Plasmid used to generate transfection reagents for bulk calling card library generation to validate redirection of HyPB insertions by ERα 69
Table 2.11: A table of Transfection conditions used to generate follow-up bulk calling card libraries with greater RNA yield, particularly for N-terminal transfection cells 69
Table 2.12: Reverse transcription mix components 70
Table 2.13: Bulk Calling Cards Amplification PCR components 71
Table 2.14: Bulk Calling Cards Amplification PCR programme 71
Table 2.15: Final Bulk Calling Card Library Amplification PCR Programme 72
Table 2.16: A table of Transfection conditions used to generate single-cell long-read calling card libraries**.** 73
Table 2.17: First strand synthesis reverse transcription programme following GEM encapsulation. 74
Table 2.18: Post-GEM scRNA reverse transcription reagents 75
Table 2.19: Reverse transcription programme for transcriptomic RNA 76
Table 2.20: Follow-up up Calling card SRT amplification primer cocktail 77
Table 2.21: Calling card SRT amplification PCR mix 78
Table 2.22: Calling card SRT amplification PCR 78
Table 8.1: A list of cloning primers used to create N- and C-terminal HyPB-ERα fusion proteins. 237
Table 8.2: RNA recovery for N-terminal fusion calling cards and C-terminal replicate 4 is insufficient to carry forward. 243
Table 8.3: Final cDNA libraries fall within the expected concentration of 2-4 ng/µl within the 500-700 bp fragment range. 245
Table 8.4: Hypergeometric test statistics for overlap between HyPB-ERα calling card peaks and ERα ChIP-Seq datasets. 252
Table 8.5: Hypergeometric enrichment statistics for overlap of HyPB-ERα calling card peaks with BRD4 ChIP-Seq datasets. 253
Table 8.6: Hypergeometric test statistics for overlap between HyPB-ERα calling card peaks and ERα ChIA-PET anchors. 254

Table 8.7: A table detailing the primers used in site-directed mutagenesis and single-cell calling cards 256
Table 8.8: Calling card SRT amplification primer cocktail 258
Table 8.9: RNA concentrations of C-terminal fusions were increased following optimisations, but N-terminal fusion transfections failed to generate sufficient mRNA. 259
Table 8.10: Hypergeometric test statistics for overlap between long-read HyPB-ERα calling card peaks compared with ERα ChIA-PET anchors and ChIP-Seq datasets..... 264

Abbreviations

3D	Three-dimensional
AF	Activation Function domain
AI	Aromatase inhibitors
AKT	Protein kinase B
ANOVA	Analysis of Variance
AR	Androgen Receptor
ATAC-seq	Assay for Transposase-Accessible Chromatin with high-throughput Sequencing
AUC	Area Under the Curve
BAP1	BRCA1-associated protein
BCS	Breast-conserving surgery
BRCA1	Breast cancer type 1 susceptibility protein
BRCA2	Breast cancer type 2 susceptibility protein
Ca^{2+}	Divalent calcium ions
CCL2	C-C motif chemokine ligand 2
CCND1	Cyclin D1
CDK4/6	Cyclin-dependent kinases 4 and 6
ChIA-PET	ChIP-Seq and Chromatin Interaction Analysis by Paired-End Tag Sequencing
ChIC	Chromatin Immuno-cleavage
ChIL	Chromatin integration labelling
ChIP	Chromatin immunoprecipitation
ChIP-Seq	Chromatin immunoprecipitation and next-generation sequencing
$CoCl_2$	Cobalt Chloride
cAMP	Cyclic Adenosine Monophosphate
c-MYC	MYC proto-oncogene
D	Hinge region
DBD	DNA-binding domain
DMEM	Dulbecco's Modified Eagle Medium
DPT	Diffusion pseudotime
E2	17β-estradiol
ECM	extracellular matrix
EDHR	ERα-Dependent Hypoxic Response
EGFR	Epidermal Growth Factor Receptor
EMT	Epithelial-mesenchymal transition
ER	Estrogen receptor

EREs	Estrogen response elements
ERK	Extracellular signal-Regulated Kinase
ESR1	Estrogen receptor 1
ESR2	Estrogen receptor 2
F domain	C-terminal domain
FBS	Fetal Bovine Serum
FOXA1	Forkhead box protein A1
FPR	False Positive Rate
GATA3	GATA-binding protein 3
GEO	Gene Expression Omnibus
GFP	Green fluorescent protein
GPER	G-protein-coupled Estrogen Receptor
GREB1	Growth Regulation by Estrogen in Breast Cancer 1
GR	Glucocorticoid receptor
HER2	Human epidermal growth factor receptor 2
HIF	Hypoxia Inducible Factor
HRE	Hypoxia response element
HyPB	Hyperactive PiggyBac
IHC	Immunohistochemistry
IDR	Irreproducible Discover Rate
IL1B	Interleukin 1 beta
IL-8	Interleukin 8
IP	Immunoprecipitation
IPM	Insertions Per Million
Ki-67	Cell proliferation marker
LBD	Ligand Binding Domain
LTR	Long terminal repeat
MAPK	Mitogen-Activated Protein Kinase
MeanFrac	Mean fluorescence intensity
MBC	Metastatic breast cancer
MNase	Micrococcal nuclease
MOW	Microfluidic oscillatory washing
NANOG	Nanog homeobox protein
NCoA1	Nuclear receptor coactivator 1
NCoA2	Nuclear receptor coactivator 2
NET	Neoadjuvant Endocrine Therapy
NR	Nuclear receptor

NTD	N-terminal domain
O2	Molecular Oxygen
OCT4	Octamer-binding transcription factor 4
ODD	Oxygen degradation domain
pA-MN	Protein A and MNase
pA-Tn5	Protein A–Tn5 fusion
PB	PiggyBac
PDX	Patient-derived xenografts
PET	Paired-end tag
PGR	Progesterone receptor
PHD	Prolyl-4-hydroxylase domain proteins
PI3K	Phosphoinositide 3-kinase
PTM	Post-translational modification
Puro^{res}	Puromycin resistance
RELACS	Restriction enzyme-based labelling of chromatin in situ
RIN	RNA integrity number
S118	Serine 118
scCC	Single-Cell Calling Card
scCUT&Tag	Single-cell CUT&Tag
SERD	Selective Estrogen Receptor Degraders
SERMs	Selective Estrogen Receptor Modulators
SRC	Steroid receptor coactivator
SRT	Self-reporting transposon
TAMs	Tumour-associated macrophages
TF	Transcription factor
TFF1	Trefoil Factor 1
TNBC	Triple-Negative Breast Cancer
TNF	Tumour Necrosis Factor
TP53	Tumour protein p53
TPR	True Positive Rate
TSS	Transcription start sites
UMI	Unique Molecular Identifier
uli-	Ultra-low input
ULI-N	Ultra-low input native
VHL	Von Hippel–Lindau protein

Acknowledgments

Firstly, I would like to thank my supervisors – Dr Andrew Holding, Prof. Dave Kent and Dr Simon Baker – for providing me the opportunity to undertake this project, in addition to their guidance and support over the last four years. I am especially thankful for your efforts, Andy. Your daily presence and mentorship have been instrumental to the completion of this project; Thank you. I will always be grateful for having been able to work with such insightful and supportive mentors. I would also like to thank my TAP member, Dapne Ezer, for your kindness and generosity throughout my PhD, including inviting me to your coding club!

I am very grateful to all the help we have had from other labs, including from Dr Robi Mitra (and the Mitra group) for kindly gifting us the calling card plasmids and providing technical support for the calling card pipeline.

I would also like to thank Dr Katherine Bridge and the Bridge lab for their support and guidance, both in the transcription factor club and in setting up the Cryosectioning and EF5 hypoxia staining methods, kindly gifting us the immunostaining materials. Similarly, I would like to thank everyone in the JBU, Coverly, and Brackenbury labs for being such friendly and helpful neighbours who are always willing to help through their insight and their equipment.

I owe an incredible amount of thanks to the staff of the technology facilities, genomics and I&C teams, as your expertise and guidance have been invaluable. Sally James, Lesley Gilbert and Samantha Donniger are such incredible scientists who have brought ERα calling cards to life. I am so thankful for your guidance throughout this project. Furthermore, Grant Calder, Graeme Park, and Karen Hogg have been incredibly helpful in getting setting up on the microscopes at their amazing facility.

Of course, none of this would have been possible without the members of the Holding lab, in particular Susanna and Weiye. You welcomed me into the lab warmly, supported my learning, and provided great company during most of my time at York. Your support will always be appreciated. Thank you.

I want to thank my family for their unwavering support over the last four years and beyond. Thank you, Mum, for always being there for me and believing in me no matter what. Thank you, Dad, for constantly reminding me of my worth. Thank you, George, for constantly cheering me on and supporting me whenever I needed it. I want to thank everyone in my family for believing in me and letting me know that I am loved, no matter what happens.

I would like to express the sincerest of gratitude to my found family – The bois, and the Bio-Goths – because without your ardent support and consistent affection, this wouldn’t have been the incredible experience it has been over the last four years. Ben, Harri, Martin, and Reece, we go way back; you are my ride or dies. Rock and Stone brothers. I have been so blessed to meet such wonderful humans at York during my time here, including but not limited to: Imran, Isaac, Mags, Gemma, Maud, M, Sabina, Oliver, Charlie and Callum. I will always cherish our memories at our daily larks, playing D&D after a hard day's work in the lab and blowing off steam at the pub.

Finally, I want to give a very special thank you to the most important person in my life—the woman I can’t wait to welcome into my family next year—my wonderful fiancée, Beth. You are an invaluable source of strength and light in my life, shining brightly like a lighthouse no matter how strong the approaching storm. None of this would have been possible without you. Thank you. I love you.

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.

I acknowledge that I have used generative AI (ChatGPT5 and CoPilot) responsibly in the preparation of this thesis in line with the University of York guidance and policy. Its role was limited to supporting my learning by helping me explore complex concepts, clarify my understanding, and identify relevant resources. At all times, I critically evaluated and refined any insights gained, ensuring that the analysis, interpretation, and writing presented here are entirely my own work.

Three publications have arisen during the time of my thesis research:

Malcolm, Jodie R., **Jack Stenning**, Jacob Pope, Jakub Lukaszonek, Susanna F. Rose, Taylor E. Smith, Lesley Gilbert, et al. 2025. "**An ERα-Dependent Hypoxia Response Defines EMT-Adjacent Tumour Regions and Suppresses the pro-Survival Effects of Amiloride in Estrogen Receptor-Positive Breast Cancer.**" BioRxiv. https://doi.org/10.1101/2025.10.03.680339. (Malcolm et al., 2025)

Capatina, Alina L., Jodie R. Malcolm, **Jack Stenning**, Rachael L. Moore, Katherine S. Bridge, William J. Brackenbury, and Andrew N. Holding. 2024. "**Hypoxia-Induced Epigenetic Regulation of Breast Cancer Progression and the Tumour Microenvironment.**" Frontiers in Cell and Developmental Biology 12 (August): 1421629. (Capatina et al., 2024)

Zhao, Weiye, Susanna F. Rose, Ryan Blake, Aňze Godicelj, Amy E. Cullen, **Jack Stenning**, Lucy Beevors, et al. 2024. "**ZMIZ1 Enhances ERα-Dependent Expression of E2F2 in Breast Cancer.**" Journal of Molecular Endocrinology 73. https://doi.org/10.1530/JME-23-0133. (Zhao et al., 2024)

1. Introduction

1.1. Overview of ERα+ breast cancer

1.1.1. Introduction to Breast Cancer

As of 2025, breast cancer remains the second most diagnosed and the fourth largest cause of cancer death globally (Bray et al., 2024; Arnold et al., 2022; Sung et al., 2021). Out of the 20 million new cancer cases in 2022, 11.6% of those cases were female breast cancer, meaning that one woman was given a breast cancer diagnosis every 13 seconds that year. The global incidence of breast cancer has risen markedly since 1990, when 875,000 cases were reported, compared to 2.32 million cases in 2022 (Sha *et al.* (Eds), 2024; Bray et al., 2024). Several factors are thought to contribute to increased breast cancer incidence, including better detection methods, altered risk profiles, population growth and ageing (Xu et al., 2023b; Sha *et al.* (Eds), 2024; Arnold et al., 2022).

Breast cancer is not a single disease, but a heterogeneous group of malignancies defined by molecular and clinical differences (Guiu et al., 2012; Abubakar et al., 2019; Liu et al., 2017). Classification is achieved using immunohistochemistry (IHC) for the estrogen receptor (ER), progesterone receptor (PGR), and human epidermal growth factor receptor 2 (HER2), alongside proliferation markers such as Ki-67 (Figure 1.1)(Guiu et al., 2012; Abubakar et al., 2019; Liu et al., 2017; Parker et al., 2009; Pan et al., 2023). These markers underpin the four canonical subtypes:

Luminal A; ER+/PGR+, HER2-, and low Ki-67 (<14%). These tumours are generally less aggressive, respond well to endocrine therapy and have the most favourable prognosis within 5 years (Feeley et al., 2014; Dai et al., 2015).

Luminal B: ER+/PGR+, HER2-, but with higher Ki-67 (>14%). Luminal B cancers are more proliferative, are associated with poorer outcomes than Luminal A, and often require combined endocrine therapy and chemotherapy (Cheang et al., 2009; Maisonneuve et al., 2014).

HER2-positive: Defined by HER2 overexpression or amplification, although these tumours can be ER/PGR positive or negative (Wolff et al., 2023; Ahn et al., 2020). Therefore, these cancers may share some features of luminal cancers, but they are predominantly driven by HER2 signalling (Ahn et al., 2020; Baselga et al., 2012). HER2+ cancers historically had poor outcomes, but targeted therapies such as trastuzumab and pertuzumab have dramatically improved survival.

Triple-Negative Breast Cancer (TNBC): is defined by the lack of expression of ER, PGR, and HER2 (Chacón and Costanzo, 2010; Lehmann et al., 2011). This subtype lacks the three major therapeutic targets, making it the most challenging to treat due to the lack of targeted therapies (Zagami and Carey, 2022; Chacón and Costanzo, 2010). TNBC is often the most aggressive and is most associated with early relapse within 5 years (Dent et al., 2007; Yin et al., 2020). This subtype is typically managed with chemotherapy, through emerging

immunotherapeutic strategies (Obidiro, Battogtokh and Akala, 2023; Valencia et al., 2022; Cortes et al., 2022).

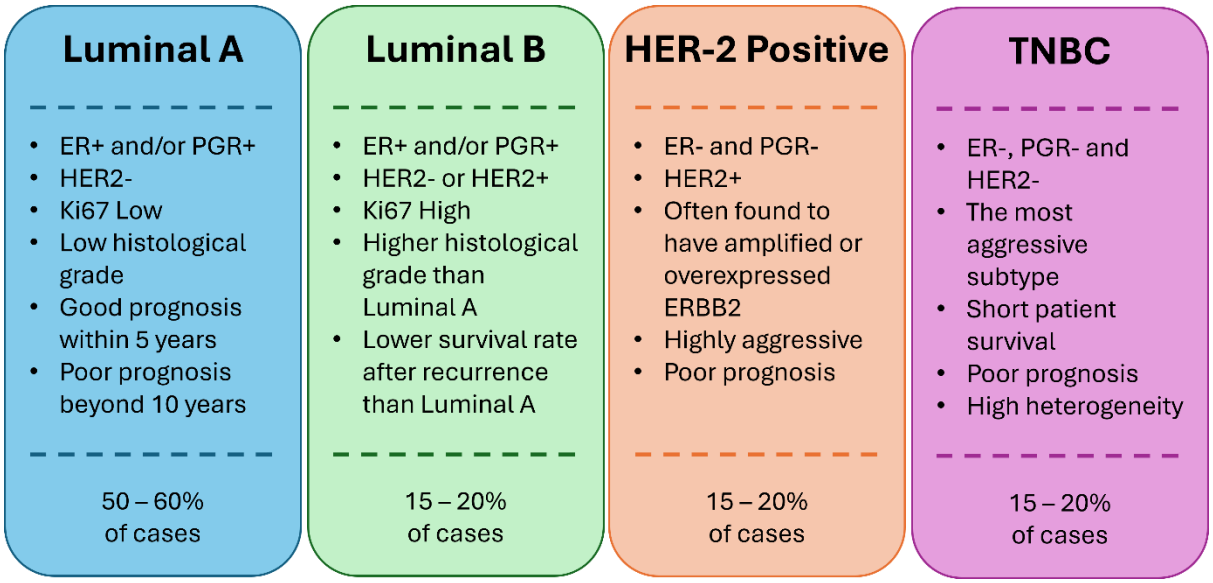


Figure 1.1 Four breast cancer subtypes are defined by the expression of ER, PR and HER2 with distinct disease progression and outcomes.

The ER, PR and HER2 define breast cancer subtype and can be further stratified based on the expression of other molecular markers such as Ki67. Most breast cancer cases are ER+, as ~80% of patients' tumours across luminal A and B patients are found to express ERα. While the aggressiveness of each subtype increases from left to right in this image within 5-10 years, beyond that, the recurrence of ER+ breast cancers are some of the most aggressive breast cancers, with very poor prognosis.

Despite ongoing efforts to refine stratification using genomic (Curtis et al., 2012; Ali et al., 2014; Lehmann et al., 2011; Burstein et al., 2015) and transcriptomic profiling (Liu et al., 2016; Wang et al., 2024; Chen et al., 2024; Turova et al., 2025), these four canonical subtypes remain the most clinically relevant categories for guiding treatment.

Approximately 70-80% of new breast cancer cases are ER+, comprising of both Luminal A and B patients (Lumachi, Santeufemia and Basso, 2015; Ozyurt and Ozpolat, 2022; Ferreira Almeida et al., 2020). Endocrine therapy is the standard of care for both, but disease recurrence and resistance to therapy remain critical challenges, particularly in Luminal B disease, where proliferation rates are higher (Burstein et al., 2014; Cheang et al., 2009; Razavi et al., 2018). The development of resistance to endocrine therapy does not occur uniformly across patients; instead, tumour heterogeneity, clonal evolution and microenvironmental signals can influence the development of resistance (Robinson et al., 2013; Razavi et al., 2018; Mavrommati et al., 2021). Around 15-20% of patients are thought to show little *de novo* response to endocrine therapy, with an additional ~30-40% of patients acquiring resistance during or after treatment, leading to ~30-50% of patients relapsing (Szostakowska et al., 2019; Clarke, Tyson and Dixon, 2015; Ottenbourgs and Van Nieuwenhuysen, 2024; Piggott et al., 2018).

The highly heterogeneous nature of breast cancer confounds a comprehensive understanding of endocrine resistance initiation, as there are many mechanisms by which heterogeneity drives reduced sensitivity to treatment (Robinson et al., 2013; Brett et al., 2021; Ahn et al., 2022; Hashimoto et al., 2023; Walens et al., 2020; Pan and Jia, 2021; Chen et al., 2023; Zhu, Jiang and Ding, 2023; Wang et al., 2014b; O'Leary et al., 2018; Pejerrey et al., 2018; Spoerke et al., 2016; Lv et al., 2025). Single-cell methods like whole-genome/exome/transcriptomic sequencing have revealed extensive intratumor heterogeneity and the presence of many distinct subpopulations and malignant subclones within primary and distant tumours (Walens et al., 2020; Pan and Jia, 2021; Chen et al., 2023; Zhu, Jiang and Ding, 2023; Wang et al., 2014b). Mutations within the ESR1 gene are a primary example. Mutations within the ESR1 locus, especially hotspot mutations such as Y537S and D538G, have been shown to reprogram cells to more aggressive phenotypes (Figure 1.2)(Li et al., 2022a, 2022b). Furthermore, primary tumours usually display a low mutation burden, but advanced metastatic breast cancers (MBC) often utilise activating mutations and mutational burden is linked to clinical outcomes (Robinson et al., 2013; Brett et al., 2021; Ahn et al., 2022; Hashimoto et al., 2023). These ESR1 mutations give a survival advantage to the cells that possess them, allowing them to outgrow other populations of cells, resulting in clonal evolution and the development of further mutations that enable resistance to therapies (O'Leary et al., 2018; Pejerrey et al., 2018; Spoerke et al., 2016; Lv et al., 2025).

MBC represents the most advanced stage of disease (Stage IV), in which tumour cells spread beyond the breast and regional lymph nodes to distant organs such as bone, lung, liver, or brain (Loibl et al., 2021; Guo et al., 2020; Park et al., 2022). Approximately 20-30% of patients with early-stage breast cancer will eventually develop metastatic disease, either through recurrence or progression, despite initial treatment success (Wang et al., 2019b; Redig and McAllister, 2013). MBC is considered incurable but treatable, with systemic therapies aimed at prolonging survival and maintaining quality of life (Riggio, Varley and Welm, 2021; Loibl et al., 2021). Five-year overall survival for metastatic breast cancer remains poor, typically 25-35% depending on subtype and treatment access (Taskindoust et al., 2021). ER+ cancers are particularly associated with late recurrence due to tumour dormancy, where disseminated cells can remain quiescent for years before reactivating (Bushnell et al., 2021; Ovadia et al., 2020). This phenomenon highlights the importance of understanding heterogeneity and microenvironmental interactions, as they contribute directly to therapy resistance and metastatic relapse (Ovadia et al., 2020; Senthebane et al., 2017; Xu et al., 2024).

1.1.2. The estrogen receptor family and their relevance to breast cancer research

Understanding the estrogen receptors is central to ER+ breast cancer progression, as subtypes contribute distinct signalling functions that influence proliferation, therapy response, and recurrence. Clarifying what is meant by "ER" is therefore essential before focusing on ERα, the primary driver of ER+ disease (Farcas et al., 2021; Fanning and Greene, 2019; Hammond et al., 2010; Robinson et al., 2013). Given that ERα is the main driver of Luminal A and B tumours (Lutz et al., 2025; Rangel and Huang, 2013; Miziak et al., 2023), while other estrogen receptors exist, its ligand-activated transcriptional activity and pro-proliferative function make it the primary focus of this thesis.

The “ER” is a term that can be used to describe three distinct receptors, the transmembrane G-protein-coupled Estrogen Receptor (GPER) and the two nuclear receptors (NR), ERα and ERβ, expressed from ESR1 and ESR2, respectively (Figure 1.2)(Walter et al., 1985; Kuiper et al., 1996; Carmeci et al., 1997; Filardo et al., 2000; Kumar et al., 2011). GPER and ERβ contribute additional signalling complexity, but the predominance of ERα in patient tumours (Xu et al., 2023a; Prossnitz and Barton, 2014; Zhou and Liu, 2020; Kinslow et al., 2022) positions it as the key ER in Luminal A/B disease.

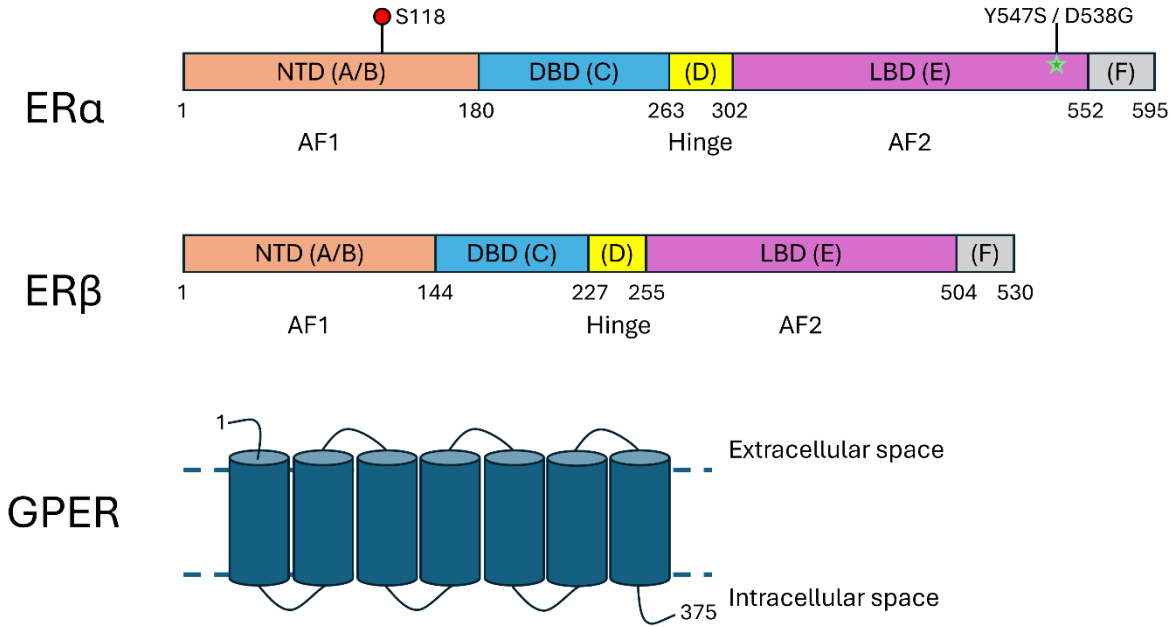


Figure 1.2 The nuclear receptors ERα and ERβ show sequence similarity in their structure and coordinate the long-term estrogen response, in contrast to the transmembrane GPER, which facilitates non-genomic signalling

ERα and ERβ show both sequence and functional similarity. The N-terminus of both proteins coordinates protein: protein interactions through the AF1 domain, as well as binding target loci through their DNA-binding domain. Both proteins encode a flexible hinge region within the middle of the polypeptide chain, followed by the ligand-binding domain, which also contains the AF2 domain. Both proteins terminate with a small c-terminal domain, also called the F domain. In contrast, GPER is a transmembrane protein that coordinates the rapid response to estrogen in preparation for the long-term signalling from the nuclear receptors, ERα and ERβ.

GPER is unlike the other two ERs, as it is a transmembrane protein comprised of 7 protein domains coupled to Gαs and Gβγ subunits (Carmeci et al., 1997; Filardo et al., 2000). GPER facilitates rapid nongenomic signalling in response to 17β-estradiol (E2), activating survival and proliferation pathways such as PI3K/AKT, MAPK/ERK, EGFR, cAMP, and Ca^{2+} mobilisation (Revankar et al., 2005; Thomas et al., 2005; Simoncini et al., 2000; Haynes et al., 2000; Filice et al., 2009; Filardo et al., 2002, 2000). PI3K/AKT promotes survival and therapy resistance (Filice et al., 2009; Haynes et al., 2000; Simoncini et al., 2000); MAPK/ERK drives proliferation (Filardo et al., 2000, 2002); EGFR enhances growth factor signalling; and cAMP/Ca^{2+} cascades add rapid nongenomic control (Revankar et al., 2005; Thomas et al.,

2005). These pathways illustrate how ER signalling integrates with broader oncogenic networks.

This integration, known as crosstalk, is particularly important in ER+ breast cancer because it complicates therapeutic targeting and contributes to heterogeneity in patient response (Osborne et al., 2005; Schiff et al., 2004; Arpino et al., 2008; Jeselsohn et al., 2015). Crosstalk between ER and these pathways underpins recurrence and resistance, making it directly relevant to the mechanistic focus of this project.

NRs are a family of transcription factors that differ from constitutively active TFs, as these molecules require binding by specific ligands for activation of transcriptional activity (Weikum, Liu and Ortlund, 2018; Sever and Glass, 2013). However, they can also be activated through post-translational modifications (PTMs), extending their regulation beyond ligand binding (Kato et al., 1995; Chen et al., 2000; Anbalagan et al., 2012). Both ERα and ERβ, like other NRs such as the glucocorticoid receptor (GR)(Stortz et al., 2017; Monje et al., 2001; Dauvois, White and Parker, 1993), remain inactive until ligand binding or PTM activation triggers a conformational change (Devin-Leclerc et al., 1998; Fliss et al., 2000; Björnström and Sjöberg, 2005). Upon activation, Hsp90 dissociates following the ER's conformational change, allowing the ER to form homo-dimers, which are then translocated to the nucleus to directly bind Estrogen response elements (EREs) at target genes (Heldring et al., 2007; Levin, 2002; Björnström and Sjöberg, 2005).

In the healthy breast, ERα promotes cell proliferation and growth, whereas ERβ has been proposed to act as a regulator of proliferation and differentiation, though this remains controversial due to limitations in reagents and antibody specificity (Lee, Kim and Choi, 2012; Lindberg et al., 2003; Dago et al., 2015; Trukhacheva et al., 2009; Yi et al., 2002).

ERα's role in driving cell proliferation positions it as the key receptor exploited during breast cancer progression. This role is supported not only by its overrepresentation in patient tumours (Paakkola et al., 2021; Lumachi, Santeufemia and Basso, 2015) but also by its localisation to the luminal compartment, and the cellular origin of Luminal A and Luminal B cancers (Yersal and Barutca, 2014; Hua et al., 2018). In contrast, ERβ is more abundant in basal and stromal compartments, and its expression can be lost in ER+ tumours (Bardin et al., 2004; Hua et al., 2018), underscoring the dominance of ERα in luminal disease. Together, these features explain why ERα is the principal driver of ER+ breast cancer and the focus of this thesis.

While multiple ER proteins exist, the ligand-activated transcriptional activity, pro-proliferative function, and luminal cell localisation of ERα underlie its role in ER+ breast cancer progression and highlight the importance of fundamental research into ERα signalling for identifying new opportunities for intervention.

1.1.3. The structure and function of ERα in breast cancer

The ERα is a 66 kDa protein that consists of 595 amino acids, which form five distinct domains with the ERα primary structure (Hewitt and Korach, 2018; Ponglikitmongkol, Green and Chambon, 1988; Habara and Shimada, 2022). The domains in ERα are called: the N-terminal domain (NTD), the DNA binding domain (DBD), the hinge region (D), the Ligand Binding

Domain (LBD) and the C-terminal domain (F) (Schwabe et al., 1993; Shiau et al., 1998; Brzozowski et al., 1997; Habara and Shimada, 2022).

The NTD is highly variable and lacks rigid secondary structure; it contains the first Activation Function (AF) domain in the ERα in AF1, a key interaction site for many of ERα's binding partners (Cagnet et al., 2018; Liu et al., 2024b; Bevan et al., 1999; Benecke, Chambon and Gronemeyer, 2000). The AF1 domain is utilised in ligand-independent ER activation, through the phosphorylation of Serine 118 (S118) by MAPK, leading to steroid receptor coactivator (SRC) binding (Figure 1.2)(Ali et al., 1993; Tremblay et al., 1999). Phosphorylated S118 then becomes a binding site critical for Pin1 binding, resulting in isomerisation of the peptidyl bond with Proline 119 that increases transcriptional activity of the ERα (Rajbhandari et al., 2012). In addition, FOXA1 and GATA3 act as pioneer factors that open chromatin and stabilise ERα binding, forming an essential part of the ERα transcriptional complex and contextualising how ERα gains access to its target sites (Cirillo et al., 2002; Asselin-Labat et al., 2007; Robinson and Carroll, 2012; Kong et al., 2011).

Genome-wide studies have shown that many ERα binding sites co-localise with motifs and occupancy of the pioneer factor FOXA1 and the lineage transcription factor GATA3 (Theodorou et al., 2013; Hurtado et al., 2011). FOXA1 facilitates chromatin accessibility and is required for efficient ERα recruitment to many regulatory regions, while GATA3 contributes to enhancer selection and maintenance of the luminal transcriptional program. Enhancers are regulatory DNA elements that contain binding sites for transcription factors and can interact with promoters via chromatin looping to increase transcription of target genes in a context-specific manner (Palstra and Grosveld, 2012; Barshad and Danko, 2025). Together, these factors form a cooperative regulatory network that shapes ERα genomic binding and estrogen-responsive gene expression in luminal breast cancer cells.

The DBD of ERα is a highly conserved region of the NR that is comprised of two zinc fingers to facilitate direct binding of ERα to EREs (Nishide et al., 2025; Steiner et al., 2022). EREs are specific DNA sequence motifs that are used to specifically recruit ERα to its target genes for regulation (Nishide et al., 2025; Steiner et al., 2022).

The hinge region is the protein domain in the centre of the ERα primary sequence that links the DBD and LBDs and facilitates transcriptional synergy between the two domains (Zwart et al., 2010). This domain also facilitates nuclear localisation on activation and assists in receptor dimerisation (Burns et al., 2014).

The activation of ERα by E2 is facilitated by the LBD of the protein, through the action of the AF1 and AF2 domains; this section of the protein coordinates critical binding interactions for controlled ERα functions (Billon-Galés et al., 2011; Benecke, Chambon and Gronemeyer, 2000).

N-terminally truncated ERα subtypes that retain the C-terminal LBD/AF-2 (for example, ERα46, ERα36 and recent LBD-only variants) can promote proliferation and endocrine resistance by shifting signalling from canonical AF-1-driven genomic programmes to AF-2-mediated or non-genomic signalling. This enables ligand-independent coactivator recruitment, activates MAPK/PI3K pathways, and produces altered responses to SERMs/SERDs (Strillacci et al., 2022; Chantalat et al., 2016). Due to the critical function of the LBD, it has been used

as a target for a class of therapies known as selective estrogen receptor modulators (SERMs) (Arao and Korach, 2021; Swaby, Sharma and Jordan, 2007).

Finally, the function of the unstructured F domain at the C-terminal end of the ERα affects the conformational changes upon ligand binding and has been shown to contribute to species-specific response to SERMs (Arao and Korach, 2021, 2018).

It is important to contextualise the function of ERα in the human breast, as although it is a major driver of disease, it relies on several other key binding partners to carry out its function (Kong et al., 2011). As we previously highlighted, the ERα is an NR that homo-and-heterodimerises upon activation, leading to trafficking into the nucleus for target gene transcription (Weikum, Liu and Ortlund, 2018; Sever and Glass, 2013).

While ERα is the main driver of cell proliferation in the breast without the function of two other proteins, GATA3 and FOXA1, the ERα cannot bind to its target genes to promote transcription (Figure 1.3) (Hurtado et al., 2011; Bernardo et al., 2010; Sadaf et al., 2023; Kong et al., 2011; Eeckhoute et al., 2007; Albergaria et al., 2009). Before ERα can bind to EREs at its target genes the pioneer factor FOXA1 must coordinate the opening of the compacted chromatin at ERα target sites (Hurtado et al., 2011; Bernardo et al., 2010; Kong et al., 2011). FOXA1 is believed to 'pre-mark' ERα binding sites, playing a critical role in determining where ERα can bind in the genome (Hurtado et al., 2011). FOXA1 is not required at all ER target genes, but loss of FOXA1 abolishes global ER binding, underscoring its critical role in maintaining ERα chromatin occupancy (Hurtado et al., 2011; Bernardo et al., 2010).

Working in conjunction with FOXA1 is GATA3, another key pioneer factor that forms part of the triumvirate of main factors in the ERα transcription complex (Kong et al., 2011; Eeckhoute et al., 2007; Albergaria et al., 2009). GATA3 acts to stabilise ERα binding as well helping to maintain accessibility of chromatin for ERα to bind to enhancer elements, cooperating with FOXA1 to act as a key facilitator of ERα target gene transcription (Theodorou et al., 2013; Kong et al., 2011).

By initiating chromatin opening, pioneer factors effectively determine the genomic regions that are accessible to ERα binding, establishing a relationship in which FOXA1 and GATA3 act upstream to specify the estrogen-responsive regulatory elements. These pioneer factors open these regulatory elements by recruiting remodelling complexes and histone-modifying enzymes to deposit enhancer-associated markers, H3K4me1 and H3K27ac (Theodorou et al., 2013; Takaku et al., 2020; Fu et al., 2019). FOXA1 and GATA3 then recruit co-activators from the SRC family of proteins, like NCoA2 and p300, to enhance ERα activity and transcription (Ahmed, Al-Saigh and Matthews, 2012; Kong et al., 2011).

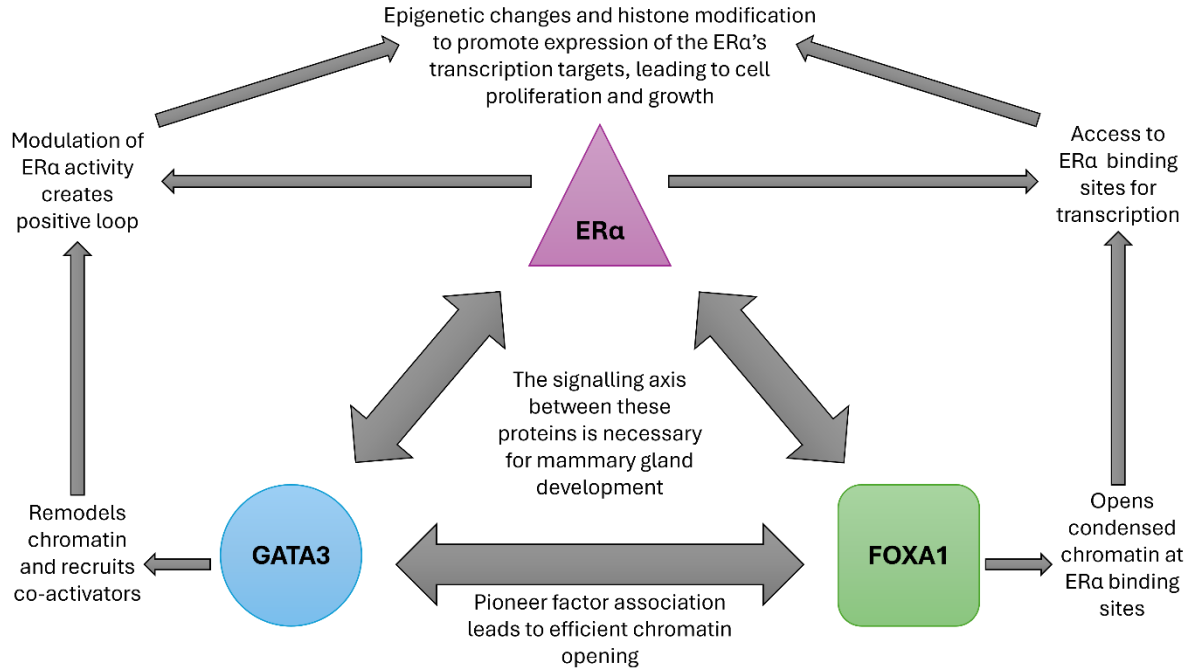


Figure 1.3 The pioneer factors FOXA1 and GATA3 are necessary to open chromatin for ERα access to bind target loci and promote cell proliferation.

FOXA1 and GATA3 act as pioneer factors for ERα, as they bind to condensed chromatin and facilitate remodelling through the recruitment of other remodellers that promote deposition of H3K4me1 and H3K27ac at enhancers that contain ERE target sites. These open chromatin markers increase accessibility for ERα to bind target sites, which is then positively reinforced by GATA3 recruitment of co-activators. Together, these three proteins work to open chromatin, facilitate ERα binding and reinforce epigenetic signals to maintain ERα signalling.

Genome-wide studies show that ERα binding occurs predominantly within accessible chromatin regions (Hurtado et al., 2011; Fiorito, Katika and Hurtado, 2013) and is strongly enriched at distal enhancer elements rather than promoters (Lin et al., 2007). Following estrogen stimulation, ERα recruitment promotes further chromatin remodelling and enhancer activation to regulate estrogen-responsive gene expression.

The structure of the ERα enables sensitive control of transcriptional output through the recruitment of transcription modulators via AF1/2 domains (Benecke, Chambon and Gronemeyer, 2000; Kobayashi et al., 2000; Métivier et al., 2001; Zwart et al., 2010; Rajbhandari et al., 2012). Both NCoA2 and p300 have been shown to simultaneously bind to the AF1 and AF2 domains of the ERα, resulting in increased target gene transcription and cell proliferation (Benecke, Chambon and Gronemeyer, 2000; Kobayashi et al., 2000; Métivier et al., 2001). The AF1 and AF2 domains of the ERα are also able to make direct physical contact with each other, enhancing the transcriptional activity of both domains through cooperativity with the hinge domain (Métivier et al., 2001; Zwart et al., 2010).

Transcriptional cooperativity between ERα, FOXA1, GATA3 and other enhancers results in the transcription of target genes that promote cell proliferation, like GREB1 (Sun, Nawaz and

Slingerland, 2007; Haines et al., 2020) , TFF1 (Baron et al., 2007; Sun et al., 2005) and CCND1 (Cicatiello et al., 2004; Jeffreys et al., 2022).

Growth Regulation by Estrogen in Breast Cancer 1 (GREB1) is directly regulated by ERα through an ERE at the promoter of the gene (Sun, Nawaz and Slingerland, 2007). Upon expression, GREB1 regulates the PI3K/AKT signalling pathway (Haines et al., 2020), promoting E2-driven cell proliferation. This activity is similar to that of GPER's activation of PI3K/AKT signalling, demonstrating cooperativity between the long-term and short-term E2-mediated cell response to maintain cell proliferation (Filice et al., 2009; Haynes et al., 2000; Simoncini et al., 2000). Importantly, GREB1 also functions as an ERα co-activator, reinforcing ERα signalling loops. This dual role highlights GREB1 as both a transcriptional output and a functional amplifier of ERα signalling.

The Trefoil Factor 1 (TFF1) gene also contains EREs within its promoter, and its expression is further enhanced by stabilisation from FOXA1 and GATA3 (Baron et al., 2007) . Although TFF1 is not universally expressed across all ER+ models (e.g. low in T47D cells), it remains a historically important ER target gene, extensively studied since its early discovery, and continues to serve as a useful marker of ER activity (Amiry et al., 2009; Yi et al., 2020; Spadazzi et al., 2021).

Finally, CCND1, which expresses cyclin D1, is also upregulated by ERα through an ERE and is functionally crucial to the G1 to S phase transition (Cicatiello et al., 2004). Overexpression of CCND1 is common in ER+ cancers and is associated with poor prognosis in patients, which is why ER+ cancers are clinically targeted with CDK4/6 inhibitors (Jeffreys et al., 2022).
This published work demonstrates how the ERα's structure enables the binding of key transcriptional regulators to induce transcriptional synergy across the protein to drive target gene expression. However, ER does not act alone in oncogenesis; rather, its transcriptional programmes cooperate with broader genetic and epigenetic alterations (e.g. PIK3CA, GATA3, TP53 mutations) to drive malignant progression, as highlighted in recent genomic studies (Nik-Zainal et al., 2016; Cancer Genome Atlas Network, 2012).

Interestingly, there is a large body of evidence that demonstrates that while ERα is one of the best lineage markers for luminal cancers, most tumour-initiating events do not involve ERα mutation (Banerji et al., 2012; Lopez et al., 2019; Jiang et al., 2014; Ciriello et al., 2013). Typically, mutations in genes like PIK3CA (Lopez et al., 2019), GATA3 (Li et al., 2022a; Banerji et al., 2012) and TP53 (Ciriello et al., 2013), precede ESR1 mutations in the progression of breast tumours (Li et al., 2022; Banerji et al., 2012). Inherited germline mutations can also contribute to an individual's risk of developing breast cancer; BRCA1 and 2 have been shown to predispose individuals to developing breast cancer (Miki et al., 1994; O'Donovan and Livingston, 2010; Somasundaram, 2010). The two genes are associated with different subtypes of breast cancer, with BRCA1 generally showing as mutated in TNBC, whereas BRCA2 mutations generate a heterogeneous mix but are generally ERα+ (Fountzilas et al., 2020; Lakhani et al., 1998; Da Silva and Lakhani, 2010). Single-cell studies of patient tumours find that BRCA status, along with a variety of other factors, significantly affect the composition of breast tissue, demonstrating how mutations in genes other than ERα contribute to tumour initiation (Reed et al., 2024).

Taken together, these findings support the view that the major breast cancer subtypes arise from distinct genetic and epigenetic alterations, and that clinically observed plasticity and heterogeneity occur primarily within, rather than across, these subtypes (Banerji et al., 2012; Lopez et al., 2019; Jiang et al., 2014; Ciriello et al., 2013; Caswell-Jin, Lorenz and Curtis, 2021; Turashvili and Brogi, 2017; Guo et al., 2023). This sustains the relevance of ERα as a lineage marker while highlighting the complexity of heterogeneity within ERα+ tumours.

1.2. Treatment of breast cancer with a focus on ERα+ disease

1.2.1. Treatment of ERα+ cancer specifically and how it differs from other subtypes

NET reduces ERα activity to shrink tumours before surgery, producing similar clinical responses to chemotherapy but with reduced toxicity, making it an attractive option for patients (Zhang et al., 2021b; Bischoff, Espié and Petit, 2024; Tremont, Lu and Cole, 2017).

Hormone-responsive breast cancers are generally treated with Neoadjuvant Endocrine Therapy (NET) before surgery and radiotherapy to remove the initial tumour, followed by adjuvant therapy (Park-Simon et al., 2023; Johnson, Conant and Soo, 2021; Onitilo et al., 2009). NET reduces ERα activity to shrink tumours before surgery, producing similar clinical responses to chemotherapy but with reduced toxicity, making it an attractive option for patients (Spring et al., 2016; Barroso-Sousa et al., 2016; Martínez-Pérez et al., 2023).

One of the advantages of neoadjuvant therapy before tumour resection is that it increases patient eligibility for breast-conserving surgery (BCS) instead of mastectomy (Christiansen et al., 2022; Zaborowski et al., 2023; De La Cruz et al., 2016). While clinical outcomes support BCS as having improved OS for breast cancer patients, oncoplastic approaches that improve cosmetic results have shown no compromise in survival and low complication rates (Christiansen et al., 2022; Li et al., 2024).

Following surgery, adjuvant endocrine therapy for ~5 years remains the backbone of ERα+ treatment, aiming to prevent recurrence by disrupting ERα-driven transcriptional programmes (Park-Simon et al., 2023; Ditsch et al., 2022; Schiavon and Smith, 2014).

Several agents are used in anti-ERα adjuvant therapy to reduce ERα signalling through various mechanisms (Figure 1.4)(Hanker, Sudhan and Arteaga, 2020). SERMS and Selective Estrogen Receptor Degraders (SERDs) are molecules that target the LBD of the ER to either sequester the protein and prevent gene transcription, or tag the protein for degradation (Wolff and Davidson, 2001; Craig Jordan et al., 2014; Johnson, 2006; Sharaf et al., 2024; Bross et al., 2002; Boér, 2017).

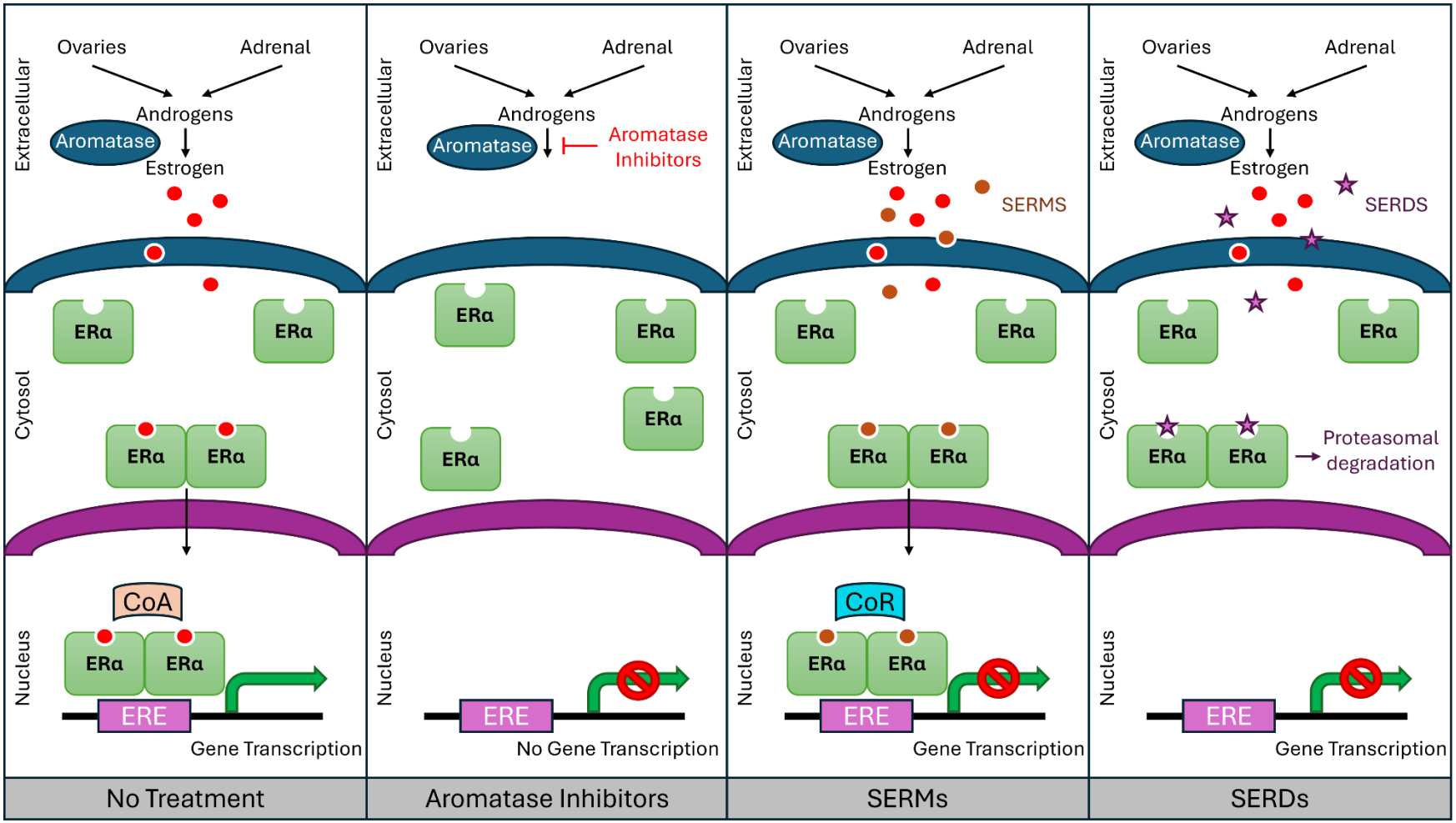


Figure 1.4 The main therapies given to ERα+ breast cancer patients are aromatase inhibitors, SERMs and SERDs, that all aim to reduce ERα signalling to kill the cancer cells.

Aromatase inhibitors aim to prevent the synthesis of estradiol from androgens, to starve ERα of its agonist, preventing transcription. SERMs and SERDs on the other hand bind directly to the LBD of ERα and serve to modulate ERα activity through the recruitment of co-repressors or induce proteasomal degradation respectively

The primary SERM used in therapy is Tamoxifen, a molecule that revolutionised the treatment of breast cancer (Fowble et al., 1996; Davies et al., 2013; Johansson et al., 2025). Tamoxifen was first identified as a potential contraceptive in 1971 (Klopper and Hall, 1971) and was approved for use in metastatic breast cancer in 1977 (Wolff and Davidson, 2001). Eventually, tamoxifen was increasingly used as an adjuvant, as the invasive breast cancer risk of patients receiving tamoxifen was reduced by 49% in high-risk women(Craig Jordan et al., 2014; Lippman and Brown, 1999). SERMs act as selective modulators, functioning as antagonists in breast tissue but partial agonists in other tissues. Second-generation SERMs such as Raloxifene were developed to refine this balance, demonstrating that modulation of ERα activity can produce significant clinical benefit (Craig Jordan et al., 2014).

SERDs such as Fulvestrant represent a distinct design, promoting ERα degradation rather than modulation. Their use in resistant disease highlights the unmet need for a deeper mechanistic understanding of ERα biology (Wang and Tang, 2022; Sharaf et al., 2024). The first approved SERD for the treatment of advanced breast cancer was Fulvestrant in 2002, 15 years later, it was approved as the first-line therapy for advanced ERα+ tumours (Bross et al., 2002, 2003; Boér, 2017).

Another key agent in anti-ERα endocrine therapy is aromatase inhibitors (AI) like aminoglutethimide that inhibit aromatase, the enzyme that converts androgens into estrogens in peripheral tissues (Santen et al., 1977; Asbury et al., 1981; Kaye et al., 1981; McArthur and Morris, 2010; Schneider et al., 2011). Aminoglutethimide was the first AI developed for use in the treatment of metastatic breast cancer in the late 1970s (Santen et al., 1977; Smith et al., 1978; Asbury et al., 1981). Today, third-generation AIs such as anastrozole (Baum et al., 2002; Cuzick et al., 2010), letrozole (Breast International Group (BIG) 1-98 Collaborative Group et al., 2005; Regan et al., 2011) and exemestane (Coombes et al., 2004; Morden et al., 2017) are widely used in ERα+ adjuvant therapy, reducing estrogen synthesis and thereby limiting ERα activation.

Certain targeted agents can be added to endocrine therapy regimens depending on tumour expression profiles, such as CDK4/6 and HER2 inhibitors (Sherr, Beach and Shapiro, 2016; Niu, Xu and Sun, 2019; Dziadkowiec et al., 2016; Helleday, 2011). CDK4/6 inhibitors target Cyclin D1-driven cell cycle progression in ERα+ cancers (Shanabag et al., 2025; Pernas et al., 2018; Goetz et al., 2017; Neven et al., 2023; O'Leary et al., 2018), while HER2-targeting compounds such as trastuzumab address HER2-driven signalling (Romond et al., 2005; Lin et al., 2023). BRCA1/2 mutations also influence ERα+ tumour biology and therapeutic response, underscoring the genetic context of treatment (Helleday, 2011; Dziadkowiec et al., 2016).

The type of main adjuvant agents used depends on several factors, one being whether the patient is pre-or-post menopausal (Partridge, 2015; Santen et al., 2017; Early Breast Cancer Trialists' Collaborative Group (EBCTCG) et al., 2012). Post-menopausal women no longer produce ovarian Estrogen, meaning the peripheral conversion by aromatase is a major source of E2 for tumours, whereas pre-menopausal women still produce ovarian estrogen (Lee et al., 2025; Sella et al., 2022; Burstein et al., 2014). This difference in circulating E2 results in pre-menopausal women typically receiving tamoxifen for five years, with AIs reserved for high-risk cases, while post-menopausal women are preferentially treated with AIs (Lee et al., 2025; Sella et al., 2022; Burstein et al., 2014).

Together, these therapies illustrate how ERα signalling is targeted across the treatment pipeline. Despite advances, resistance remains a significant challenge in ERα-positive disease, creating an urgent need for a deeper understanding of ERα biology.

1.2.2. Prognosis in ERα-positive breast cancer

The prognosis of patients within 5 years of receiving anti-ERα endocrine therapy is the highest among breast cancer subtypes, with the lowest levels of recurrence compared to HER2+ and TNBC (He et al., 2019; Palmi et al., 2025; Lee et al., 2023c; Romond et al., 2005; Pan et al., 2017).

However, there is a steady, lifelong increase in risk of recurrence every year beyond 5 years for ERα+ breast cancer, significantly worsening prognosis for patients from year 5 to 20 after diagnosis (Pan et al., 2017; Sparano et al., 2024). This late recurrence is a defining feature of ERα+ disease and reflects the biology of tumour dormancy, where disseminated cells can remain quiescent for years before reactivating (Pan et al., 2017; Zhang et al., 2013b).

Prognosis is influenced by predictive markers rather than direct causal factors. These include receptor discordance between the primary tumour and distant metastases, loss of ERα/PGR expression (Shiino et al., 2022; Grinda et al., 2021), and tumour grade (Elston and Ellis, 1991). High-grade tumours are defined by poor differentiation and high proliferation rates, and are consistently associated with worse long-term outcomes (Elston and Ellis, 1991; de Azambuja et al., 2007; Inwald et al., 2013).

Progression occurs through mechanisms such as ESR1 mutations, which generate activating variants of ERα that reduce sensitivity to endocrine therapy (Robinson et al., 2013; Jeselsohn et al., 2015; Chandarlapaty et al., 2016). Clonal evolution further compounds this process, as resistant subclones expand under therapeutic pressure, producing more aggressive disease phenotypes (Osborne and Schiff, 2011; Hanker, Sudhan and Arteaga, 2020; Lee et al., 2023d; Shiino et al., 2022).

Some subgroups of ERα+ cancers are more susceptible to aggressive late recurrence than others. Patients with Luminal B and node-positive, high-grade ERα+ tumours have a larger risk than other patients (Lee et al., 2023c; Pan et al., 2017). Node positivity reflects disease stage at diagnosis rather than a separate subgroup, but remains a strong predictor of recurrence risk (Carter, Allen and Henson, 1989; Galea et al., 1992; Teichgraeber, Guirguis and Whitman, 2021). Patients with these features are therefore offered intensified adjuvant strategies, such as CDK4/6 inhibitors like abemaciclib, to mitigate their increased risk(Johnston et al., 2020; Pan et al., 2017).

Inherited germline mutations also influence prognosis. BRCA2 mutations, for example, are frequently associated with ERα+ tumours and contribute to heterogeneity in clinical outcomes, whereas BRCA1 mutations are more commonly linked to TNBC (Helleday, 2011; Dziadkowiec et al., 2016). These genetic contexts highlight how prognosis is shaped by multiple layers of tumour biology beyond ERα status alone.

Taken together, prognosis in ERα+ breast cancer is favourable in the short term but complicated by late recurrence, receptor loss, mutational evolution, and genetic background. Post-relapse treatment options remain an unmet need across all breast cancer subtypes but are particularly pressing in ERα+ disease due to the high incidence of late recurrence and the limited durability of current endocrine therapies.

1.2.3. Resistance to endocrine therapy

As previously mentioned, the loss of ERα results in resistance to endocrine therapy (Shiino et al., 2022; Grinda et al., 2021), but this is only one of many mechanisms that drive ~20 to 40% of patients receiving endocrine therapy to develop resistance (Zhou et al., 2019; Gao et al., 2024; Pan et al., 2017; Pedersen et al., 2022). Because all endocrine therapies, including AIs, are targeted agents against ER, aberrations in its regulation represent a clear pathway to subverting sensitivity to treatment (Jeselsohn et al., 2015; Fanning et al., 2016; Brett et al., 2021; Nagy and Jeselsohn, 2022; Yates et al., 2024).

Dozens of recurrent mutations in the LBD have been identified, most clustered at residues Y537 and D538, which together account for the majority of metastatic ERα+ tumours (Jeselsohn et al., 2015; Fanning et al., 2016). These mutations stabilise the active conformation of the protein in the absence of E2, resulting in a constitutive activity. These mutations are almost always found in metastatic samples and demonstrate how these mutations contribute to aggressive ERα recurrence (Brett et al., 2021). Patients with these mutations benefit from treatment with newer SERDs to overcome the constitutive activation (Bidard et al., 2022; Brett et al., 2021; Jeselsohn et al., 2015; Tolaney et al., 2023).

Fusion mutations involving the NTD occur in ~1% of ERα+ patients, producing constitutively active fusion proteins (Nagy and Jeselsohn, 2022; Yates et al., 2024). Examples such as ESR1-YAP1 and ESR1-DAB2 lack the LBD, removing the binding site for SERMs/SERDs and driving uncontrolled ERα transcription (Lei et al., 2018; Hartmaier et al., 2018; Li et al., 2013b). Genomic instability caused by mutations in DNA repair machinery, such as BRCA1/2, facilitates fusion events and rearrangements that generate ERα variants insensitive to endocrine therapy (Latysheva and Babu, 2019; Liu et al., 2025). While BRCA1 mutations are more often linked to TNBC, BRCA2 mutations contribute to heterogeneity in ERα+ tumours, underscoring how genetic context shapes resistance mechanisms.

Another, less common mechanism for endocrine therapy resistance is the complete loss of ERα and a switch to alternative survival pathways, resulting in insensitivity to endocrine therapy (Hanker, Sudhan and Arteaga, 2020; Shiino et al., 2016; Belachew and Sewasew, 2021). This mechanism represents a minority of cases, with around 10% of patients with resistance to therapy showing a complete loss of ERα (Hanker, Sudhan and Arteaga, 2020; Shiino et al., 2016).

The key co-regulatory proteins of ERα also contribute to endocrine therapy resistance in patients, particularly through the SRC family of proteins (Browne et al., 2018; Walsh et al., 2012; Zhou et al., 2020; Giuliano et al., 2011; Hou et al., 2020; Yi et al., 2013). SRC-1 or nuclear receptor coactivator 1 (NCoA1) promotes resistance by activating SIRT1-mediated deacetylation of ERα, stabilising its activity under therapy pressure (Browne et al., 2018;

Walsh et al., 2012; Zhou et al., 2020). The overexpression of SRC-3 is also associated with worse outcomes for patients, driven by destabilised p53 activity through TRAF4 upregulation (Yi et al., 2013; Giuliano et al., 2011; Hou et al., 2020).

Endocrine therapy resistance is also induced through mutation in the peripheral signalling pathways that complement the ERα transcriptional programme (Skolariki et al., 2022; Dong et al., 2021; André et al., 2019; Hanker, Sudhan and Arteaga, 2020). In particular, aberrant activation of the PI3K/AKT signalling axis has been associated with reduced dependence on the ERα signalling pathway, enabling endocrine therapy survival (Skolariki et al., 2022; Dong et al., 2021). Amplification of receptor tyrosine kinases such as HER2 and EGFR or their downstream targets (e.g. MAPK) also stimulates E2-independent activation of ERα through phosphorylation (Hanker, Sudhan and Arteaga, 2020; Presti and Quaquarini, 2019). The overexpression of ERα targets like CyclinD is also associated with resistance, as the potent growth signal from this key cell cycle regulator is strong enough to promote E2-independent proliferation (Hanker, Sudhan and Arteaga, 2020).

FOXA1, while a pioneer factor, also acts as a driver of resistance when overexpressed. It reprograms the ERα cistrome, leading to IL-8 upregulation and activation of hypoxic signalling via HIF-2α, reinforcing resistance mechanisms (Fu et al., 2019, 2016).

Taken together, these mechanisms illustrate the diverse ways in which ERα+ tumours can evade endocrine therapy. Resistance may arise through mutations, fusions, co-regulator changes, or rewiring of signalling pathways, each producing distinct resistant phenotypes. This diversity is precisely what makes endocrine resistance so challenging to treat: no single mechanism dominates, and patients often harbour multiple overlapping drivers. These observations naturally lead to the concept of heterogeneity, where resistance is not uniform but emerges through clonal competition and expansion within tumours.

1.2.4. Resistance and heterogeneity

The multitude of resistance mechanisms in ERα+ breast cancer underscores a deeper problem: heterogeneity. Each mutation, fusion, or signalling adaptation generates distinct subclones within a tumour, creating a patchwork of resistant populations (Saha et al., 2018; Lindström et al., 2018; Beumers et al., 2023; O'Leary et al., 2018). Rather than a single pathway of escape, resistance emerges as a heterogeneous process, confounding precision treatment and making generic therapeutic strategies inadequate. Understanding how resistance mechanisms drive clonal expansion is therefore central to explaining why ER+ tumours relapse despite prolonged endocrine therapy.

Heterogeneity is a key confounding factor in ERα+ breast cancer treatment. Precision medicine relies on predictable tumour biology, yet intra-tumour diversity means that therapies effective against one clone may fail against another. This complexity is compounded by the tumour microenvironment, which actively reinforces heterogeneity (Li et al., 2020a; Joffroy et al., 2010; Murray et al., 2015; Diaz Bessone et al., 2019). Tumour-associated macrophages (TAMs) secrete cytokines such as CCL2 (Li et al., 2020a), TNF, and IL1B (Murray et al., 2015; Diaz Bessone et al., 2019), stimulating PI3K/AKT signalling and aromatase activity, thereby fuelling resistant clones with excess estrogen.

Among microenvironmental factors, hypoxia plays a particularly large role: by reshaping the epigenome, hypoxia promotes persistent changes in enhancer accessibility and transcriptional programmes, driving the expansion of resistant epigenetic clones (Capatina et al., 2024; Batie et al., 2022; Ma et al., 2022; Salminen, Kaarniranta and Kauppinen, 2016; Yang, Harris and Davidoff, 2018).

Together, genetic, epigenetic, and microenvironmental heterogeneity converge to undermine the efficacy of endocrine therapy. Given its profound impact on the epigenome and its persistence across tumour evolution, hypoxia provides a powerful model to investigate how microenvironmental pressures drive epigenetic clone expansion in ERα+ breast cancer. In the next section, we will examine hypoxia in detail as a lens through which to understand the interplay between resistance, heterogeneity, and tumour progression.

1.3. Hypoxia and microenvironment in breast cancer

1.3.1. Primary mechanisms of the cellular hypoxic response

Hypoxia is defined as a low oxygen tension (pO_2) in the tumour compared to the surrounding tissue, measuring between 2.5 and 10 mmHg pO_2, or 0.2 to 1% O_2 (Bredell et al., 2016; Telarovic, Wenger and Pruschy, 2021; Bai et al., 2022). Tumours experience hypoxia for several reasons; it can be due to improper vasculature formation resulting in impaired oxygen perfusion, or due to increased metabolic demand due to uncontrolled proliferation (Liao et al., 2023; Sebestyén et al., 2021). Intermittent hypoxia is generated through repeated hypoxia and re-oxygenation cycles and has been shown to display unique effects on cell behaviour distinct from continuous hypoxia (Liu et al., 2022b; Chen et al., 2018).

There are many consequences of hypoxia on cell fate, but the classical driver of the cellular hypoxic response is the stabilisation of Hypoxia Inducible Factor (HIF)-1α and HIF-2α in low oxygen (Wang and Semenza, 1995; Semenza, 2004; Tirpe et al., 2019; Chua et al., 2010). The HIF-α proteins are transcription factors that act as the master regulators of the cellular response to low oxygen (Liu, Semenza and Zhang, 2015; Kozal and Krześlak, 2022).

Both HIF-1α and HIF-2α can induce target gene expression in response to changing oxygen concentrations through the PHD and VHL proteins, coordinating their destruction (Yu et al., 2001; Haase, 2009; Fong and Takeda, 2008). During normoxia, the prolyl-4-hydroxylase domain proteins (PHD) hydroxylate conserved proline residues in the oxygen degradation domain (ODD) of HIF-α (Fong and Takeda, 2008). This residue is then recognised by the VHL domain of an E3 ligase that coordinates the degradation of HIF-α subunits (Yu et al., 2001; Haase, 2009). PHD hydroxylation of HIF-α requires O_2 as a substrate, meaning that in low oxygen conditions, PHD hydroxylation fails and HIF-α proteins evade degradation (Yu et al., 2001; Haase, 2009; Fong and Takeda, 2008).

HIF-1α and HIF-2α are two subunits of the HIF proteins that coordinate the short and long-term hypoxic response, respectively, through their dimerisation with HIF-1β (Loboda, Jozkowicz and Dulak, 2010; Shi and Gilkes, 2025; Jaśkiewicz et al., 2022a, 2022b; Koh and

Powis, 2012). Studies have shown HIF-1α peaks early at ~4 hours into hypoxia and expression drops after around 8 hours (Jaśkiewicz et al., 2022a), at which point HIF-2α becomes more prominent and remains elevated in prolonged hypoxia. Both HIF-α proteins dimerise with the constitutively expressed HIF-1β and are then translocated into the nucleus, where they coordinate the transcription of genes through binding their hypoxia response element (HRE) binding site (Kaluz, Kaluzová and Stanbridge, 2008; Dengler, Galbraith and Espinosa, 2014). HIF-2α expression is restricted to certain cell types and displays cell-specific functions, but generally upregulates angiogenesis, erythropoiesis and stem-cell-like transcription programmes (Hoefflin et al., 2020; Albadari, Deng and Li, 2019). Whereas HIF-1α is broadly expressed and preferentially upregulates genes driving glycolysis, glucose transport and metabolic programming to shift cellular processes toward functioning in low oxygen environments (Hoefflin et al., 2020; Albadari, Deng and Li, 2019).

1.3.2. Cell fate consequences of HIF-α expression

HIF-1α is generally better studied across cancer research than HIF-2α for several reasons, from its ubiquitous expression to its alignment with classical cancer hallmarks (Masoud and Li, 2015; Poon, Harris and Ashcroft, 2009; Shi and Gilkes, 2025; Kao et al., 2023). By upregulating glycolytic enzymes and glucose transporters, HIF-1α supports the Warburg effect, increasing glucose metabolism and providing a rapid growth advantage to tumour cells (Cairns, Harris and Mak, 2011; Poon, Harris and Ashcroft, 2009). This makes HIF-1α a potent regulator of the early hypoxic response, defined as the initial stage when oxygen levels first fall below normoxic thresholds and positions it as a short-term mediator of tumour adaptation to hypoxia (Shi and Gilkes, 2025; Kao et al., 2023). By sustaining glycolysis and energy production, HIF-1α-expressing cells gain a proliferative advantage over other populations, intensifying competition in heterogeneous tumours (Semenza, 2012; Muz et al., 2015).

Hypoxia has long been established as a driver of tumour progression through its role in stimulating angiogenesis, metastasis and therapy resistance, leading to more aggressive disease (Vaupel, 2004; Muz et al., 2015; Semenza, 2012; Wang et al., 1995). A fundamental consequence of hypoxia is impaired drug delivery: when vascularisation is inadequate, oxygen tension is low and bloodborne therapies cannot reach tumour cells effectively (Bridges and Harris, 2015; Mou et al., 2024). This physical barrier reinforces the molecular mechanisms of resistance described below.

Specifically, HIF-α expression predicts poor response to endocrine therapy in ERα+ patients by promoting highly plastic, stem-like cell populations (Generali et al., 2006; Yan et al., 2018; Ryu, Park and Lee, 2011; Heddleston et al., 2010; Padró et al., 2017). These stem-like states increase intra-tumour heterogeneity and directly contribute to endocrine resistance. Hypoxia employs multiple routes to promote chemoresistance, including increasing cell robustness by altering the cell cycle, promoting anti-apoptotic survival signals and expanding stem-like populations (Yan et al., 2018; Muz et al., 2015). HIF-1α also downregulates ESR1 transcription, reducing ERα expression and sensitivity to anti-ERα therapies, while simultaneously activating survival programmes that sustain tumour growth(Ryu, Park and Lee, 2011; Generali et al., 2006).

These findings highlight hypoxia as both a physical and molecular driver of therapy resistance. By stabilising HIF-α proteins, tumours gain access to transcriptional programmes that promote stemlike plasticity, metabolic reprogramming, and reduced drug delivery. This convergence of mechanisms explains why hypoxia is such a powerful amplifier of heterogeneity in ERα+ breast cancer, and why it represents a critical barrier to effective endocrine therapy.

1.3.3. Hypoxia-driven heterogeneity's effect on breast tumorigenesis

Hypoxia induces both spatial and temporal heterogeneity in breast cancer through chronic vs intermittent hypoxia, which puts selective pressure on cells and can push cells into a plastic reprogramming (Yan et al., 2018; Heddleston et al., 2010; Yun and Lin, 2014; Chen et al., 2018). Plasticity refers to the ability of cancer cells to dynamically shift their transcriptional and phenotypic states in response to environmental stress, thereby enabling adaptation and survival. The sustained action of HIF-2α promotes the activation of Notch and Wnt signalling, key stem-cell master regulators that result in increased expression of the stem markers OCT4, NANOG and c-MYC (Yan et al., 2018).

In contrast to the general ERα downregulation carried out by HIF-1α-expressing cells, there are specific and direct interactions between HIF-1α and ERα that contribute to disease progression (Ryu, Park and Lee, 2011; Generali et al., 2006; Padró et al., 2017; Min et al., 2022; Cho et al., 2005). The proposed mechanisms surrounding the interaction are mixed, with some publications reporting it as promoting the degradation of ERα (Padró et al., 2017) and others claiming it promotes ligand-independent activation (Yang et al., 2015; Min et al., 2022).

What is clear, however, is that HIF-1α activity alters the ERα transcriptional programme, changing its binding localisation from canonical targets, which is then reinforced through chromatin rearrangements from pioneer factors (Min et al., 2022; Ma et al., 2022; Capatina et al., 2024; Padró et al., 2017; Malcolm et al., 2025). ERα has been found to localise with HIF at promoters with HREs (Min et al., 2022) and data from our lab shows that during hypoxia, ERα coordinates a unique transcriptional profile from normoxia (Malcolm et al., 2025).

Collectively, these findings suggest mechanisms through which hypoxia may promote cellular heterogeneity in breast cancer, by generating populations of cells with enhanced survivability and reduced reliance on ERα signalling (Padró et al., 2017; Jehanno et al., 2022; Xie et al., 2016; Zhang et al., 2016). Evidence from xenograft models supports this, showing that hypoxic tumour regions display distinct transcriptomic profiles enriched for stemlike features, including ERα downregulation and tumorigenic reprogramming (Kim et al., 2018; Jehanno et al., 2022) Hypoxia can promote spatial heterogeneity by creating pockets of low ERα-expressing cells at the tumour core, which survive and expand under selective pressure. Evidence from experimental models supports this: repeated cycles of hypoxia and reoxygenation in metastatic breast cancer cell lines select for a stable, stemlike, EMT-positive subpopulation that is highly metastatic in xenograft mouse models (Louie et al., 2010). These findings demonstrate that hypoxic niches do not simply reduce ERα signalling but actively reprogramme cells toward stem-like states with enhanced migratory and invasive capacity. Similarly, HIF-2α has been shown to drive conversion to a stem-cell phenotype and induce chemoresistance, conferring a selective advantage under hypoxic stress (Yan et al., 2018).

Together, these studies highlight how hypoxia generates distinct transcriptomic profiles within tumour regions, enriching for stemlike features and reinforcing spatial heterogeneity (Jehanno et al., 2022; Zhang et al., 2016).

These advantages – metabolic reprogramming, stemlike plasticity, and reduced ERα dependence – create competition among subclones, driving clonal selection and expansion of more aggressive traits (Walens et al., 2020; Wang et al., 2014b; Beumers et al., 2023). This process compounds the pro-tumorigenic effects until cells begin to migrate to distant sites in the body.

These observations point to the importance of examining both the spatial heterogeneity of low ERα-expressing cells at the hypoxic tumour core and the temporal dynamics of individual cells over time. Spatial heterogeneity is increasingly recognised as a critical feature of breast cancer, and 3D culture systems are being developed to better replicate patient tumours (Ji et al., 2019; Szostakowska et al., 2019; Kinnel et al., 2023; Fisher and Rao, 2020; Zhang et al., 2020; Lamichhane and Tavana, 2024). In parallel, single-cell methodologies are rapidly advancing, allowing transcriptional programmes to be resolved at unprecedented resolution and providing new opportunities to study how hypoxia shapes ERα signalling (Wang et al., 2023; Jovic et al., 2022; Wen and Tang, 2025).

Hypoxia is a central force shaping heterogeneity in ERα+ breast cancer. By reprogramming transcriptional networks, reducing ERα dependence, and fostering stem-like traits, hypoxia generates resistant subclones that undermine therapy. Recognising this complexity underscores the need for approaches that capture both the spatial organisation of tumours and the temporal dynamics of single-cell behaviour.

1.4. Three-dimensional cell culture methods used to study breast cancer

A promising route towards addressing the needs of patients with resistant cancer is three-dimensional (3D) cell culture methods. Recent developments in 3D culturing techniques have provided opportunities to study many factors contributing to breast cancer acquiring resistance to treatment, such as heterogeneity, hypoxia, stem-cell enrichment and many others that are not possible in 2D cell culture (Ji et al., 2019; Szostakowska et al., 2019; Kinnel et al., 2023; Fisher and Rao, 2020; Zhang et al., 2020; Lamichhane and Tavana, 2024).

In addition to patient-derived xenografts that implant human cancer cells in immunodeficient mice (Zhang et al., 2013a), many new models of studying human breast cancer in three-dimensional cultures have been developed (Redmond et al., 2021; Shi et al., 2023; Mamnoon et al., 2021; Däster et al., 2017; Bhattacharya, Calar and de la Puente, 2020; Glibetic et al., 2024; Grist et al., 2019; Pyne et al., 2024; Wang et al., 2018).

Several of these methods are highly technical or use uncommon reagents or equipment, such as bioprinting, microfluidic and other scaffold-based 3D cultures (Wang et al., 2018; Blanco-Fernandez et al., 2021; Major et al., 2024; Liverani et al., 2019; Wu et al., 2024). 3D scaffold cultures rely on embedding cells in a natural or synthetic hydrogel that attempts to more

faithfully replicate the TME extracellular matrix (ECM) (Wang et al., 2018; Blanco-Fernandez et al., 2021; Liverani et al., 2019).

The composition of scaffolds can be tailored to study microenvironmental factors such as ECM stiffness and oxygen gradients (Major et al., 2024; Whitman et al., 2019). These systems have been used to investigate how hypoxia contributes to the epithelial-mesenchymal transition (EMT) of primary breast cancer cell lines, providing mechanistic insight into how microenvironmental stress drives tumour progression (Wang et al., 2018). Bioprinting works on similar principles, except that cells and ECM bioink are directly printed into 3D structures, allowing the printing of custom vascularised breast cancer tumour models (Shi et al., 2023; Jia et al., 2016; Wu et al., 2024). The most technical iteration of this type of 3D cell culture is the microfluidic devices that use capillary action to combine cells suspended in Matrigel and collagen with molecules to encapsulate spheroids in individual beads (Grist et al., 2019; Pyne et al., 2024).

Together, these methods accurately replicate in vivo conditions well, are more accurate models for drug sensitivity assays, and critically, they capture aspects of tumour biology that cannot be modelled in 2D monolayers (Nunes et al., 2019; Imamura et al., 2015; Muguruma et al., 2020). By reproducing gradients of oxygen, nutrients, and ECM stiffness, they provide reproducible systems that better reflect the complexity of the breast cancer microenvironment.

The most accessible and least technical 3D cell culture system currently is spheroid/organoid cultures, as the simplest methods require typical cell culture reagents, but they also provide a malleable platform for more complex model systems (Däster et al., 2017; Ham et al., 2016; Raza et al., 2017; Yang et al., 2025; Sachs et al., 2018). It is important to highlight the difference between spheroid and organoid cultures, as there are large differences that have implications for the conclusions drawn from experiments using these methods. Spheroids are simpler aggregates of cells, like breast cancer cell lines or primary cells, whereas organoids are derived from stem or progenitor cells that are capable of differentiation (Gunti et al., 2021; Lee et al., 2023b; Sakalem et al., 2021). Organoids can be derived from patient material and can more faithfully recapitulate in vivo tumour progression (Mazzucchelli et al., 2024; Guillen et al., 2022), Spheroid models present one of the simplest and most reproducible systems to rapidly generate models with varying oxygen levels between cell populations. They can also be genetically modified with relative ease, making them highly adaptable for studying hypoxia, heterogeneity, and ERα signalling in breast cancer (Gunti et al., 2021; Lee et al., 2023b; Sakalem et al., 2021).

There are relatively few studies that directly examine ERα expression in spheroids. However, there is evidence to suggest that expression of ERα could be enhanced compared to 2D cultures, making spheroids an attractive model to study ERα (Whitman et al., 2019; Koletsis et al., 2025; Montani et al., 2014). In the breast, ECM stiffness has been shown to aid the maintenance of ERα expression (Munne et al., 2021; Jahin et al., 2023). Spheroid culture encourages cells to create cell-to-cell contacts that are more similar to their original tissue, and are then under different physical forces and gradients (like oxygen penetration across the spheroid) that are not present in 2D cultures (Whitman et al., 2019; Kapałczyńska et al., 2018; Breslin and O'Driscoll, 2016). Taken together, spheroid culture presents a useful method for studying differential ERα regulation during hypoxia.

While patient-derived organoids would provide more clinically relevant data, the expression level of ERα is variable from patient to patient (Makhlouf et al., 2024; Stravodimou and Voutsadakis, 2025), therefore, establishing this model would require extensive optimisation and donor selection. Instead, breast cancer cell lines could be used to establish the methods of studying differential ERα regulation and then be used as a control to select suitable donors for later establishment of patient-derived organoids.

There are many variations in the methods of establishing spheroid cultures with, as of yet, little standardisation across different research groups (Lonkwic, Zajdel and Kaczka, 2025; Raghavan et al., 2016; Jubelin et al., 2023; Beckers et al., 2025). Several categories of approaches have been developed, each with distinct strengths and limitations. Scaffold- and microfluidic-based systems allow precise oxygen control but require specialised equipment (Grist et al., 2019; Pyne et al., 2024; Arora et al., 2024; Tevlek et al., 2023). Mechanical aggregation methods, such as spinner flasks and rotating bioreactors, generate large numbers of spheroids by preventing adhesion and promoting cell-cell collisions, though they limit control over spheroid size and can introduce shear stress (Arora et al., 2024; Lonkwic, Zajdel and Kaczka, 2025; Achilli, Meyer and Morgan, 2012; Zinn, Mehner and Patel, 2023). Similarly, the use of microfabricated or micro-patterned plates can be used to reliably generate uniform spheroids and organoids (Gong et al., 2015; Chao, Ngo Le and Engelward, 2020); However, these methods also require high levels of technical expertise and equipment.

Simpler non-adherent techniques, including hanging drops and ultra-low attachment plates, rely on gravity or culture surface chemistry to force cells into aggregates (Foty, 2011; Kuo et al., 2017; Lonkwic, Zajdel and Kaczka, 2025; Jeong, Tin and Irudayaraj, 2022). Hanging-drop methods produce uniform spheroids but are low-throughput and labour-intensive (Foty, 2011; Kuo et al., 2017; Lonkwic, Zajdel and Kaczka, 2025; Jeong, Tin and Irudayaraj, 2022), while spheroids grown in ultra-low attachment plates are scalable and accessible, though some cell lines require matrix support to aggregate reliably (Beckers et al., 2025; Lonkwic, Zajdel and Kaczka, 2025; Raghavan et al., 2016).

Despite this diversity of methods, the lack of standardisation limits comparability between studies. For this thesis, the emphasis will be on developing a specific, well-defined spheroid model that is reproducible and capable of capturing spatial ERα binding heterogeneity. This approach will provide a system that critically improves on 2D culture by reflecting the complexity of cell-to-cell connections and microenvironmental factors such as hypoxia.

1.5. Methods to study transcription factor binding across mammalian genomes

1.5.1. Establishment of the gold standard - Chromatin immunoprecipitation and next-generation sequencing (ChIP-Seq)

Initial attempts to investigate the location of transcription factor binding in the genome utilised DNase I to digest open DNA (Galas and Schmitz, 1978; Tullius and Dombroski, 1986). Later developments in analysis of transcription factor binding sites would utilise technology

developed a decade before the original DNA footprinting method; The use of formaldehyde to crosslink protein to nucleic acids (Gilmour and Lis, 1985; Kahiapo and Monahan, 2023).

Chromatin immunoprecipitation (ChIP) was developed through combining crosslinking chromatin complexes with immunoprecipitation (IP) using antibodies specific against the protein of interest, to study protein binding patterns throughout the genome (Gilmour and Lis, 1985; Kahiapo and Monahan, 2023). By re-enforcing transient DNA binding interactions with crosslinking, both the TF of interest and the DNA it is bound to can be separated from a crude cell lysate through capture by an antibody raised against the TF (Kahiapo and Monahan, 2023; Gilmour and Lis, 1985). Originally, these experiments relied on PCR or hybridisation to detect enrichment at candidate loci, which limited analysis to a handful of sites at a time (Solomon, Larsen and Varshavsky, 1988; Nelson, Denisenko and Bomsztyk, 2006). The advent of next-generation sequencing transformed ChIP into ChIP-Seq by enabling genome-wide readout of all DNA fragments bound by the immunoprecipitated protein (Robertson et al., 2007; Barski et al., 2007). This step represented a major leap from targeted assays to unbiased profiling, allowing researchers to define complete cistromes rather than single binding events.

1.5.2. ChIP-Seq is the gold standard for transcription factor binding analysis

ChIP-Seq is widely regarded as the gold standard for transcription factor binding analysis to date (Xu et al., 2021; Jaini et al., 2014; Narlikar and Jothi, 2012; Jothi et al., 2008; Johnson et al., 2007). Since its invention, ChIP-Seq has been an invaluable tool to study the occupancy of chromosomal loci by target proteins and has been used to further understand gene regulation, epigenetics and more (Mundade et al., 2014; Bardet et al., 2013; Hu et al., 2010; Johnson et al., 2007; Jothi et al., 2008).

ChIP-Seq has been used extensively to define the ERα cistrome, its co-dependency with pioneer factors and how these insights are translated into clinical understanding by deepening understanding of endocrine therapy resistance (Carroll et al., 2006; Welboren et al., 2009; Hurtado et al., 2011; Magnani et al., 2011; Hah et al., 2011). ChIP-Seq revealed the requirement for FOXA1 binding before ERα (Hurtado et al., 2011), as well as revealing ERαs binding preference for distal enhancers over promoter regions (Carroll et al., 2006; Welboren et al., 2009). A combination of ChIP-Seq and other methods aids in creating compelling evidence of transcriptional output and enhancer activity, like GRO-Seq and PolII mapping to display rapid, transient waves of ERα-regulated transcription (Hah et al., 2011). Application of ChIP-Seq has also been applied to patient tumours to demonstrate the consequences that ERα binding profiles have on patient outcome and endocrine therapy resistance (Ross-Innes et al., 2012)

1.5.3. Limitations of ChIP-Seq

However, there are several issues a researcher using the method may face when applying it to their protein of interest (Wardle and Tan, 2015; Gilfillan et al., 2012).

The first and the hardest to overcome is that ChIP-Seq is heavily reliant on high-quality, specific antibodies used to capture the protein of interest (Wardle and Tan, 2015; Gilfillan et al., 2012). A poor-quality antibody will not effectively capture its target in the sample, resulting in significant sample loss during an experiment or, worse, significant off-target capture and obfuscation of the true signal. Issues with poorly optimised antibodies are not a problem for highly studied proteins such as ERα; however, researchers studying more obscure proteins may encounter issues conducting ChIP-Seq.

Similarly, IP is inherently challenging, as samples can be easily lost due to inefficient capture (Gilfillan et al., 2012). Consequently, standard ChIP-Seq protocols require high input material, which makes their application to rare populations and patient samples challenging. As a result of these challenges, several methods have been developed to modify ChIP-Seq protocols to generate high-quality data from low-input samples (Brind'Amour et al., 2015; Zhu et al., 2019; Cao et al., 2015; Rotem et al., 2015; Grosselin et al., 2019).

Beyond input requirements, inefficiencies in data normalisation also highlight the limitations of standard ChIP-Seq. For example, genome-wide changes in transcription factor binding cannot be accurately captured using transcriptomic normalisation strategies, which assume total occupancy remains constant. This issue has been demonstrated in studies of ERα binding, where experimental variability and global shifts in occupancy confound analysis.

The most common way to overcome issues with data normalisation is to include spike-in controls, however there are more recent attempts to improve normalisation, such as Parallel-factor ChIP (Guertin et al., 2018; Orlando et al., 2014; Vale-Silva, Markowitz and Hochwagen, 2019; Wu, Wang and Huang, 2021). Spike-in normalisation relies on adding a fixed quantity of exogenous chromatin to each sample, providing an external reference for scaling (Orlando et al., 2014; Vale-Silva, Markowitz and Hochwagen, 2019). Given that the added DNA is exogenous, spike-in chromatin acts as a fixed external reference that is unaffected by biological changes in the sample. The disconnection from biological shifts in the sample makes spike-in controls a useful tool, especially in contexts where large global shifts in chromatin occupancy is expected.

Parallel-factor ChIP was proposed as one alternative corrective approach, introducing internal reference peaks to normalise differential binding (Guertin et al., 2018). While spike-ins can be effective, they are sensitive to technical variability in spike-in preparation, mixing efficiency, and antibody cross-reactivity with the spike-in material. Parallel-factor ChIP avoids these issues by using an endogenous reference factor captured within the same IP, improving consistency across samples; however, it requires the presence of a suitable cofactor with stable genome-wide occupancy, which is not always available. Both methods therefore offer partial solutions to the limitations of standard ChIP-Seq. In practice, both approaches have their own strengths, and a researcher's choice depends on whether an external or internal reference is more appropriate for the experimental context.

The need for such methods underscores that conventional ChIP-Seq alone is insufficiently quantitative for single-cell TF analysis, and that additional strategies are required to improve reproducibility and accuracy.

1.5.4. Optimisation of ChIP-Seq for low input samples

Some methods offer simple optimisation steps that improve the data quality from low input conditions, such as ultra-low input native (ULI-N) ChIP-Seq (Brind'Amour et al., 2015), whereas others utilise oscillating washes with magnetic beads, or encapsulation in droplets using microfluidics (Rotem et al., 2015). Other library generation improvement methods aim to increase low-input samples' yield through the combination with tagmentation (Akhtar et al., 2019).

Other methods take more drastic changes to the standard ChIP-Seq method to reduce the required input even further. One such method is ultra-low input native ULI-NChIP-Seq (Brind'Amour et al., 2015). Designed to improve upon the Native ChIP method (Mendez et al., 2018; Nitsch and Schneider, 2024), this research group developed a method that uses a detergent-based lysis buffer, fragmentation via micrococcal nuclease (MNase) and optimised library preparation to generate high-quality data from as few as 1000 cells. Although methods such as this are very effective in analysing histone markers and generating fewer artefacts from crosslinking, they are less effective at capturing transient interactions such as TF binding.

Similarly, methods such as MINT-ChIP (van Galen et al., 2016), STAR-ChIP (Zhang et al., 2021a) and TCL-Seq (Zarnegar et al., 2017) provide innovative solutions for low-input and quantitative analysis of histone markers; they have limited application to transcription factors in their current formats. This limitation is due to the use of native ChIP-Seq, while it is possible to probe TF binding, capturing transient and rapidly binding TFs like ERα can be more challenging (Gilfillan et al., 2012; Hah et al., 2011).

Another approach applied to ChIP-Seq that does aid in the dissection of TF binding sites is microfluidic devices; one of these methods uses microfluidic oscillatory washing (MOW) based ChIP-Seq (Cao et al., 2015; Zhu et al., 2019). Fragmented chromatin is then added and efficiently absorbed into the dense bed of beads in a specialised PDMS microfluidics device, where it undergoes MOW. This process is where pressure is applied to both the inlet and outlet to wash the beads back and forth to create an over 35-fold enrichment of fragments of loci known to bind this antibody, compared to the negative control (Cao et al., 2015; Zhu et al., 2019). This method has been validated for use to study histone modifications down to 100-600 cells, although the authors only report the identification of transcription factor binding sites when using 10,000 cells.

1.5.5. Optimisation of ChIP-Seq for single-cell analysis

Single-cell analysis has exploded within the last decade and has carved out a niche in the centre of modern biological study (Kalisky et al., 2018; Zappia, Phipson and Oshlack, 2018). Adapting a gold standard tool like ChIP-Seq for single-cell TF analysis would be an invaluable asset for all of biological research, including breast cancer research (Rotem et al., 2015; Grosselin et al., 2019). Several attempts at merging single-cell analysis with ChIP-Seq have been made, such as DROP-ChIP (Rotem et al., 2015) and scChIP-Seq (Grosselin et al., 2019). Single-cell ChIP is challenging due to off target capture of chromatin via non-specific interaction creates noise in results and obfuscates the true signal. This noise is compounded in low-input experiments where the epitope of interest is not abundant, making it harder to

accurately identify the true signal (Rotem et al., 2015; Grosselin et al., 2019; Dahl and Gilfillan, 2018).

DROP-ChIP attempts to overcome noise by utilising DNA barcoding in combination with microfluidics to allow easier multiplexing of samples and a greater quality of data produced at a single-cell level (Rotem et al., 2015). Single-cells are encapsulated within aqueous drops containing a weak detergent and MNase. Simultaneously, another drop-maker generates drops from a barcode library from a 384-well plate and can then be combined in a three-point merger device with a third set of droplets containing enzymatic buffer and DNA ligase. Upon merging, this ligates sequencing-ready barcodes on the free ends of chromatin, uniquely tagging DNA from individual cells. Although the authors do not provide exact percentages, the overlap is illustrated in Venn diagrams, showing that the majority of peaks (>70-80%) are shared across single-cell clusters, which supports the robustness of the method despite low input noise.

A more recent publication appears to have taken this further to study intra-tumour heterogeneity in breast cancer (Grosselin et al., 2019). Further, tests on B and T cells produced two stable clusters after consensus clustering, which could accurately map cell identity with specificity over 99.7% and 99.5% when assessing two different histone markers (Grosselin et al., 2019). However, the method still contains limitations. Although it is possible to apply this method to TF binding, the lack of fixation makes it difficult to study these transient interactions. In addition, recommendations by DROP-ChIP authors include finding an alternative to MNase to increase applicability beyond chromatin markers.

Extensions of ChIP have also been developed to improve efficiency and recovery. Restriction enzyme-based labelling of chromatin in situ (RELACS) (Arrigoni et al., 2018) builds directly on ChIP principles by introducing barcoded adaptors during nuclei fragmentation, enabling multiplexing of many samples in a single reaction. By labelling chromatin fragments in situ, RELACS increases library efficiency, reduces labour-intensive steps, and improves recovery of material from limited input.

In summary, ChIP-Seq has provided invaluable insights into the key functions and impacts of proteins that coordinate the tumorigenic ERα transcriptional programme. Innovation and approaches have allowed the method to probe deeper into the transcriptional control of cells on a genome-wide scale. However, when explicitly asking questions regarding TF binding heterogeneity, further development is needed to generate data of the same quality as histone markers and epigenetic analysis. The evidence from single-cell adaptations and barcoding strategies such as RELACS collectively suggests that ChIP is evolving toward methods capable of resolving heterogeneity at single-cell resolution. This trajectory provides a clear justification for focusing on how ERα binding heterogeneity can be studied in breast cancer, and why further optimisation of ChIP-based methods remains central to answering this question. While these optimisations are necessary to push ChIP toward single-cell resolution, there are also alternative methods that bypass immunoprecipitation entirely and aim to generate single-cell transcription factor binding profiles.

1.5.6. Alternatives to ChIP-Seq that avoid immunoprecipitation

As previously mentioned, the IP step in ChIP-Seq is one of the largest sources of data loss in the technique (Wardle and Tan, 2015; Dahl and Gilfillan, 2018). Whether from a poor-quality antibody and low input, it is easy for fragmented chromatin to be missed during capture and thus could create gaps in data collected. multiple publications have attempted to tackle this issue by removing the IP step and taking alternative approaches to study TF binding and histone modifications through immunotethering (Harada et al., 2018; Maehara et al., 2021; Skene and Henikoff, 2017; Meers et al., 2019; Kaya-Okur et al., 2019; Wu et al., 2021; Bartosovic, Kabbe and Castelo-Branco, 2021).

The first method we describe is Chromatin integration labelling (ChIL), a technique that utilises immunostaining, Tn5 transposition and *in situ* transcription to simplify the study of TF binding (Harada et al., 2018; Maehara et al., 2021). Chromatin is fixed, fragmented and incubated with a primary antibody, the same as ChIP-Seq, but it is then incubated with a secondary antibody that has been conjugated with a ChIL probe. The integration of ChIL DNA allows for *in situ* transcription via the inserted T7 promoter and adjacent downstream genomic DNA. This then enables cell lysis and RNA purification for analysis via sequencing after library preparation.

The method was also applied to single-cells for histone markers, and the results are very similar to experiments run with 100 cells, indicating future applicability to single-cell analysis (Harada et al., 2018; Maehara et al., 2021). Although the authors do not test this method for TFs, they do note that the method shows considerable increases in sensitivity compared to DROP-ChIP . This technique clearly shows that alternative methods to ChIP-Seq can produce data of a similar quality while using an *in situ* method.

However, this method still carries its limitations. The use of Tn5 brings along the biases for open chromatin that are prevalent in Assay for Transposase-Accessible Chromatin with high-throughput sequencing (ATAC-seq). In addition, the use of primary and secondary antibody incubations makes the procedure intensive with overnight incubations and buffer exchanges. This drastically lowers the throughput of the methods and makes the technique less applicable on broader scales.

Similarly, other methods utilise specific antibodies to recruit other reagents to the binding site to enable analysis, such as Cleavage Under Targets & Release Using Nuclease (CUT&RUN) (Skene and Henikoff, 2017; Meers et al., 2019). This native *in situ* method aims to overcome issues that arise with solubility and crosslinking in the study of TF binding. This publication details how it aims to improve upon Chromatin Immuno-cleavage (ChIC)(M, T and Uk, 2004) through modifications inspired by other publications.

These publications show that CUT&RUN utilise the ability to release mononucleosis and TF: DNA complexes with light digestion of MNase in an attempt to apply this directly to the study of TF binding sites (Skene and Henikoff, 2017; Meers et al., 2019). Unfixed nuclei are immobilised on lectin-coated magnetic beads and are incubated with a primary antibody against the protein of interest. Once washed, a fusion protein is added, consisting of protein A and MNase (pA-MN), a bacterial protein that has an affinity for the constant region of immunoglobulins from certain species. The fusion protein recruits the MNase directly to the site of TF binding, and then the addition of divalent calcium ions (Ca^{2+}) initiates MNase DNA

cleavage and the reaction can be stopped shortly after through chelation of Ca^{2+}. This will directly release the soluble pA-MN:TF: DNA complex directly into the nucleoplasm, which can then be collected via centrifugation.

When applied to study Abf1 and Reb1, binding sites identified in this publication were almost identical to bona fide TF binding sites within the high-resolution ORGANIC dataset (Skene and Henikoff, 2017; Meers et al., 2019). Independent motif generation was then carried out for these TFs and compared with the CUT&RUN data from these experiments. Over 90% of binding sites were occupied by fragments for their corresponding motif, finding a very high signal contrast above background and almost negligible occupancy of one TF at the other TFs' binding motif sites.

Together, these data suggest CUT&RUN is highly sensitive and specific at generating precise maps of TF binding.

There have been many developments following on from the CUT&RUN protocol, including the most cutting-edge ultra-low input (uli-) CUT&RUN, which achieves the lowest input for the CUT&RUN methodology (Patty and Hainer, 2021; Lardo and Hainer, 2022). This protocol largely remains the same as the original except for small optimisations and substitution of specific reagents; the most prominent differences can be seen in the purification and library preparation steps. After chelation of Ca^{2+} to arrest cleavage, a Phenol-chloroform isolation and alcohol-salt precipitation are used to yield purified uliCUT&RUN-enriched DNA (Patty and Hainer, 2021; Lardo and Hainer, 2022). Then, in library preparation, NEBNext stem loop adaptors were integrated into library building as a cost-effective way to multiplex the library generated.

The data yielded from experiments were of high quality, with 100,000 unique reads from single cells and 1,000,000 from blastocysts. Their methods require 15,000,000 raw reads per sample, allowing for multiplexing of many samples in a single run.

The method still carries its limitations, such as due to being carried out under native conditions could prevent transient interactions from being detected. Although the publication suggests a method for using light crosslinking to improve the detection of these interactions. It is also a low-throughput method; although other similar techniques have been adapted for high-throughput applications, the same could be done here (Wang et al., 2019a; Bartlett et al., 2021; Salma et al., 2023). Automation has been applied to CUT&RUN, but not at the same sensitivity as uli-CUT&RUN (Maier et al., 2023).

Other publications have taken more drastic changes to the original CUT&RUN method rather than attempting to improve the method for low input; One such method is Cleave under targets and tagmentation (CUT&Tag) (Kaya-Okur et al., 2019; Wu et al., 2021; Bartosovic, Kabbe and Castelo-Branco, 2021).

This method attempts to reduce steps from the CUT&RUN protocol and make it more applicable to study single cells (Kaya-Okur et al., 2019). It does so by replacing the pA-MN fusion with a Tn5-protien A fusion (pA-Tn5) that is loaded with sequencing-ready adaptors. Using Tn5 eliminates the need for DNA end polishing and adaptor ligation before preparing sequencing libraries; instead, released fragments are immediately ready for library prep.

Transposition can be initiated with Mg^{2+} addition, directly integrating sequencing adaptors onto released DNA. Fragments can then be enriched from purified DNA and pooled for multiplex sequencing.

To compare this method to ChIP-Seq and CUT&RUN, 8 million reads from histone marker profiling were selected for comparison. ChIP-Seq data had such a high background that a total of 50 million reads were required to enable comparison. Both other techniques have very little noise, with CUT&Tag showing less background than CUT&RUN, although the latter also appears to have higher signal. Another recent publication directly comparing CUT&Tag and ChIP-Seq demonstrated an overlap of 54% in the average recall of known peaks for two histone markers (Abbasova et al., 2025). CUT&Tag has demonstrated that it is a powerful tool to study single-cell epigenomics, and recent applications to study TF binding demonstrate its potential application in the study of ERα+ breast cancer (Kaya-Okur et al., 2019; Wu et al., 2021; Bartosovic, Kabbe and Castelo-Branco, 2021).

Single-cell CUT&Tag (scCUT&Tag) builds directly on the CUT&Tag framework by coupling tagmented fragments to droplet-based single-cell library preparation (Wu et al., 2021). This adaptation enables each fragment pool to be uniquely barcoded, producing single-cell resolved chromatin maps. The method has been applied to multiple histone marks (H3K4me3, H3K27ac, H3K36me3, H3K27me3) and tested for TFs such as OLIG2 and RAD21, demonstrating that single-cell TF occupancy can be recovered. Although per-cell TF signal is sparse compared to histone marks, clustering and pseudobulk aggregation revealed cell-type specific binding patterns and recovered expected motifs, validating specificity. Integration with scRNA-seq further linked occupancy to transcriptional output, showing that scCUT&Tag can connect binding profiles to gene expression at the single-cell level.

Despite these advances, scCUT&Tag faces practical limitations. TF signals are sparse compared to histone marks, requiring deeper sequencing, stringent QC, and reliance on pseudobulking to achieve robust results. Per-cell read depth remains low, and while integration with scRNA-seq is possible, transcriptomic data are not currently captured simultaneously from the same assay.

When evaluating the current options for single-cell TF binding, each method carries distinct trade-offs. DROP-ChIP demonstrates feasibility but suffers from noise and dependence on MNase. uliCUT&RUN achieves ultralow input sensitivity but remains low-throughput and struggles with transient interactions. scCUT&Tag provides higher signal-to-noise and robust motif recovery, but its sparse per-cell coverage limits resolution.

These approaches show that single-cell TF binding analysis is possible, yet none of the methods described fully overcome the challenges of sensitivity, throughput, and integration. While they represent important progress, they remain limited in their ability to connect binding heterogeneity directly to transcriptional consequences. This limitation sets the stage for approaches such as calling card reporters, which will be discussed in the following section as a strategy designed to address these gaps.

1.5.7. Self-reporting transposons generate ‘Calling Cards’ to identify transcription factor binding sites

The previously described immunotethering methods all utilise fusion proteins to direct molecules to the TF binding site (Harada et al., 2018; Maehara et al., 2021; Skene and Henikoff, 2017; Meers et al., 2019; Kaya-Okur et al., 2019; Wu et al., 2021; Bartosovic, Kabbe and Castelo-Branco, 2021). Although this is a potent tool, they all contain a similar weakness, also inherent to ChIP-Seq, in that they only provide a snapshot of transcription factor binding at the time of collection (Furey, 2012; Whitington et al., 2011; Mundade et al., 2014). This prevents observation of the cell fate consequences resulting from the TF binding profile generated. In addition, these methods remain dependent on antibodies and immunoprecipitation steps to tether reagents to the TF of interest. As discussed above, antibody quality and capture efficiency are major sources of data loss, meaning that even the most advanced CUT&RUN or CUT&Tag protocols inherit this vulnerability. Calling cards resolve this limitation by eliminating the need for antibodies, instead relying on the transcription factor itself to direct integration events.

An alternative method to study TF binding in mammalian cells that is also capable of recording TF binding over a period of time, in addition to capturing the transcriptome of single-cells, is relatively new, calling card reporting (Wang, Johnston and Mitra, 2007; Wang et al., 2011, 2012a; Moudgil et al., 2020; Lalli et al., 2022; Cammack et al., 2020).

Initially reported in 2007, calling card reporting is an *in situ* method that probes TF binding in live cells (Wang, Johnston and Mitra, 2007). After successive iterations, this technology has recently been applied to many different areas of study within TF binding analysis (Wang et al., 2011, 2012a; Moudgil et al., 2020; Lalli et al., 2022; Cammack et al., 2020).

Calling cards uses a unique method to map TF binding compared to previously discussed methods have used antibodies to either direct reagents to release a TF bound to its DNA binding site or to capture TF: DNA complexes that have already been liberated (Wang, Johnston and Mitra, 2007; Wang et al., 2011, 2012a; Moudgil et al., 2020; Lalli et al., 2022; Cammack et al., 2020). The calling card methodology avoids chromatin fragmentation by using the TFs' inherent natural affinity for their targets to direct targeted DNA release.

1.5.8. First applications of calling cards in yeast

Initial experiments used a fusion protein with the transcription factor of interest itself and a Sir4 domain that can recruit a Ty5 transposase in yeast (Wang, Johnston and Mitra, 2007; Wang et al., 2011). This method uses co-transfection of a plasmid containing this TF-Sir4 fusion protein, a plasmid with Ty5 and another with the transposon in which a self-reporting transposon (SRT) gene (GFP, $Puro^{res}$, Gal4) and barcode are present. Successful co-transfection will induce transient expression of the fusion protein, localisation of the transposon to the binding site and insertion of the self-reporting transposon into the site of TF binding.

The integrated transposon acts as a ‘Calling card’, a permanent record that indicates a binding event occurred at that site of the genome (Wang, Johnston and Mitra, 2007; Wang et al., 2011). The permanent integration allows for the collection of DNA, and digestion by restriction

enzymes and self-ligation of fragments for use as a template for inverse PCR. The products can then be sequenced, and the reads can be mapped back to a reference genome to identify the location of TF binding sites.

By running a concurrent control that does not contain the TF-Sir4 fusion protein, quantitative analysis can reveal any insertions that are likely due to chance interaction of an undirected transposon. Using receiver-operator curves tests the method's sensitivity against the false positive rate, where the area under the curve (AUC) indicates accuracy. A score of 1 indicates it is perfectly accurate, and an area of 0.5 indicates performance is due to chance. When testing three TFs in yeast, two received a score of 0.99 and another 0.84, in addition to identifying 100% of the known binding sites of one of these TFs.

1.5.9. Application of Calling Cards to mammalian systems

Later publications simplified the protocol through direct fusion of TF and transposase, in addition to assessing an alternate transposase – *PiggyBac* (PB) (Wang et al., 2012a; Moudgil et al., 2020; Lalli et al., 2022; Cammack et al., 2020). PB is a cut-and-paste transposon that was likely selected due to its high levels of activity in mammalian genomes and has been widely used in Drosophila genomic research (Zhao et al., 2016; Woodard and Wilson, 2015). The SRT constructs were modified to contain a reporter gene flanked by the PB long terminal repeat (LTR). After the characterisation of undirected PB insertion, this publication attempts to use this system to map SP1 binding sites in the human colon adenocarcinoma cell line HCT116 (Wang et al., 2012a).

40,000 insertions were identified using this system, with drastically different distribution from the undirected PB. Insertions were also more clustered, particularly around transcription start sites (TSS) and in open chromatin (Wang et al., 2012a). A statistical model was created to assess loci with more insertions than chance, which revealed very high confidence clusters of CC, of which 83% contained a SP1 binding site nearby.

In addition to identifying the SP1 binding motif within these reads, these data clearly show that directed PB can mark known SP1 binding sites by transposing into the nearby genome (Wang et al., 2012a). The TTAA insertion preference does not significantly impact results as CC centres were on average 6 bp away from SP1 binding motifs, well within the range observed in ChIP-Seq (0-20 bp). However, this lack of impact reflects the specific distribution of SP1 motifs, which are frequently located near TTAA sites (Moudgil et al., 2020; Cammack et al., 2020; Wang et al., 2012a). The preference remains an inherent limitation of PiggyBac: loci that lack TTAA sequences cannot be marked, reducing sensitivity. While SP1 binding was largely unaffected, other transcription factors with different motif distributions may be more constrained by this bias.

There was a high degree of overlap in loci identified, as 80% of CCs had a ChIP-Seq peak identified within 250 bp. In addition, 93% of CCs had either a ChIP-Seq peak or an SP1 motif nearby, clearly showing a strong predictive power (Wang et al., 2012a).

Next, the sensitivity of the method was tested by identifying 'high confidence' targets. These targets must appear in the TRANSFAC database, contain an SP1 site, and be identified in the

ENCODE project dataset, and the author's own ChIP-Seq dataset run concurrently. The method shows 79% sensitivity to TTAA-containing sites compared to 39% at TTAA-deficient loci (Wang et al., 2012a).

At this point, calling cards are also still limited by the use of restriction enzymes, as recognition sites must be at a suitable distance from CCs; as such, alternative recovery protocols could improve this. Another consideration is the fusion of the transposase to the TF itself. Depending on whether the transposase is fused to the N- or C-terminus, steric hindrance may interfere with DNA binding or disrupt interactions with cofactors. This means that pilots must be carried out with both orientations, and with different linker lengths, to ensure that the fusion does not compromise the natural activity of the TF.

There are also several distinct advantages to this method, providing a way to record TF binding events in live cells permanently, enabling the use of assays to mark TF binding and subsequent culture of cells to assess cell fate consequences of this binding (Wang et al., 2012a). This provides information that cannot currently be assessed using other methods.

1.5.10. Single-cell calling cards

The development of single-cell calling cards (scCC) provides a path toward resolving the central question posed in this thesis: whether transcription factor binding heterogeneity can be mapped at single-cell resolution and linked to cell identity (Moudgil et al., 2020). With targeted adaptations, scCC could be applied to ER, enabling the study of ER binding patterns across heterogeneous tumour populations.

Single-cell RNA sequencing is the go-to method for dissecting cell identity in a heterogeneous cell population, as it serves as a readout for genes being expressed within the cells (Haque et al., 2017; Jovic et al., 2022; Choi and Kim, 2019). In scCC, typical transcriptomic RNA is captured alongside distinct self-reporting transposon (SRT) transcripts; the former reflects cell state, while the latter encodes transposon-genome junctions originating from TF-directed insertion events. This pairing allows TF binding profiles to be mapped and directly associated with cell identity within the same assay.

Validation demonstrated mapping of BRD4 at super-enhancers, high reproducibility (Pearson correlation of 0.997 between normalised insertions across replicates); the undirected transposase has a natural affinity for the BRD4 protein (Moudgil et al., 2020; Cammack et al., 2020; Yen et al., 2023). The approach was extended to additional TFs (SP1, FOXA2, BAP1), recovering canonical motifs, expected binding profiles and significant enrichment of ChIP-Seq signal at calling card peaks ($p = 1 \times 10^{-30}$) across diverse cell lines.

Circularisation is required to bring long SRT transcripts, barcodes, and genomic junctions together for short-read identification, which can reduce recovery efficiency in low-input contexts (Moudgil et al., 2020). Subsequent work adapted scCC for in vivo delivery via AAVs, enabling continuous deposition of calling cards in the mouse brain. Cohorts collected at day 10 and day 28 demonstrated temporal recording: early-activated genes were marked at both times, late-activated genes only at day 28, and constitutive genes accumulated more

insertions over a period of time. This capacity to record binding history highlights the relevance of scCC for studying the temporal effects of hypoxia-induced heterogeneity on ERα binding.

An important consideration in interpreting calling card data are the key differences in how the binding data is recorded and stored (Moudgil et al., 2020, 2021; Lalli et al., 2022; Yen et al., 2023). Calling card peaks act as historical record of a TF binding event rather than a quantitative measure of TF occupancy like ChIP-Seq. More recent developments of the method use barcoded SRTs to both multiplex TF binding collection in addition to identifying unique binding events (Lalli et al., 2022), but this method is inherently limited by the number of barcodes available. There are only 23 barcodes available, as any mutation in the HyPB terminal repeat region prevents recognition by the transposase. This key difference compared to the gold-standard ChIP-Seq makes calling cards a useful binary indicator tool for studying TF binding but gives more limited information regarding the strength of TF binding.

Taken together, scCC is the first method to simultaneously capture TF binding and transcriptomic identity at single-cell resolution, with the added potential for temporal recording in vivo. While practical challenges remain (transposase biases, fusion optimisation, recovery efficiency, capturing TF binding strength), scCC directly advances the aims of this thesis by enabling ERα binding heterogeneity to be mapped to cell identity and fate in breast cancer.

1.6. Project introduction

1.6.1. Hypothesis

Building on the single-cell calling card framework outlined above, we hypothesise that heterogeneity of ERα binding may contribute to endocrine therapy resistance in breast cancer, and that this effect could be further exacerbated by signalling crosstalk and hypoxia. To test this hypothesis, we created a system to enable Calling Card reporting with self-reporting transposons (SRTs) in breast cancer cell lines to investigate how heterogeneous ERα binding contributes to resistance in endocrine therapy. We will be focusing on ERα as it is a major driver of breast cancer, and we aim to engineer an ERα Calling Card reporter platform capable of generating the first single-cell ERα binding profile in MCF7 cells.

Simultaneously, we established a reproducible three-dimensional MCF7 cell culture method that reliably generates hypoxia due to impaired oxygen diffusion that could be used in the future in combination with ERα scCC in cells resistant and sensitive to treatment. It would then be possible to employ these methods using patient samples to dissect how clinically relevant ERα-driven gene regulation during hypoxia affects endocrine therapy resistance.

1.6.2. Rationale

1.6.2.1. Single-cell ERα Calling Cards

Previous publications have shown that calling cards are an accurate, sensitive, and reliable method to probe TF interactions on a genome-wide scale (Wang et al., 2012a; Moudgil et al., 2020; Lalli et al., 2022; Cammack et al., 2020). Data from calling card experiments are

comparable to data collected from the current gold standard – ChIP-Seq – as between 83 to 90% of calling card peaks can be found within 1000 bp of ChIP-Seq peaks for multiple TFs (Wang et al., 2012a; Moudgil et al., 2020; Lalli et al., 2022; Cammack et al., 2020).

scCC presents an alternative to ChIP-Seq that can overcome several limitations of ChIP, circumventing the cell input requirements and IP limitations, enabling simultaneous transcriptomic and TF binding analysis in a single experiment (Moudgil et al., 2020). From a bulk call population, ChIP-Seq averages the signal of all cells in a culture and cannot detect differences in binding driven by heterogeneity (Rotem et al., 2015; Kaya-Okur et al., 2019). In comparison, calling cards have demonstrated single-cell applications to identify TF binding sites and directly link that information to cell identity and cell fate consequences in cell lines and live animals (Wang et al., 2012a; Moudgil et al., 2020; Lalli et al., 2022; Cammack et al., 2020).

Although single-cell ChIP exists for histone modifications (Rotem et al., 2015), Its application is limited when applied to transcription factors due to poor data sensitivity (Rotem et al., 2015; Kaya-Okur et al., 2019). Ultra-low input and single-cell ChIP methods using microfluidics (MOW-and DROP-ChIP) and other alternative protocols, such as uliCut&RUN, aim to improve resolution for transcription factors but are still limited (Harada et al., 2018; Maehara et al., 2021; Skene and Henikoff, 2017; Meers et al., 2019; Kaya-Okur et al., 2019; Bartosovic, Kabbe and Castelo-Branco, 2021; Rotem et al., 2015; Zhu et al., 2019; Brind'Amour et al., 2015; Cao et al., 2015). Single-cell Cut&Tag has demonstrated the ability to study single-cell transcription factor binding, but it does not simultaneously capture the transcriptome, nor does it historically record TF binding over a period of time like scCC (Bartosovic, Kabbe and Castelo-Branco, 2021; Moudgil et al., 2020).

To increase the recovery efficiency of SRT junctions in single-cells, we will integrate long-read sequencing to reduce circularisation burdens and improve sample capture, thereby strengthening the linkage between ERα binding profiles and transcriptomic identity in heterogeneous populations. This is a novel extension of the existing methods, serving as a platform for other calling card reporters to improve efficiency and detection of TF-directed calling card insertions.

Through engineering of a novel ERα calling card platform to study binding heterogeneity, particularly in patient material, we will be able to probe both inter-and intra-tumour heterogeneity of ER binding. In doing so, we may reveal how ER can drive clonal evolution within a tumour and why this leads to metastatic relapse in some patients and not others.

One consideration for our ERα calling card approach is mutational-burden that accompanies the transposon-mediated insertion of exogenous SRTs into the genome. Although previous studies have shown the overall burden of insertions is tolerated, they recognise that any integration event can potentially disrupt local chromatin or gene function (Moudgil et al., 2020; Wang et al., 2012a; Cammack et al., 2020; Lalli et al., 2022). Cells that carry insertions in essential promoters, genes or other key regulatory elements are less likely to survive and are therefore under-represented in the final calling card dataset. Calling card datasets therefore

are biased towards regions that are tolerant to insertion, and this should be kept in mind when interpreting the genome-wide profiles. However, with these limitations in mind, the method remains an effective tool for mapping TF binding, as the vast majority of binding events occur in non-coding regions of the genome (Slattery et al., 2014; Stergachis et al., 2013). Furthermore, CC's have been used to study TF binding in live mouse brains and no negative effects on the mice were observed, indicating that mutation-induced cell death is limited (Cammack et al., 2020).

In this thesis, we will set out how we engineered a platform capable of creating the first single-cell ER binding profile in breast cancer cells and validated its ability to map genuine binding through comparison with ChIP-Seq and Chromatin Interaction Analysis by Paired-End Tag Sequencing (ChIA-PET) datasets. Where relevant, long-read sequencing will be applied to enhance SRT recovery and mapping fidelity in low-input contexts.

1.6.2.2. MCF7 hypoxic spheroid model

Immortalised cell lines and the culture of primary tissue expanded in 2D cell culture flasks is a mainstay of breast cancer research. These methods have been widely adopted due to their relatively low cost, reliability, and reproducible in undertaking experiments. Often, before testing and then confirming results in more costly animal models (Costa et al., 2016). By more faithfully recreating in vivo conditions, cells grown using this method can respond in ways that more closely resemble the responses of cells within patient tumours.

We aim to apply 3D spheroid generation techniques to create a model to study hypoxia-induced effects on ERα biology and how that could contribute to endocrine therapy resistance. MCF7 spheroids have been used to study hypoxia in the past, and growing spheroids to a diameter of ~200 - 500 μm (for MCF7) will begin to exhibit a hypoxic core of cells (Yakavets et al., 2020). However, there is little consistency between methods, and there has not been research closely characterising the development of hypoxia and establishing a reliable model of MCF7 hypoxic spheroids (Beckers et al., 2025).

We intend to create a simple protocol to generate MCF7 spheroids and characterise how they develop hypoxia over time, and demonstrate their ability to study hypoxia-driven ERα changes through inhibiting hypoxic ERα target gene products (Malcolm et al., 2025). We demonstrate the development of hypoxia in our model through immunostaining with EF5, a molecule that is reduced in low oxygen and can be detected by immunofluorescence microscopy (Wang et al., 2012b; Pugh-Toole et al., 2022; Mikhail, Eetezadi and Allen, 2013; Lao et al., 2015; Grimes et al., 2014; Pires et al., 2012). We can then treat spheroids with inhibitors of proteins identified in our recent publication, which found a novel signature of genes that are upregulated by ERα specifically during hypoxia (Malcolm et al., 2025). This model is designed to be compatible with single-cell ER calling cards, and-with long-read sequencing for efficient SRT recovery enables temporal and spatial dissection of hypoxia-driven heterogeneity.

1.6.3. Our research aim

The overarching aim of this thesis is to engineer and validate a single-cell experimental platform that links ERα binding heterogeneity to transcriptomic identity within the breast cancer microenvironment. By integrating long-read sequencing to improve recovery of SRTs and combining this with a reproducible hypoxic spheroid model, the vision is to create a framework that can dissect how ERα binding heterogeneity drives endocrine therapy resistance in clinically relevant contexts.

1.6.4. Research Objectives

Research Strand 1: Single-cell ERα Calling Card Development

1. Create fusion proteins of ERα with the hyperactive transposase PiggyBac (HyPB) at N- and C-terminal positions.
2. Optimise a bulk calling card system capable of mapping genome-wide ERα binding in heterogeneous populations.
3. Validate ERα single-cell calling card sequencing, incorporating long-read approaches to improve circularisation efficiency and recovery, and establish the first ERα binding profiles at single-cell resolution.

Research Strand 2: Spheroid Model of ERα+ Breast Cancer

4. Establish a reproducible protocol to generate spheroids using MCF7 cells.
5. Characterise spheroid growth to determine when hypoxic cores develop without progressing to necrosis and validate hypoxia via EF5 immunostaining.
6. Assess ERα expression and hypoxia-driven changes to establish compatibility with single-cell calling cards, with longread sequencing enabling efficient SRT recovery in complex 3D samples.

2. Materials and Methods

2.1. General molecular biology methods and construct optimisation

2.1.1. Culturing competent bacteria

Competent *E.coli* Stbl3 cells (RecA-) (ThermoFisher, Ref: C737303) were cultured on autoclave-sterilised LB agar plates prepared from LB Agar powder (Formedium, PN: LMM0204). The appropriate antibiotics for selection were added to molten agar at a temperature of 50-60 °C: ampicillin sodium (Cambridge Bioscience, PN: A0425g; 100 µg/mL) or kanamycin sulphate (Merck, PN: K40005G; 50 µg/mL). Plates were poured into Thermo Scientific petri dishes (PN: 101IRR) and stored at 4 °C until use.

Liquid cultures were established in autoclave-sterilised 0.02 w/v LB media (Sigma, PN: L3022). Cultures were inoculated from single colonies in Falcon 14 mL tubes (PN: 352001) or Sarstedt 15 mL tubes (PN: 62.554.502) and were incubated at 37 °C with shaking at 250 rpm. Antibiotics were added at the appropriate concentrations to maintain plasmid selection.

Glycerol stocks were prepared for long-term storage by mixing overnight cultures with sterile glycerol (Sigma, PN: G9012) to a final concentration of 25% and storing samples in Greiner cryovials (PN: 122263) at -80 °C.

QIAprep spin miniprep kits (Qiagen, PN: 27106) or QIAprep midiprep kits (Qiagen, PN: 12143, lot: 169055838) were used according to the manufacturer's instructions to purify plasmid DNA, depending on the amount required. When higher concentrations of plasmid were needed, the DNA was further purified by ethanol precipitation. Sodium acetate (ThermoFisher, PN: R1181) was added to a final concentration of 0.3 M, followed by three volumes of absolute ethanol (Sigma, PN: 20821.330). Samples were incubated at -80 °C for three hours, centrifuged at 4 °C at maximum speed, and pellets were resuspended in nuclease-free water (VWR International, PN: PD092) for downstream applications.

2.1.2. Fusion protein cloning

Cloning primers were designed using SnapGene (Version: 8.2.0) and the ThermoFisher Tm calculator, typically 18-25 nucleotides in length with melting temperatures between 55-60 °C (Appendix Table 8.1). The ERα coding sequence was PCR amplified from the pCMVhERalpha plasmid (Addgene, PN: 1011410) and cloned into the pRM1258 WT-HyPB vector (Moudgil et

al., 2020). Both N- and C-terminal fusion constructs were generated to evaluate the positional effects of a fusion partner at either end of the ERα on protein function. We designed primers to incorporate the linker sequence of previously published calling card reporters (Moudgil et al., 2020) to ensure an in-frame fusion between the two protein-coding sequences. The reverse primer designed to generate the N-terminal removed the ESR1 stop codon to extend into the added linker sequence, while the C-terminal reverse primer retained the stop codon; instead, the forward primer removed the HyPB stop codon before the linker sequence. The translated linker sequence was “KLGGGAPAVGGGPKAADK”.

PCR amplification was performed using the 2x PCRBIO HS VeriFi Mix Red master mix (PCRBIO Systems, PN: PB10.4701) with primers (Merck) added from 10 µM working stocks in nuclease-free water (VWR International, PN: PD092). The reaction was carried out on a DNA Engine Dyad thermocycler (Model PTC220, Serial DY001978), and the composition is summarised in Table 2.1, and cycling conditions are shown in Table 2.2

Table 2.1: General PCR components

Reagent	**Volume (µL)**
2X PCRBIO HS VeriFi Mix Red (PCRBIOSTSTEMS, PN: PB10.47-01)	10
DI Water	8
Primer Pair mix (10 µM forward and reverse primers)	1
Sample DNA (1-100 ng)	1

Table 2.2: Fusion protein cloning PCR programme

PCR Step	**Temp (ºC)**	**Time (s)**
Initial denaturation	95	180
Amplification Cycles (30 total)	95	30
	55	30
	72	60
Final extension	72	600
Hold temp (ºC)	4	Forever

PCR products were verified by agarose (Melford, PN: A20090) gel electrophoresis with a Sub-Cell GT tank (Bio-Rad, PN: 1704401) using 1X TAE buffer prepared from Tris base (Sigma, PN: 302050), EDTA (Santa Cruz Biotechnology, Ref: sc-203932), and glacial acetic acid (Sigma, PN: 332092.5L). DNA was visualised by incorporating SYBR Safe stain into gels (Invitrogen, PN: S33102) and was contextualised with the DNA Ladder II (PCRBIO Systems,

PN: PB40.1205). DNA fragments were purified using Monarch DNA gel extraction kits (NEB, PN: T1020S), with the optional wash step included.

Restriction digests were performed with XhoI (NEB, PN: R0146S), SalI-HF (NEB, PN: R3138S), and HindIII-HF (NEB, PN: R3104S) following the manufacturer's protocols. Inserts were assembled into the HyPB backbone using NEBuilder HiFi DNA Assembly (NEB, PN: E5520S) at a 1:2 vector-to-insert ratio, incubated at 50 °C for 15 min.

Chemically competent E. coli Stbl3 (RecA-) cells (ThermoFisher, Ref: C737303) were transformed with ligation products by heat shock, recovered in SOC media, and plated on LB agar supplemented with the appropriate antibiotic. Plates were incubated at 37 °C to obtain single colonies for downstream culture. Plasmid recovery was performed as described in Section 2.1.1, using QIAprep spin or midiprep kits (Qiagen, PN: 27106 or 12143).

Constructs were verified by Sanger sequencing (SourceBioscience). Sequence traces were analysed using SeqScanner 2 (Applied Biosystems, Version 2.0) and were used to align with template maps in SnapGene Viewer. For downstream expression, entry clones were recombined into destination vectors using Gateway LR Clonase II enzyme mix (Invitrogen, Ref: 11791020) in TE buffer (ThermoFisher, Ref: J75793.AE), following the manufacturer's protocol. The resulting expression constructs are shown schematically in (Figure 3.2).

2.1.3. Construct optimisation

2.1.3.1. Site-directed mutagenesis to add a Kozak sequence to the C-terminal fusion expression plasmid

To enhance translation efficiency of the C-terminal HyPB-ERα fusion, a Kozak sequence was introduced upstream of the HyPB coding region by site-directed mutagenesis. Forward and reverse primers complementary to the target region (5 bp before the HyPB start codon) were synthesised to incorporate a 3 bp mutation to generate a complete Kozak consensus sequence (Appendix Table 8.1).

The template C-terminal HyPB-ERα expression plasmid was diluted to 5 ng/µL in nuclease-free water (VWR International, PN: PD092). PCR reactions were assembled using the 2x PCRBio HS VeriFi RedMix master mix (PCR Biosystems, Ref: PB10.4701) with 1 µM of each mutagenic primer. Reaction volumes and composition followed the general PCR conditions in

Table 2.1 and were performed on a DNA Engine Dyad thermocycler (Model PTC220, Serial DY001978) using the cycling parameters used in Table 2.3,

Following amplification, the template DNA was digested with DpnI (NEB, Ref: R0176S) according to the manufacturer's instructions. The digested products were transformed into chemically competent E. coli Stbl3 cells as described in Section 2.1.1, and colonies were selected on LB agar containing the appropriate antibiotic. Plasmid recovery was performed using QIAprep kits (Qiagen, PN: 27106/12143), and successful incorporation of the Kozak sequence was confirmed by Sanger sequencing (SourceBioscience). Sequence analysis was carried out using SeqScanner 2 (Applied Biosystems, Version 2.0) and aligned to template maps in SnapGene Viewer.

Table 2.3: Site-directed mutagenesis PCR programme

PCR Step	**Temp (ºC)**	**Time (s)**
Initial denaturation	98	30
Amplification	98	10
Cycles *(18 total)*	55	30
	72	270
Final extension	72	600
Hold temp (ºC)	4	Forever

2.2. MCF7 cell culture, transfection, and puromycin selection

2.2.1. Cell recovery and maintenance

MCF7 cells were thawed from liquid nitrogen storage by rapid warming at 37 °C and resuspension in DMEM (Gibco, PN: 41966029) supplemented with 10% heat-inactivated fetal bovine serum (FBS-HI; Gibco, PN: 10500064). The cells were counted using a hemocytometer (Marienfeld, PN: 8100103), and viability was determined using trypan blue exclusion (ThermoFisher, PN: T10282). Cells were then pelleted at 400 xg for 5 min and were resuspended in fresh DMEM for seeding at appropriate densities depending on the culture vessel used (e.g., 0.7 x10^6 cells in T25 flasks; Corning, PN: 430639).

The cells were maintained in DMEM + 10% FBS-HI at 37 °C and 5% CO2, and cultures were passaged every 3-4 days at 60-80% confluence. Cells were washed twice with PBS (Fisher, PN: 10209252) and detached with 0.05% trypsin-EDTA at 37 °C (Gibco, PN: 25300054). Following cell detachment, trypsin was neutralised with fresh media and pelleted by centrifugation at 400 xg for 5 min. Cells were then resuspended in media, counted as described above, and seeded at appropriate densities (e.g., 2.1 $x10^6$ cells in T75 flasks; Corning, PN: 430641U).

2.2.2. Lipofectamine 3000 transfection

DNA transfections were performed using Lipofectamine 3000 (ThermoFisher, PN: L3000001) according to the manufacturer's instructions, with conditions optimised for MCF7 cells. The cells were seeded at 1.5 $x10^5$ cells/well in 24-well plates (Corning, PN: 10380932) or 0.45 $x10^6$ cells/well in 6-well plates (Corning, Ref: 3516) to reach ~70% confluence after 24 hours.

In 24-well plate transfections, complexes were prepared in 25 µL OptiMEM (ThermoFisher, PN: 11524456) containing 500 ng total plasmid DNA and 1 µL P3000 reagent. This mix was combined with 1.5 µL Lipofectamine 3000 diluted in 25 µL OptiMEM. For 6-well plates, complexes were prepared in 125 µL OptiMEM with 1 µg total plasmid DNA and 2 µL P3000 reagent, which was then combined with 5 µL Lipofectamine 3000 diluted in 125 µL OptiMEM. Across both plates, complexes were incubated for 15 min at room temperature before being added to cells dropwise. The cell culture media were replaced after 24 hours to reduce cytotoxicity.

This framework was applied across multiple experiments in the following chapters. Table 2.4 shows an example of plasmid combinations used in Kozak SDM transfections. Each experiment contains a table specifying the constructs used (e.g. HyPB-ERα fusion variants, Kozak-mutated plasmids, puromycin self-reporting transposons) and the experimental context (Western blot validation, Kozak optimisation, puromycin dose-response). Unless otherwise stated, all transfections followed the general conditions described above.

Table 2.4: A table of Transfection conditions for MCF7 cells, subsequently used for Western Blotting, Kozak-mutated plasmids

Transfection condition name	**Plasmid 1 (500 ng)**	**Plasmid 2 (500 ng)**
ERα N-terminal fusion protein	ERα-HyPB N-terminal expression plasmid	Puromycin Self-reporting Transposon
ERα C-terminal fusion protein Minus Kozak Sequence	ERα-HyPB C-terminal expression plasmid - Kozak	Puromycin Self-reporting Transposon
ERα C-terminal fusion protein Plus Kozak Sequence	ERα-HyPB C-terminal expression plasmid + Kozak	Puromycin Self-reporting Transposon
Puromycin control + pUC19	Puromycin Self-reporting Transposon	pUC19
No Transfection control	None	None

2.2.3. Puromycin dose-response curve for MCF7 cells

To determine the optimal puromycin concentration for selection, MCF7 cells were seeded at 0.07 x10^6 cells/well in 24-well plates (Corning, PN: 10380932) and treated with a range of puromycin concentrations (Sigma, PN: P8833) from 0 to 10 µg/mL prepared from a x1000 stock. Cells were imaged daily for 7 days, and the culture media were refreshed every 2-3 days. Images were analysed in ImageJ (Version 2.16.0, Fiji Version 1.54P) using the Cell Counter plugin, and cell survival was normalised to the untreated controls to identify the concentration that achieves complete cell death over 7 days.

For later optimisation in Lipofectamine-transfected cells, MCF7 cells of different passage numbers (P#15 and P#31) were seeded at 0.45 x10^6 cells/well in 6-well plates (Corning, Ref: 3516). After 24 hours, the cells were transfected with HyPB-ERα fusion constructs and puromycin SRTs (Table 2.5). The cells were then expanded into 10 cm plates (VWR, Ref: 10062880) and allowed to reach 70% confluence before treatment with puromycin at 1.25, 1.5, or 1.75 mg/mL. The cell culture media were refreshed every 2-3 days until no viable cells remained in the SRT-only control; cells were harvested for RNA.

Table 2.5: A table of Transfection conditions used to generate transfection reagents for the optimisation of puromycin treatment in cells transfected with the N-terminal fusion protein plasmid and puromycin resistance SRT.

Transfection condition name	**Plasmid 1 (500 ng)**	**Plasmid 2 (500 ng)**
ERα N-terminal fusion protein Calling Cards	ERα-HyPB N-terminal expression plasmid	Puromycin Self-reporting Transposon
Puromycin control + pUC19	Puromycin Self-reporting Transposon	pUC19

2.3. Protein expression and HyPB functional validation

2.3.1. Expression confirmation via Western Blotting

To confirm full-length HyPB-ERα expression, MCF7 cells were transfected with either N- or C-terminal HyPB-ERα fusion constructs for subsequent Western blotting (Table 2.6). The cells were directly lysed in the plate wells using 2x Laemmli buffer (BioRad, PN: 1610737) diluted 1:1 in PBS containing cOmplete protease inhibitor (Roche, PN: 118735080001). Lysates were scraped until viscous, and samples were frozen at -20 °C before use. When needed, samples are defrosted, and to each 100 µL aliquot, 6.7 µL 2-mercaptoethanol (BioRad, PN: 1610710) was added. The samples are then denaturation at 95 °C for 5 min, followed by briefly vortexing, and centrifuging to prepare them for SDS-PAGE.

Table 2.6*:* Plasmids used to transfect MCF7 cells in preparation for subsequent Western Blot

Transfection condition name	**Plasmid (500 ng)**
Wild Type (WT) + Puromycin Calling Cards	WT *HyPB* expression plasmid
ERα N-terminal fusion protein Calling Cards	*ER*α-*HyPB* N-terminal expression plasmid
ERα C-terminal fusion protein Calling Cards	*ER*α-*HyPB* C-terminal expression plasmid
Empty Vector	pUC19

Proteins were resolved on 4-20% Mini-PROTEAN TGX gels (BioRad, PN: 4561095) alongside a PAGE Ruler Plus ladder (BioRad, PN: 526619). Electrophoresis was performed in 1x Tris/Glycine buffer (BioRad, PN: 1610737), initiated at 40 V for 10 min and continued at 170 V for 40-50 min. The gels were then transferred to PVDF membranes using iBlot 2 stacks (ThermoFisher, PN: IB24001) according to the manufacturer's instructions.

Membranes were blocked in either 4% milk (Marvel) or 3% BSA (Sigma, PN: B4287), with wash buffers supplemented with 0.01% Tween20 (Sigma, PN: P2287). Antibody protocols, including dilutions, blocking conditions, and wash steps, are summarised in Table 2.7. Detection was performed using ECL reagent (ThermoFisher, PN: 32109) and was imaged on the iBright system (ThermoFisher). After anti-ERα blotting, membranes were washed, for GAPDH probing was carried out following the ER antibody protocol but omitting the overnight block, treating membranes as freshly transferred.

Table 2.7*:* A list of primary and secondary antibodies (Ab), their concentrations, buffers used, and protocols followed after iBlot2 transfer.

Primary Antibody and dilution	**Steps after the iBlot transfer**	**Secondary Antibody**	**Blocking Buffer**	**Wash Buffer**
Estrogen Receptor Alpha *(Abcam, PN: ab3575)* *1: 1,000*	- 1-hour Blocking Buffer - Overnight Primary Ab Incubation (4°C) - 1 hour Wash Buffer - 1-hour Secondary Ab Incubation - 2 minutes Blocking Buffer - 30 minutes Wash Buffer	Anti-Rabbit Secondary Ab (Jackson ImmunoResearch, PN: 111-035-144) 1: 15,000	4% Milk in PBS	4% Milk in PBS + 0.01% Tween20
Estrogen Receptor Alpha *(Santa Cruz, PN: sc-8002)* *1: 600*	- Overnight Blocking Buffer (4°C) - 1-hour Primary Ab Incubation - 1 hour Wash Buffer - 1-hour Secondary Ab Incubation - 2 minutes Blocking Buffer - 30 minutes Wash Buffer	Anti-Mouse Secondary Ab (Jackson ImmunoResearch 115-035-174) 1: 10,000	3% BSA in PBS	3% BSA in PBS + 0.01% Tween20

GAPDH *(Proteintech, PN:10494-1-AP)* *1: 5,000*	- 1 hour at room temperature - 1-hour Primary Ab Incubation - 4x 15 minutes Wash Buffer - 1-hour Secondary Ab Incubation - 2 minutes Blocking Buffer - 2x 5 minutes Wash Buffer	Anti-Rabbit Secondary Ab (Jackson ImmunoResearch, PN: 111-035-144) 1: 10,000	4% Milk in PBS	4% Milk in PBS + 0.01% Tween20

Western blotting of Kozak-mutated constructs was performed as described above, with antibody conditions detailed in Table 2.7 except for ab3575, which was not used.

2.3.2. HyPB fusion partner functional validation with TdTomato and BrokenHeart

Lipofectamine 3000 transfections were performed as described in Section 2.2.2, using plasmid combinations shown in Table 2.8. For BrokenHeart assays, cells were imaged at 24 and 48 hours to assess co-transfection efficiency through confirmation of red fluorescence in MCF7 cells. TdTomato assays required longer-term culture; cells were therefore reseeded after 24 hours at a 1:5 ratio into 24-well plates to promote clonal expansion. Fluorescence imaging was performed from 48 h post-transfection and every 48 h for one week to determine if red fluorescent colonies formed over time.

Table 2.8*:* Plasmids used to generate transfection reagents for functional validation of fusion proteins using the BrokenHeart and TdTomato SRTs

Transfection condition name	**Plasmid 1 (500 ng)**	**Plasmid 2 (500 ng)**
Wild Type (WT) + BrokenHeart Calling Cards	WT *HyPB* expression plasmid	BrokenHeart Self-reporting Transposon
ERα N-terminal fusion protein Calling Cards	*ER*α-HyPB N-terminal expression plasmid	BrokenHeart Self-reporting Transposon
ERα C-terminal fusion protein Calling Cards	*ER*α-HyPB C-terminal expression plasmid	BrokenHeart Self-reporting Transposon
BrokenHeart OR TdTomato control + pUC19	BrokenHeart OR TdTomato Self-reporting Transposon	pUC19
Positive control	pmaxGFP	pUC19
Negative control (no lipofectamine)	pUC19	pUC19

Later optimisation of co-transfection efficiency using BrokenHeart was conducted by seeding MCF7 cells between 0.35-0.55 10^6 cells/well in 6-well plates (Corning, Ref: 3516), with each density tested in duplicate. Transfections were performed as described in Section 2.2.2, using plasmid combinations listed in Table 2.9. Co-transfection efficiency was assessed by BrokenHeart reporter activity.

Table 2.9: A table of Transfection conditions used to generate transfection reagents for seeding optimisation using the BrokenHeart SRT

Transfection condition name	**Plasmid 1 (500 ng)**	**Plasmid 2 (500 ng)**
Wild Type (WT) + BrokenHeart Calling Cards	WT HyPB expression plasmid	BrokenHeart Self-reporting Transposon

2.4. ERα Calling Card Sequencing

2.4.1. Bulk ERα Calling Cards

2.4.1.1. Cell seeding and transfection

MCF7 cells were harvested and counted as described in Section 2.2.1 so they could be seeded into 6-well plates at densities between 0.45-0.55 $x10^6$ cells/well. Plates were tilted forward, backwards, and side-to-side to distribute cells evenly before being left to settle for 15 minutes in the biosafety cabinet. Once the cells had settled, the plate was returned to the incubator.

Lipofectamine 3000 transfections were performed as described in Section 2.2.2, using plasmid combinations listed in Table 2.10 (Chapter 2) and
Table **2.11** (Chapter 3). These included wildtype HyPB, N- and C-terminal HyPB-ERα fusion constructs, puromycin self-reporting transposons, and appropriate controls.

Table 2.10*:* Plasmid used to generate transfection reagents for bulk calling card library generation to validate redirection of HyPB insertions by ERα

Transfection condition name	**Plasmid 1 (500 ng)**	**Plasmid 2 (500 ng)**
Wild Type (WT) + Puromycin Calling Cards	WT *HyPB* expression plasmid	Puromycin Self-reporting Transposon
ERA N-terminal fusion protein Calling Cards	*ERA-HyPB* N-terminal expression plasmid	Puromycin Self-reporting Transposon
ERA C-terminal fusion protein Calling Cards	*ERA-HyPB* C-terminal expression plasmid	Puromycin Self-reporting Transposon
Puromycin control + pUC19	Puromycin Self-reporting Transposon	pUC19

Table 2.11: A table of Transfection conditions used to generate follow-up bulk calling card libraries with greater RNA yield, particularly for N-terminal transfection cells

Transfection condition name	**Plasmid 1 (500 ng)**	**Plasmid 2 (500 ng)**
Wild Type (WT) + Puromycin Calling Cards	WT HyPB expression plasmid	Puromycin Self-reporting Transposon
ERA N-terminal fusion protein Calling Cards	ERA-HyPB N-terminal expression plasmid	Puromycin Self-reporting Transposon
ERA C-terminal fusion protein + Kozak Sequence Calling Cards	ERA-HyPB C-terminal + Kozak Sequence expression plasmid	Puromycin Self-reporting Transposon
Puromycin control + pUC19	Puromycin Self-reporting Transposon	pUC19
No Transfection control	None	None

2.4.1.2. Post-transfection culture and puromycin selection

After 24 hours post-transfection, each well was split into its own 10 cm dish (VWR, PN: 10062880). Cells were detached and washed in the same way as Section 2.2 before transferring into labelled dishes. Residual cells were recovered by rinsing wells twice with

media and pooled into the corresponding dishes. Plates were swirled in a figure-eight pattern to ensure even distribution before incubation.

Once cultures reached ~80% confluence, the culture media were replaced with DMEM + 10% FBS supplemented with puromycin (Sigma, PN: P8833) at 2 µg/mL. The selection media was refreshed every 2-3 days, and selection continued until all puromycin-only controls had died (typically 7 days).

2.4.1.3. RNA harvest and quality assessment

Cells from each condition were harvested independently to avoid cross-contamination. Detached cells were pelleted at 400 xg, washed in PBS (Fisher, PN: 10209252), and resuspended in 1 mL PBS before RNA extraction. RNA was purified using the NEB RNA extraction kit (PN: T2010), following the manufacturer's instructions. The Purified RNA was eluted in 50 µL nuclease-free water (VWR, PN: PD092) and stored at -80 °C. RNA integrity and concentration were assessed using Agilent high-sensitivity bioanalyzer chips (PN: 50671513).

2.4.1.4. cDNA generation and SRT amplification

Two µg of RNA were reverse transcribed using SuperScript IV (Thermo, PN: 18090050) with SMART_dT18VN primers and dNTPs (Thermo, PN: 18427013), RT reagents in Table 2.12 were combined with annealed RNA and incubated at 55°C for 10 minutes, followed by 80°C for 10 minutes. cDNA was treated with 1 microliter RNase H (Cell Signalling, PN: 7013) cDNA was treated with RNase H (Cell Signalling, PN: 7013) before amplification.

Table 2.12*:* Reverse transcription mix components

Reagent	**Volume (µL)**
5x SSIV buffer (Thermo, PN: 18090050)	4
SuperScript IV (Thermo, PN: 18090050)	1
RNaseOUT (Thermo, PN: 10777019)	1
DTT (100 mM)	1

SRT amplification was performed using PCRBIO HS VeriFi Mix Red (PCRBIO Systems, PN: PB10.4701) with primers listed in the Appendix Table 8.7. Reaction composition is shown in Table 2.13 and cycling conditions in Table 2.14. Control reactions amplified actin. PCR products were visualised on 1% agarose gels in 1x TAE buffer, stained with SYBR Safe (Invitrogen, PN: S33102), and compared against DNA Ladder II (PCRBIO Systems, PN: PB40.1205).

Table 2.13: Bulk Calling Cards Amplification PCR components

Reagent	**Volume (µL)**
2X PCRBIO HS VeriFi Mix Red (PCRBIOSTSTEMS, PN: PB10.47-01)	12.5
DI Water	9.5
10 µM Forward Primer (SRT_PAC_F1)	1
10 µM Reverse Primer (SMART)	1
cDNA	1

Table 2.14*:* Bulk Calling Cards Amplification PCR programme

PCR Step	**Temp (ºC)**	**Time (s)**
Initial denaturation	95	180
Amplification Cycles (20 total)	98	20
	65	30
	72	300
Final extension	72	600
Hold temp (ºC)	4	Forever

2.4.1.5. PCR cleanup and quantification

PCR products were purified using AMPure beads (Beckman Coulter, PN: A63880) at a 0.6 ratio. Samples were washed twice with 70% ethanol and eluted in 20 µL nuclease-free water. The DNA concentration was measured using Agilent high-sensitivity DNA chips (PN: 50674626).

2.4.1.6. Final library preparation

Libraries were prepared using the Nextera XT kit (Illumina, PN: FC1311024) by the University of York Technology Facility Genomics Team via the following method.

Tagmentation and PCR amplification were performed according to the manufacturer's instructions, with barcoded piggyBac primers and indexed Nextera N7 primers assigned uniquely to each replicate. Cycling conditions are shown in Table 2.15. Libraries were cleaned with AMPure beads at a 0.7 ratio and quantified using Agilent D5000 screen tape (PN: 50675588) on the TapeStation. Final libraries were sent to Novogene for sequencing.

Table 2.15*:* Final Bulk Calling Card Library Amplification PCR Programme

PCR Step	**Temp (ºC)**	**Time (s)**
Initial denaturation	95	180
Amplification Cycles (13 total)	98	10
	50	30
	72	30
Final extension	72	300
Hold temp (ºC)	4	Forever

2.4.2. Long-Read ERα Calling Cards

2.4.2.1. General calling cards workflow

The long-read calling card assay was carried out in the same way as bulk calling cards (Section 2.4.1.1 and 2.4.1.2), with modifications. The seeding density of MCF7 cells was changed to 0.45 x10^6 cells/well in 6-well plates compared to 0.55 x10^6 in bulk assays. During transfection, only the undirected HyPB and the C-terminal HyPB-ERα constructs were tested (Table 2.16). Finally, the puromycin selection concentration was reduced to 0.5 µg/mL to improve the viability of co-transfected cells. The cells were cultured until puromycin SRT-only controls were dead (after 6 days), for harvest and single-cell barcoding.

Table 2.16: A table of Transfection conditions used to generate single-cell long-read calling card libraries.

Transfection condition name	**Plasmid 1 (500 ng)**	**Plasmid 2 (500 ng)**
Wild Type (WT) + Puromycin Calling Cards	WT HyPB expression plasmid	Puromycin Self-reporting Transposon
ERA C-terminal fusion protein Calling Cards	ERA-HyPB C-terminal expression plasmid	Puromycin Self-reporting Transposon
Puromycin control + pUC19	Puromycin Self-reporting Transposon	pUC19
No Transfection control	None	None

2.4.2.2. Preparation for 10X Chromium GEM encapsulation and single-cell barcoding

Cells were carefully harvested after puromycin selection and were gently mixed with a wide-bore pipette tip to ensure homogeneity. Each replicate was counted with a hemocytometer (Marienfeld, PN: 8100103) as described in Section 2.2. Suspensions were further mixed 10-15 times with regular pipette tips and passed through Flowmi cell strainers (Fisher, Ref: 15342931) to remove clumps.

Cells were centrifuged at 300 xg for 5 minutes, and the supernatant was carefully removed without disturbing the pellet. All removed supernatant was retained in a separate tube to recover lost cells if required. Pellets were washed in 1 mL PBS (Fisher, Ref: 10209252) + 1%

BSA (Sigma, Ref: A147025G), followed by a second centrifugation at 300 xg for 5 min. After the wash, ~50 µL of supernatant was deliberately left in the tube to maintain cell concentration.

Pellets were resuspended thoroughly but gently in the remaining supernatant using regular-bore tips to ensure clumps were broken and single-cells remained. Cell counts were repeated with a hemocytometer, and if counts were too low, the saved supernatant was spun down at 450 xg for 5 minutes to recover additional cells. Volumes were adjusted to achieve a final concentration of 700-1200 cells/µL in 25 µL, corresponding to ~10,000 cells per replicate. Once the target concentration was reached, suspensions were kept on ice and immediately processed by the University of York Technology Facility Genomics Team using the Chromium Next GEM Single-cell 3 Reagent Kits v3.1 Dual Index (10X, Ref: PN1000268). The Genomics team carried out all steps from this point forward.

2.4.2.3. scRNA-seq library preparation

GEM generation and barcoding were performed using the Chromium Next GEM Single-cell 3 Reagent Kits v3.1 Dual Index (10X, Ref: PN1000268), following the manufacturer's protocol with one modification: the RT primer was replaced with an equivalent volume of Low TE Buffer (ThermoFisher, Ref: J75793.AE). GEMs were loaded onto the controller, and reverse transcription was carried out under the conditions shown in Table 2.17.

Table 2.17: First strand synthesis reverse transcription programme following GEM encapsulation.

PCR Step	**Temp (ºC)**	**Time (mins)**
Lid temperature	53	N/A
Extension	53	45
Melting	85	5
Hold temp (ºC)	4	Forever

The GEM-RT mixture was cleaned using Recovery Agent and DynaBeads MyOne Silane (ThermoFisher, Ref: 37002D). At the final elution stage, 36.5 µL of Elution Solution I was added, incubated for 1 minute at room temperature, and placed on a 10X Magnetic Separator

until clear. 36 µL of eluate was recovered and split into two 18 µL aliquots: one for scRNA-seq library preparation, the other for single-cell calling card (scCC) libraries.

2.4.2.4. Single-cell calling card (scCC) library preparation

One aliquot of the GEM-RT product was processed for scRNAseq. Reverse transcription reagents are listed in Table 2.18 and cycling conditions in Table 2.19. Cleanup, cDNA amplification, and library construction followed the standard 10X protocol (Steps 2.23.7). Libraries were quantified on Agilent TapeStation High Sensitivity D1000 ScreenTape (Ref: 50675584) and pooled for sequencing. Sequencing was performed on a NovaSeq X with paired-end 150 bp reads, trimmed to 28 bp (read 1), 90 bp (read 2), and 10 bp index reads for downstream analysis.

Table 2.18*:* Post-GEM scRNA reverse transcription reagents

Reagent	**Volume (µL)**
Maxima 5X RT buffer (Fisher, Ref: 10899010)	20
20% w/v Ficoll PM-400 (Sigma, Ref: F4375)	20
10 mM dNTPs (ThermoFisher, Ref: 18427-013)	10
RNaseOUT (ThermoFisher, Ref: S10777-019)	2.5
500 µM 10x_TSO	2.5

Table 2.19: Reverse transcription programme for transcriptomic RNA

PCR Step	**Temp (ºC)**	**Time (mins)**
Annealing	25	30
Extension	50	90
Melting	85	5
Hold temp (ºC)	4	Forever

The second aliquot was used for scCC library preparation. Primer cocktails for long-read and short-read calling cards (Table 2.20 and Appendix Table 8.8, respectively) and PCR mixes (

Table ***2.21***) were assembled, and amplification was performed under the programme in Table 2.22. PCR products were cleaned using AMPure beads (Beckman Coulter, PN: A63880) as described in Section 2.4.1.5.

Table 2.20*: Follow-up up Calling card SRT amplification primer cocktail*

Reagent	**Volume (µL)**
100 µM IlluminaSeq1_scCC_10X_3xPT primer	5
100 µM Universal_Long_PB_LTR primer	5
Low TE Buffer (ThermoFisher, Ref: J75793.AE),	10

Table 2.21*:* Calling card SRT amplification PCR mix

Reagent	**Volume (µL)**
2X Kapa HiFi Hotstart Readymix (Roche, Ref: KK2601)	25
10x First-strand synthesis product	18
ddH2O	6
Primer cocktail from *Table 2.20*	1

Table 2.22: Calling card SRT amplification PCR

PCR Step	**Temp (ºC)**	**Time (s)**
Initial denaturation	98	180
Amplification Cycles (20 total)	98	20
	67	30
	72	300
Final extension	72	600
Hold temp (ºC)	4	Forever

2.4.2.5. Long-read sequencing

Final scCC PCR products were prepared with the ONT NBD114.24 native DNA barcoding kit (Oxford Nanopore Technologies, Ref: SQKNBD114.24) according to the manufacturer's instructions. Sequencing was performed on R10.4 flow cells (FLOPRO114M) using a PromethION 24. Runs were controlled with MinKNOW (Version 24.06.14), and basecalling was carried out with Dorado (Version 7.4.13) in super accuracy mode with barcode trimming enabled.

2.5. Computational analysis

2.5.1. Bulk ERα Calling Card pipeline (Chapter 3)

Raw FASTQ files were adaptor-trimmed with cutadapt (Version 4.2) (Martin, 2011) and aligned to the UCSC hg38 reference genome using Samtools (Version 1.16.1-GCC-11.3.0)(Li et al., 2009) before conversion to BAM format. BAM files were indexed and processed with calling card-specific scripts (Moudgil et al., 2020) to assign primer barcodes and indices, confirm integration at TTAA tetranucleotides, and annotate adjacent sequences. BAM files were converted to qBed format for peak calling with PyCallingCards (Version 1.0.0). As CC peaks can be inserted within 1000 bp of a TF binding site, Bedtools (2.30.0-GCC-11.2.0) window was used to expand calling card peaks by 1000 bp at both ends. Overlap with ChIP-Seq and ChIA-PET data was then carried out using Bedtools intersect. The number of overlapping CC peaks was divided by the total number of CC peaks and multiplied by 100 to calculate the percentage of calling card peaks that overlap with ChIP-Seq. This method does not consider TTAA motifs present in the region other than at the insertion site, but this is beyond of the scope of this work and should be investigated in the future. Downstream analysis was performed in Bash and R using Bedtools (Version 2.30.0-GCC-11.2.0)(Quinlan and Hall, 2010) and custom scripts (Stenning, 2026). Supporting ChIP-Seq datasets were downloaded from ENCODE and processed in parallel for comparison.

2.5.2. Long-read ERα Calling Card pipeline (Chapter 4)

FASTQ files were processed using a modified protocol adapted for nanopore sequencing. Reads were either (i) adaptor trimmed in two steps with cutadapt (Version 4.2) or (ii) passed directly to alignment and designated as “full-length” datasets. To standardise orientation, all reads were reformatted so that the HyPB terminal repeat sequence was positioned at the beginning of the read and then removed to facilitate genomic alignment and compatibility with existing calling card scripts (Moudgil et al., 2020). Trimmed and untrimmed reads were aligned to hg38 using the NexFlow (Di Tommaso et al., 2017) Epi2Me single-cell pipeline (Version 23.04.2). BAM outputs were processed with the annotation code from the published calling card protocol, converted to qbed, and peaks were called with PyCallingCards (Version 1.0.0). As CC peaks can be inserted within 1000 bp of a TF binding site, Bedtools (2.30.0-GCC-11.2.0) window was used to expand calling card peaks by 1000 bp at both ends. Overlap with ChIP-Seq and ChIA-PET data was then carried out using Bedtools intersect. The number of overlapping CC peaks was divided by the total number of CC peaks and multiplied by 100 to calculate the percentage of calling card peaks that overlap with ChIP-Seq. This method does not take into account TTAA motifs present in the region other than at the insertion site, but this

is beyond of the scope of this work and should be investigated in the future. Multiplex information was not used, and samples were treated as pseudobulk datasets. Downstream analysis was performed in Bash and R using Bedtools (Version 2.31.0-GCC-12.3.0) and custom scripts (Stenning, 2026).

2.5.3. Downstream statistical analyses

To ensure comparability across both bulk and longread ER calling card datasets, we applied a consistent suite of downstream analyses.

Genome-wide reproducibility (Bulk Chapter 3 only): Reproducibility between biological groups was assessed using Pearson correlation coefficients. Statistical significance of correlations was evaluated using the rcorr function (Hmisc package, R), which computes two-tailed p-values for each pairwise correlation.

Peak overlap with independent datasets: To assess the statistical significance of overlap between ERα calling card peaks and independent datasets (ERα ChIP-Seq, BRD4 ChIP-Seq, and ERα ChIA-PET), we performed hypergeometric testing. The universe was defined as all canonical TTAA sites in hg38. For each dataset, TTAA sites overlapping peaks were enumerated (n for calling card, K for ChIP/ChIA-PET), and the observed intersection (k) was compared to the expectation under random sampling. P-values were calculated using the cumulative hypergeometric distribution, and fold enrichment was defined as (k/n)/(K/N). Fisher's exact test was additionally applied to confirm robustness. All analyses were performed in R using the GenomicRanges and IRanges packages (Lawrence et al., 2013).

Signal enrichment at calling card peaks: To assess whether the ChIP-Seq and ChIA-PET signal was enriched at calling card peaks, we extracted the mean ChIP-Seq signal intensity from bigWig tracks within 1 kb of each peak. ChIA-PET bigWigs were downloaded from the UCSC Genome Browser (Fullwood et al., 2009; Li et al., 2014). For background, we generated three shuffled peak sets matched for chromosome distribution and peak width. Observed signal distributions were compared to shuffled controls using Wilcoxon rank-sum tests, and empirical p-values were calculated by comparing the observed mean signal to the distribution of shuffled means. Signal extraction was performed using rtracklayer (Lawrence, Gentleman and Carey, 2009) and GenomicRanges in R.

2.6. Spheroid culture and hypoxia modelling

2.6.1. General workflow

MCF7 cells were seeded into ULA 96-well plates (Corning, P/N: 10023683) at defined densities in 200 µL of medium per well, with the values used defined for each experiment. After seeding cells, the ULA plates were centrifuged at 120 xg for 5 minutes to concentrate cells at the bottom of the well before incubation at 37 °C. The culture media were refreshed every 2-3 days by replacing 80 µL of media with fresh media in each well. Imaging was performed using a Zeiss Observer 7 microscope (Zen2 Blue v3.9.101, University of York Imaging and Cytometry Facility), and spheroid diameters were measured in ImageJ (Version 2.16.0, Fiji v1.54P).

2.6.2. Seeding density optimisation

To determine the optimum density for spheroids of 200-500 µm diameter after 7 days, cells were seeded at 500, 1000, 5000, and 10,000 cells/well across 12 wells of a ULA 96-well plate (Corning, P/N: 10023683), following the procedure in Section 2.6.1. Spheroids were imaged daily for 7 days, and diameters were measured in ImageJ.

2.6.2.1. Defining standard media composition

To determine the influence of media components on spheroid formation, phenol red-free DMEM (Gibco, Ref: 11054020) was supplemented with 1x Glutamax (Gibco, Ref: 35050 0038) and B27 (Gibco, Ref: 17504044). Phenol red-containing DMEM (Gibco, P/N: 11995 065) with 10% FBS was tested in parallel. Cells were seeded as described in Section 2.6.1, according to the layout in Figure 2.1. Spheroids were imaged after 7 days, and diameters were measured in ImageJ.

Figure 2.1: Ultra-low attachment 96-well plate layout for phenol Red-free DMEM spheroids

Each condition was seeded twice across two rows of a 96-well plate, one with 500 and the other with 1000 cells/well as indicated by the colour of the text boxes across plate rows. An additional plate was made that seeded 500 and 1000 cells/well in Phenol Red DMEM without supplements.

2.6.3. Validating MCF7 spheroids as a hypoxia model

2.6.3.1. Cell culture conditions

Spheroids were seeded at 1000 cells/well as described in Section 2.6.1 and were maintained for 13 days. On days 7, 9, 11, and 13, replicate rows were harvested: one treated with 200 µM EF5 (Sigma, P/N: EF5-30A4M) for 4 hours before fixation, the other with EF5 vehicle control.

2.6.3.2. Fixation and antigen retrieval

Spheroids were collected in a wide-bore pipette tip and pooled into a single tube per condition. Samples were then washed twice in PBS (Fisher, PN: 10209252), and fixed in 4%

paraformaldehyde (ThermoFisher, P/N: 43368.9M) for 1 hour at room temperature. Following fixation, the spheres were washed and incubated overnight at 4°C in 1x sodium citrate antigen retrieval buffer (Merck, P/N: C9999). A fresh aliquot of buffer was pre-warmed to 95°C, and a 30% sucrose solution (Fluorochem, P/N: F492423) was pre-chilled to 4°C. Antigen retrieval was performed the next day with a heat shock step with pre-heated sodium citrate buffer at 95 °C for 10 minutes, followed by incubation in cooled 30% sucrose (Fluorochem, P/N: F492423) at 4 °C for 2 hours.

2.6.3.3. Embedding in freezing medium

Spheroids were embedded in Polyfreeze blue medium (Sigma, P/N: SHH0025120ML) in small cryomolds (PolySciences, P/N: 25374500) and frozen in a dry ice/propan-2-ol slurry. Cryoblocks were sectioned at 10 µm thickness on a Leica CM1950 cryostat and mounted on Superfrost Plus slides (ThermoFisher, P/N: J1800AMNZ).

2.6.3.4. Cryosectioning

Cryostat chamber and specimen stage temperatures were set to 26 °C and 15 °C, respectively, before sectioning. Sample blocks were transferred on dry ice to the Leica CM1950 cryostat and equilibrated in the chamber for at least 30 minutes before processing. Freezing medium was applied to the chuck, and the embedded sample block was mounted with the mould base facing upwards.

Sections were cut at 10 µm thickness using a cryotome blade and were transferred to Superfrost Plus glass slides (ThermoFisher, P/N: J1800AMNZ). Spheroid-containing sections were identified with light microscopy during cutting. To maximise reproducibility and allow comparative staining, sequential sections were alternated between two groups of four per slide, positioned on opposite sides with a central gap. This strategy permitted staining with two different AF488-conjugated antibodies on the same slide while controlling for section variability. Completed slides containing eight sections were stored on dry ice in slide boxes for storage at -80 °C or processed immediately for immunostaining.

2.6.3.5. Immunostaining with conjugated antibodies

Frozen slides were equilibrated to room temperature for 1 hour before cryosection areas were circled with a hydrophobic barrier pen (Merck, P/N: Z6725481EA). The slides were washed

with 200 µL PBS (Fisher, PN: 10209252) within the marked area; this volume was used consistently for all subsequent steps unless otherwise stated.

Sections were incubated with 0.1 M glycine (Fisher, PN: 10209252), pH 7.4, in PBS for 15 minutes, followed by a PBS wash. Samples were permeabilised in 0.2% Triton X-100 (Sigma, P/N: X100) for 5 min, then washed three times in PBS. Blocking was performed in 5% BSA (Sigma, P/N: B4287) in PBS for 1 hour at 4 °C in a humidified chamber.

After blocking, sections were incubated overnight at 4 °C in 1% BSA in PBS containing primary antibodies: Anti-EF5 (Sigma, P/N: EF530A4M, 75 µg/mL), Anti-ERα (Abcam, P/N: ab93021250ul, 1:20), or Anti-CAIX (Fisher, P/N: 18797463, 1:350). Negative and secondary-only controls omitted the primary antibody. Slides were protected from light from this stage onward to prevent photobleaching, as Anti-EF5 is directly conjugated to AF488.

The following day, primary antibody solutions were removed, and sections were washed three times in PBS. Secondary antibodies were applied for 1 hour at room temperature: anti-mouse (ThermoFisher, P/N: A11001, 1:2000) or anti-rabbit (ThermoFisher, P/N: A11001, 1:1000) in 1% BSA in PBS. Negative controls omitted the secondary antibody. After incubation, slides were washed three times in PBS before adding ProLong Gold antifade reagent (ThermoFisher, P/N: 15260719). No. 1.5 coverslips (Expredia, P/N: SGJ3030) were placed carefully to avoid air bubbles. Slides were cured in the dark for 24 hours before short-term storage at 4 °C or long-term storage at 20 °C.

2.6.4. Treating MCF7 spheroids with Amiloride

MCF7 spheroids were seeded at 1000 cells/well in two ULA 96-well plates (Corning, P/N: 10023683) using the method described in Section 2.6.1 and were maintained in culture for 13 days. Amiloride (Sigma, P/N: A74101g) was prepared from a 100 mM stock in DMSO (Sigma, P/N: D2650) and serially diluted 1:10 to final concentrations ranging from 100 µM to 1 nM. Treatments were applied according to the plate layout in the Appendix Figure 8.18.

One plate was treated with Amiloride on day 2 to model intervention during spheroid formation. The second plate was treated on day 9 to model the treatment of established, hypoxic spheroids. On days 7 and 13, spheroids treated with 100 M Amiloride were harvested, exposed to EF5, fixed, cryosectioned, and immunostained as described in Section 2.6.3.

To minimise artefacts associated with plate position, spheroids cultured in the outer wells (rows A and H, columns 1 and 12) of 96-well plates were excluded from analysis. These wells consistently produced smaller spheroids compared to central wells, a phenomenon attributed to edge effects caused by differential evaporation and altered oxygen diffusion gradients across the plate (Mansoury et al., 2021; Grosch et al., 2016; Das et al., 2016; Maxwell et al., 2023; Hass et al., 2023; Al-Ani et al., 2018). Media depth variation in outer wells likely increased oxygen availability, confounding hypoxia development.

Growth curves generated from these edge wells showed an apparent increase in spheroid diameter at low amiloride concentrations close to the reported ENaC IC50 (~100 nM), consistent with literature suggesting ENaC inhibition can increase MCF7 proliferation (Ware et al., 2021). However, this apparent increase was not reproducible once edge wells were excluded, confirming that the observation was driven by plate artefacts rather than treatment response. For all analyses presented in the Results, only spheroids from central wells were included to ensure reproducibility and eliminate positional bias.

2.6.5. Image data analysis

Images were exported from the Zeiss Observer 7 microscope using Zen2 Blue (Version 3.9.101). Spheroid diameters were measured in ImageJ (Schneider, Rasband and Eliceiri, 2012)(Version 2.16.0, Fiji v1.54P (Schindelin et al., 2012)) as the maximum straight line across the spheroid surface; detached cells were excluded.

Differences in spheroid diameter between media conditions were assessed by one-way ANOVA followed by Tukey's post hoc test (Tukey, 1949).

For fluorescence analysis, ER and CAIX signals were weak and close to background, so contrast was enhanced by adjusting black/white values. Thresholds were determined using secondary-only controls to avoid amplifying the background signal. Default values in ImageJ are 0 (Black) and 16384 (White). For ER AF488 staining, values were adjusted to 250 (Black) and 6500 (White). For CAIX AF568 staining, the Black value was set to 400, with White remaining at 6500.

Further analysis was performed in RStudio (R Version 4.5.1; RStudio Version 2025.05.1+513 Mariposa Orchid for Windows) and CellProfiler (Version 4.2.6)(Stirling et al., 2021). ImageJ diameter measurements were processed and plotted in R. Fluorescence intensity was quantified in CellProfiler using the MeasureObjectIntensityDistribution function across

concentric bins from the spheroid centre to the surface. Outputs were processed in R using the same workflow as the diameter data.

For high signal analysis, CellProfilers' Threshold function was applied to remove regions with intensity <0.09 units. Remaining regions were measured for size and intensity, and only images with regions above this threshold were carried forward in the High Signal mask analysis. All processed datasets and analysis code are available on GitHub (https://github.com/JackStenning/Thesis).

3. Bulk HyPB-ERα Calling Cards demonstrates that our novel fusion protein can replicate ERα binding across the MCF7 genome

3.1. Introduction

We have explored how ERα is a central driver of breast cancer, coordinating transcriptional programmes through direct DNA binding and recruitment of co-regulators (Sandström et al., 2024; Lee et al., 2017a; Hoch et al., 1999; Fang, Chen and Weigel, 2009; Paakkola et al., 2021; Lumachi, Santeufemia and Basso, 2015). Mapping ERα binding patterns has traditionally relied on ChIP-Seq, which provides high-resolution binding profiles but requires large cell inputs and averages signal across populations, obscuring heterogeneity (Rotem et al., 2015; Kaya-Okur et al., 2019).

Recent advances, such as ChIA-PET, have revealed ERαs role in long-range chromatin looping (Theodorou and Carroll, 2010; Fullwood et al., 2009). These studies demonstrated enhancer-promoter contacts at ERα target loci, but even ChIA-PET and related methods, such as HiChIP, remain constrained in their ability to capture the binding dynamics of TFs at single-cell resolution (Li et al., 2017a, 2017b; Mumbach et al., 2016).

Calling card reporters, which use transposase-mediated DNA insertions to create a historical record of TF binding events, have emerged as a powerful alternative (Wang et al., 2012a; Moudgil et al., 2020; Lalli et al., 2022; Cammack et al., 2020). Previous studies have demonstrated their accuracy and reproducibility, with 83% of calling card insertions and 72% of called peaks overlapping within 1 kb of ChIP-Seq sites for multiple TFs (Wang et al., 2012a; Moudgil et al., 2020; Lalli et al., 2022; Cammack et al., 2020). This overlap was first quantified in yeast and mammalian systems, establishing calling cards as an alternative to ChIP-Seq for discovering TF binding sites. It is important to clarify however, calling cards report that a binding event occurred rather than providing a quantitative measure of binding strength, as insertion frequency reflects both biological and technical factors and cannot be interpreted as a direct proxy for occupancy. Nuclear receptors such as ERα have not yet been adapted for calling card analysis, leaving a gap in our ability to probe their binding heterogeneity.

The objective of this chapter is to engineer and validate the first nuclear receptor-based calling card reporter by using ERα to the HyPB. Both N- and C-terminal fusion constructs were generated, reflecting prior literature where C-terminal fusions are common (Appendix Figure

8.1)(Hayashi and McMahon, 2002; Kalaitzidis et al., 2004; Kurata et al., 1999; Powell and Xu, 2008; Wen et al., 1999; Shim et al., 1997; Kolligs et al., 2002), but N-terminal variants have also been reported (Zhou et al., 2014; Larson et al., 2021; Powell and Xu, 2008). Bulk ERα calling card experiments were performed to confirm that the fusion proteins can deposit insertions near genuine ERα binding sites.

To validate the system, calling card peaks were compared against the gold standard ChIP-Seq datasets, with additional analysis including motif enrichment, signal profiling, and overlap with ChIA-PET anchors. Standard tools such as PyCallingCard (Guo et al., 2024) and bedtools (Quinlan and Hall, 2010) were used for motif and overlap analysis. Peaks adjacent to known ERα target genes such as GATA3, FOXA1 and GREB1 were examined to confirm biological relevance (Cirillo et al., 2002; Asselin-Labat et al., 2007; Robinson and Carroll, 2012; Kong et al., 2011; Hodgkinson et al., 2018; Sun, Nawaz and Slingerland, 2007; Cheng, Michalski and Kommagani, 2018; Haines et al., 2020).

This chapter aims to engineer and validate the first nuclear receptor-based calling card reporter by fusing ERα to the HyPB, thereby establishing a bulk calling card system capable of mapping genome-wide ER binding in heterogeneous cell populations.

Objectives to achieve this aim are:

- Generate N- and C-terminal HyPB-ERα fusion constructs and assess their ability to deposit insertions near genuine ERα binding sites.
- Benchmark bulk ERα calling card peaks against gold-standard ChIP-Seq datasets to evaluate overlap, motif enrichment, and signal profiling.
- Assess the biological relevance of calling card peaks by examining localisation to known ERα target genes (e.g., GATA3, FOXA1, GREB1).
- Compare calling card data with ChIA-PET anchors to confirm compatibility with long-range chromatin interaction profiles.

Together, these analyses establish whether the HyPB-ERα fusion proteins can replicate endogenous ERα binding patterns and provide the foundation for subsequent single-cell calling card experiments described in the next chapter.

3.2. Results

3.2.1. The ERα coding sequence was cloned at the C-and N-terminus of the HyPB coding sequence to generate a fusion protein expression plasmid

To create ERα-HyPB and HyPB-ERα fusion proteins, we used a 3-step process. First, *ESR1* was cloned from a donor plasmid, followed by insertion into the pRM1258 WT-HyPB vector. Finally, these HyPB-ERα constructs were transferred into an expression vector using Gateway cloning.

The constructs, called the N- and C-terminal ERα fusion protein plasmids, are named based on the position of the ERα domain within the expressed fusion protein sequence. The N-terminal plasmid contains the ERα coding sequence at the beginning of the HyPB coding sequence, positioning the ERα protein partner at the N-terminus of the fusion protein following translation (ERα-HyPB). Whereas the C-terminal plasmid fusion protein coding sequence begins with the HyPB sequence, positioning the ERα partner of the fusion protein at the C-terminus (HyPB-ERα).

The coding sequence for ERα (UniProt Sequence #P03372 (Apweiler et al., 2004; Wu et al., 2006)) gene was cloned into the pRM1258 WT-HyPB vector in frame with frame with both ends of the HyPB coding sequence to generate the two HyPB-ERα fusion constructs. Working on similar principles to Gibson Assembly, NEBuilder HiFi assembly DNA inserts require a short sequence homologous to the vector to be added to the end of the *ESR1* gene to enable cloning. The ESR1 insert was prepared for subcloning through PCR, which included the addition of an amino acid linker sequence and overlapping homology with the HyPB coding region and surrounding DNA sequence in the pRM1258 WT-HyPB plasmid. The resulting NEBuilder HiFi assembly reaction product from the ESR1 insert and linearised pRM1258 WT-HyPB vector was then used to transform into Stbl3 competent E. coli.

The fusion protein plasmids were purified from these cultures the following day, and insertion was confirmed by restriction site mapping with SalI-HF, XhoI and HindIII-HF (Figure 3.1).

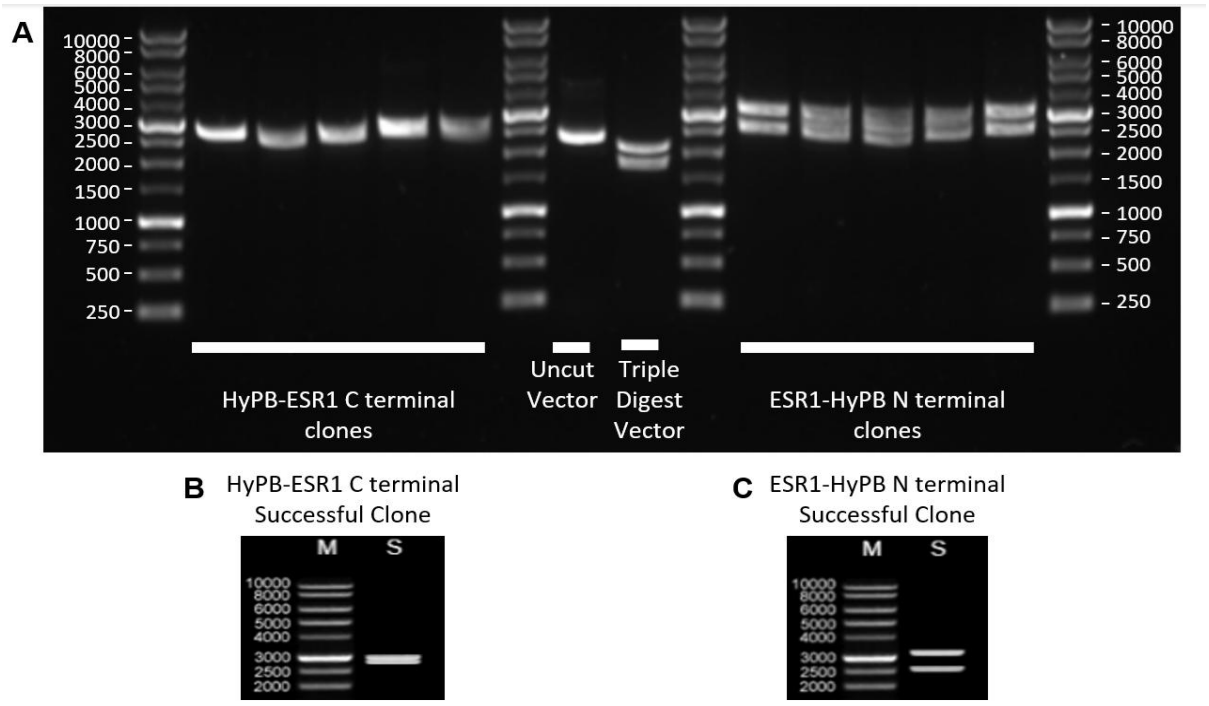


Figure 3.1: Restriction mapping with SalI, XhoI and HindIII on a 1% agarose gel shows successful insertion of the *ESR1* gene upstream and downstream of the *HyPB* gene in the vector plasmid 'pRM 1258 pENTR_myc-hypPBase'.
The *ESR1* gene was PCR amplified with a region encoding an 18 amino acid linker in addition to compatible ends to the 5' and 3' end of the *HyPB* gene for insertion after digestion with either SalI (5') or XhoI (3'). Ligation of the insert destroys the original restriction enzyme recognition site, resulting in digestion with both enzymes and HindIII, creating two fragments of different lengths. Potential clones were digested with all three enzymes and run on a 1% agarose gel **(A)** alongside the uncut original pRM 1258 pENTR_myc-hypPBase and the triple-digested vector. The vector produces three fragments when digested with all three enzymes; however, one of the fragments is ~50 bp and cannot be visualised on the gel. The expected size of fragments from successful clones was calculated using Snapgene for the 3' insertion generating the C-terminal fusion gene **(B),** as well as the 5' insertion generating the N-terminal fusion gene **(C)**. All potential clones created fragments of the expected sizes following digestion, indicating that all cloning reactions were successful.

Successful cloning was indicated by a ~1000 bp increase in the band produced by the clones than was produced from the corresponding empty vector (~2000 bp shifts to ~3000 bp). As all bands generated from the digested clones match the prediction, samples with the highest concentrations and A260/280 values were prioritised for Sanger sequencing.

Analysis of sequencing runs confirmed successful generation of vectors containing the ESR1 coding sequence inserted in the frame with HyPB and without mutation for all the fusion protein sequence variants. Together, these data show that we have determined the ESR1 coding

sequence had been successfully cloned in frame with the HyPB gene at the N- and C-terminal positions.

After confirmation of successful N- and C-terminal fusion cloning by Sanger sequencing, Gateway cloning was used to create an expression vector with the N- and C-terminal fusion coding sequence downstream of the CMV promoter to initiate expression. Plasmids were then recovered and Sanger sequenced to confirm that no mutations had been incorporated, and the N- and C-terminal fusion expression plasmids were constructed successfully (Figure 3.2).

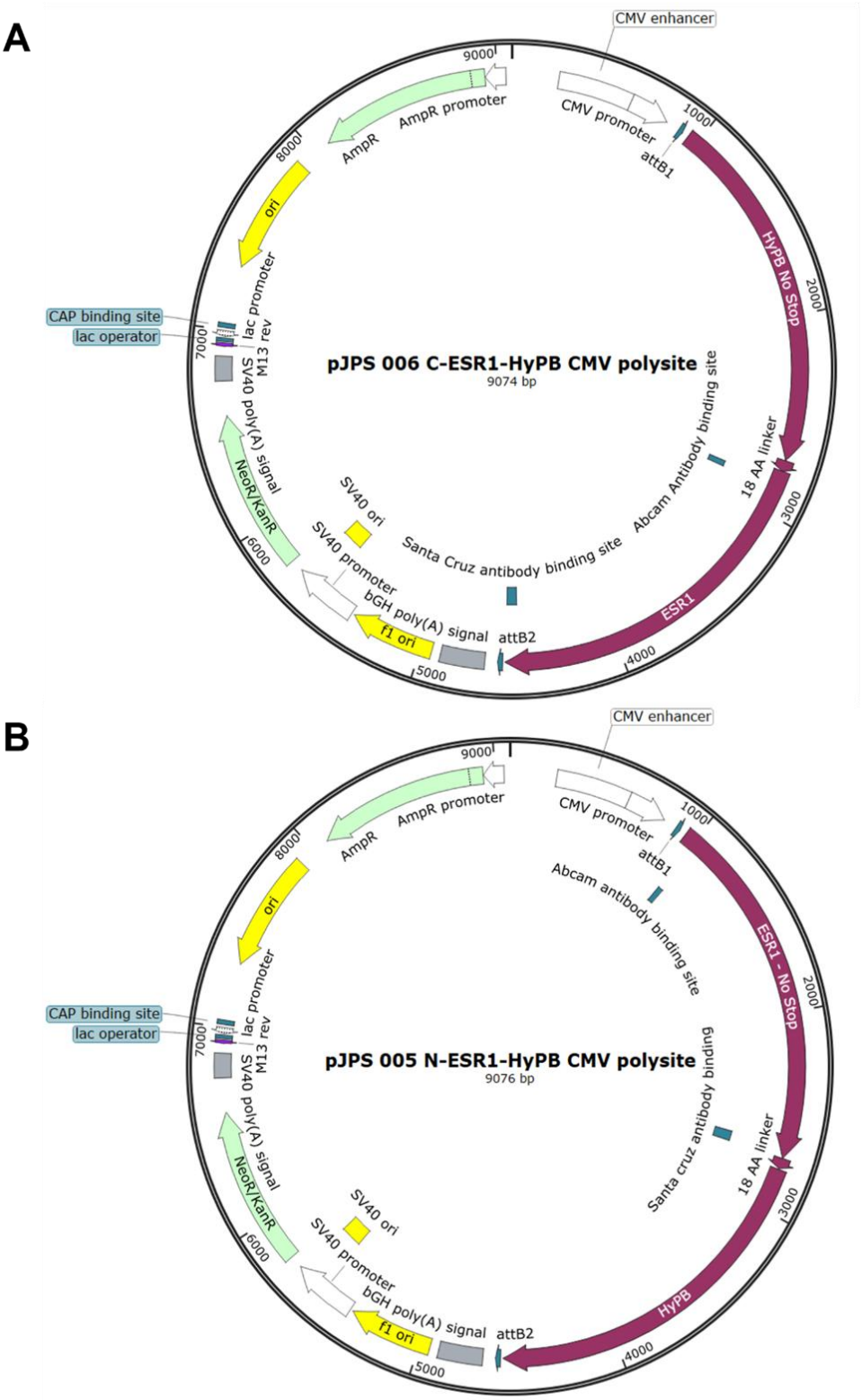


Figure 3.2: The predicted sequence between attB1 and attB2 gateway cloning sites in the N-and C-terminal fusion protein plasmids maps matched to Sanger-Sequencing of clones plasmids exactly, confirming successful expression plasmid generation. Plasmid maps were generated, predicting the expected sequence of successful clones for the C-terminal **(A)** and N-terminal **(B)** fusion protein genes. Sanger sequencing within the coding region of these genes matched exactly to the plasmid maps, showing that the cloning up and downstream of the *HyPB* gene was successful

3.2.2. Full-length HyPB-ERα fusion protein expression was confirmed via Western Blot following transfection into MCF7 and MDA-MB-231 cell lines.

Western blot analysis for the expression of ERα was carried out in both MCF7 and MDA-MB-231 to confirm expression of the HyPB-ERα fusion proteins following transfection (Figure 3.3). MCF7 cells were selected as this is the cell type that is carried forward in calling card experiments. MDA-MB-231 cells were selected as they do not express ERα and thus act as an endogenous ERα negative control. Both cell lines were transfected with both fusion protein plasmids, and lysates were collected for analysis. Two blots were carried out using two different primary antibodies, one that targets an epitope at the N terminus of the ERα (Figure 3.3 A) and another that targets the C-terminus (Figure 3.3 B). Full Western blots can be found in the Appendix Figure 8.2 and Figure 8.3.

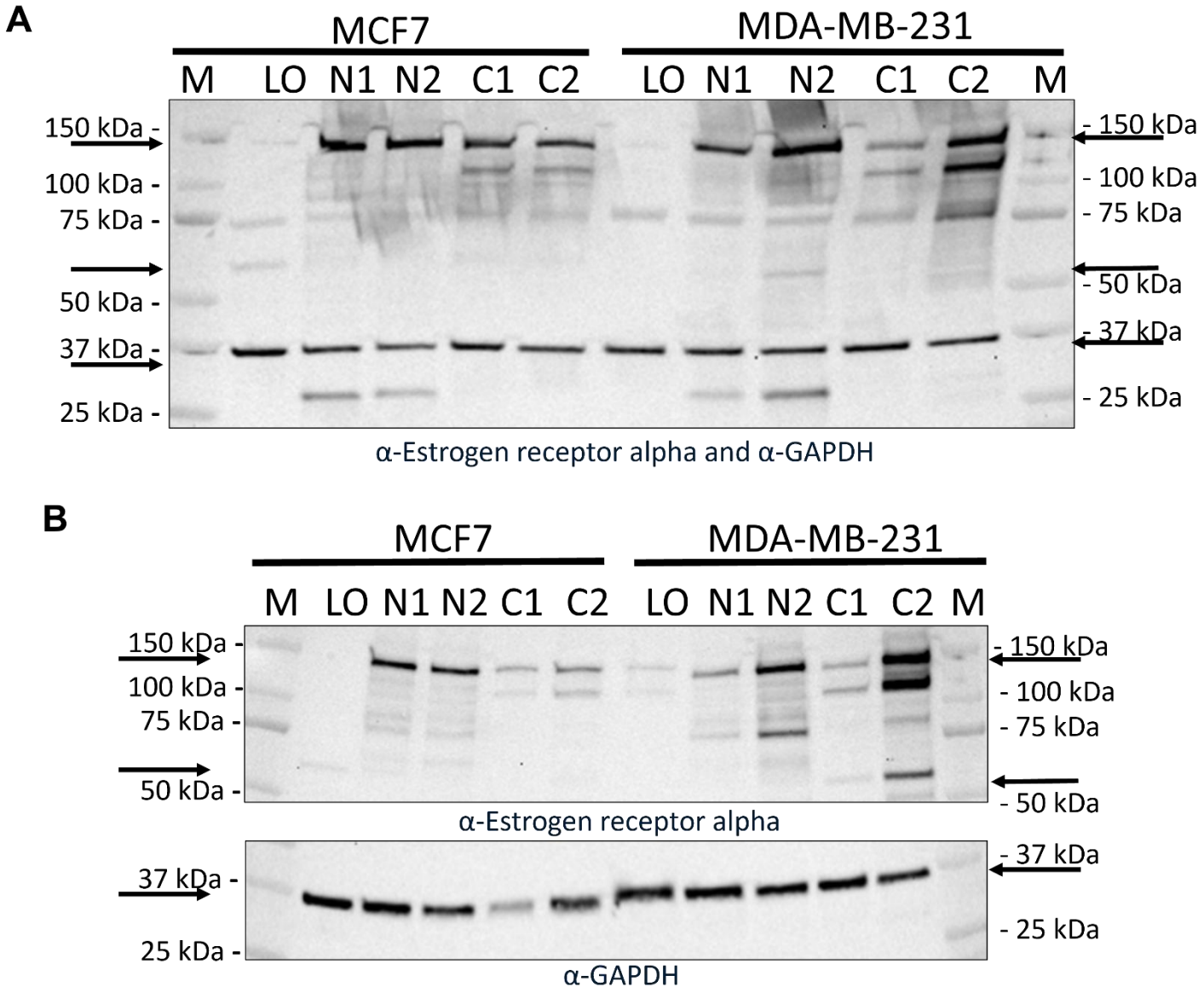

A
MCF7
MDA-MB-231
M LO N1 N2 C1 C2 LO N1 N2 C1 C2 M
150 kDa -
100 kDa -
75 kDa -
50 kDa -
37 kDa -
25 kDa -
- 150 kDa
- 100 kDa
- 75 kDa
- 50 kDa
- 37 kDa
- 25 kDa
α-Estrogen receptor alpha and α-GAPDH
B
MCF7
MDA-MB-231
M LO N1 N2 C1 C2 LO N1 N2 C1 C2 M
150 kDa -
100 kDa -
75 kDa -
50 kDa -
- 150 kDa
- 100 kDa
- 75 kDa
- 50 kDa
α-Estrogen receptor alpha
37 kDa -
25 kDa -
- 37 kDa
- 25 kDa
α-GAPDH

Figure 3.3: Fusion protein expression plasmids can express the full-length protein, as shown by the presence of a band of the expected weight on an α-ER (N-terminal) and α-ER (C-terminal) Western blot.

Following gateway cloning of C- and N-terminal fusion protein genes into expression plasmids, the plasmids were transfected into MCF7 and MDA-MB-231 cells and cell lysates were collected after 72 hours for Western blotting. **M** = Marker, **LO** = Lipofectamine only, **Nx** = N-terminal fusion protein transfection, replicate x, **Cx** = C-terminal fusion protein transfection, replicate x. **(A)** Was blotted with the N-terminal targeting Abcam α-Erα antibody (ab3575). The molecular weight of the full fusion protein is 136 kDa, as indicated by the first arrow; each calling card transfection in both bolts shows a band at this molecular weight, whereas the lipofectamine does not. There is a second, smaller band between 136 and 100 kDa in each C-terminal transfection that could indicate alternative translation initiation of the fusion construct. The second arrow indicates the endogenous ER band at 66 kDa. Although expression of this band is visible in the empty vector and N-fusion conditions, it is not apparent in the C-fusion conditions in either blot. The membrane was re-blotted with an α-GAPDH antibody as a loading control for this blot (band found at 37 kDa), demonstrating equal protein loading in each lane. α-ER (C-terminal). **(B)** Was blotted with the Santa Cruz α-ERα antibody (sc-8002). This blot is labelled in the same way as (A) and replicates the presence of the full-length fusion protein and endogenous ERα in calling card transfections. Interestingly, there appears to be induction of endogenous ER expression in both C-terminal fusion protein transfections in MDA-MB-231 cells. The membrane was re-blotted with an α-GAPDH antibody as a loading control for this blot (band found at 37 kDa), demonstrating equal protein loading in each lane, except for the C1 and C2 MCF7 lanes, in which there appears to be some misloading, as shown by the weaker band at 37 kDa.

Our western blot targeting the N-terminus of ERα showed that the N- and C-terminal HyPB-ERα are expressed in both cell lines post-transfection, identified as a band with a molecular weight of 136 kDa (Figure 3.3 A). A second, smaller band was only seen in the C-terminal HyPB-ERα transfections underneath the 136 kDa full-length protein band. It is unclear what this smaller band is and why it is specific to the C-terminal fusion protein; this band is addressed in the next chapter.

There are several other faint bands between 150 kDa and 75 kDa, some of which can be identified, while the origins of some bands are not established. The band at 74 kDa is in both the control samples and in MDA-MB-231 cells, which do not express ERα, with the Abcam antibody and should be considered off-target bands. As expected, the endogenous ERα band appears in the MCF7 lipofectamine control at 66 kDa but not in the MDA-MB-231 lysates. Finally, additional bands can be seen at approximately 28 kDa, but only in the N-terminal transfections; the origins of these bands were not established but will be discussed later. Overall, this blot confirms that the expression vectors can express the full-length fusion protein.

In the blot targeting the C-terminus of ERα, we also confirm the expression of the full-length fusion proteins across both cell lines (Figure 3.3 B). However, the C1 and C2 transfections in MCF7s are not as intense as other samples due to improper loading. Some of the samples from C2 were split into the MDA-MB-231 lipofectamine-only control. The smaller band at 66 kDa indicate the expression of endogenous ERα. Interestingly, it appears that a strong band around 66 kDa can be observed in the MDA-MB-231 cells when transfected with the C-terminal fusion protein. The band at 28 kDa is no longer apparent in the N-terminal transfections, but an additional band at 50 kDa can be observed, the origins of this band were not established, but will be discussed later.

These blots show that the *ESR1-HyPB* fusion protein expression plasmids are capable of full-length expression in MCF7 cells, confirming their ability to be used in downstream calling card experiments. With the establishment of full fusion protein expression, we proceeded with functional validation of the HyPB-ERα fusion proteins.

3.2.3. The HyPB-ERα fusion proteins are functionally active transposases.

The HyPB-ERα expression plasmids were individually co-transfected alongside the BrokenHeart transposon to functionally validate full-length fusion proteins through induction of red fluorescence in MCF7 cells. The BrokenHeart SRT contains a dsRed gene that is split by a HyPB transposon in the centre of the coding region. Therefore, in the presence of a functional transposase, the dsRed will be reactivated and cells will express the full-length protein.

Transfected cells were imaged after 48 hours with a fluorescent microscope, and all wells transfected with the HyPB-ERα fusion proteins contained red cells after a day in culture post-transfection (Figure 3.4 A). These results validate the function of the fusion protein HyPB partner, as the fusion proteins can reconstitute the full dsRed gene.

Following confirmation of transposase activity, we went on to demonstrate that the HyPB transposase activity results in the successful integration and inheritance of SRTs by daughter cells using the TdTomato SRT. The PiggyBac SRT cleavages are typically concomitant with genomic insertions (Chen et al., 2020; Mitra, Fain-Thornton and Craig, 2008; Yusa et al., 2011). However, this is not always the case (Nishizawa-Yokoi et al., 2014), and as a result, we carried out a further experiment using the TdTomato SRT. Unlike the BrokenHeart transfection, all cells transfected with the TdTomato SRT express the full-length protein

following transfection, regardless of HyPB status; resulting in red immunofluorescence, and after a week, transient expression of the TdTomato ceases due to loss of the plasmid from cell division. However, cells that are also transfected with a functional HyPB domain will retain their expression of TdTomato as the gene was deposited within the genome of cells and will be stably inherited by daughter cells. The red cells will then create colonies of red cells that are not present in the TdTomato-only control.

HyPB-ERα transfections produced red colonies that expanded over time in the same way as the undirected HyPB transfection (Figure 3.4 B). After 11 days, red colony formation was clear in all HyPB containing transfections, from both the number and intensity of colonies compared to the TdTomato only control. In contrast, in the TdTomato only control, expression is limited to isolated cells that were initially transfected, but the expression of the TdTomato gene cannot be passed on to their daughters, as the sequence was not integrated by a HyPB domain. The presence of red colonies in the N- and C-terminal fusion protein transfections indicates that the HyPB partner within the fusions is functional and inserts cleaved SRTs into the genome.

Together, we show that the HyPB partner of our HyPB-ERα fusion proteins is a functional transposase that can incorporate transfected SRTs into host genomes for stable integration and inheritance by daughter cells. The next step of the project is to validate the function of the ERα partner in our fusion proteins; we therefore continued forward with the first ERα bulk calling card sequencing experiment.

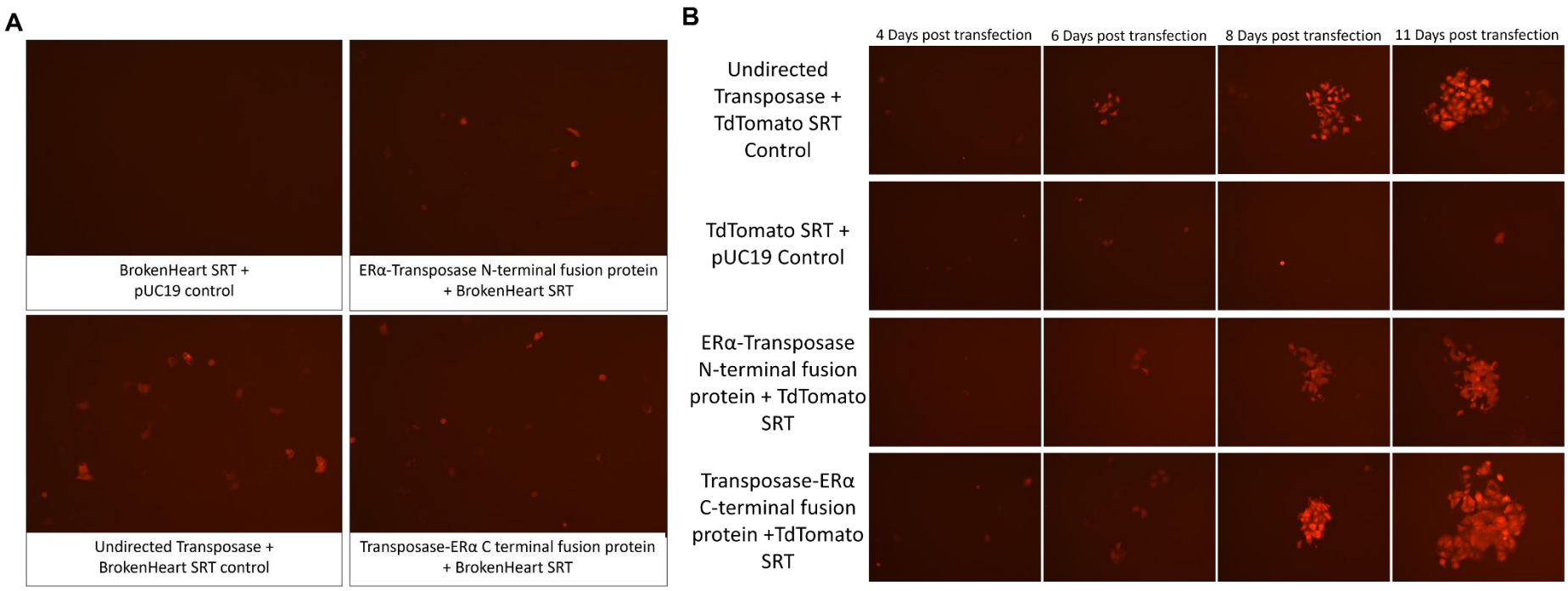


Figure 3.4: Images of cells co-transfected with ERα fusion proteins and TdTomato or BrokenHeart SRTs show successful insertion of PiggyBac Transposons.

Cells co-transfected with the BrokenHeart transposon were imaged 72 hours post-transfection **(A)**, showing that only cells transfected with a functional HyPB can express DsRed, demonstrating the function of ERα fusion proteins. Cells co-transfected with either C-or N-terminal ERα fusion expression plasmids and the TdTomato SRT were imaged every 48 hours for 11 days **(B).** These images show that while the pUC19 control cells only maintain TdTomato expression in the original transfected cell, daughter cells of ERα fusion protein transfections can inherit the TdTomato gene after insertion, like the Undirected HyPB control.

3.2.4. MCF7 cells transfected with C-terminal HyPB-ERα calling card reagents produce sufficient mRNA to generate sequencing libraries

Puromycin resistance SRTs were selected for bulk calling cards to enable the selection of functionally transfected cells and increase the proportion of calling card transcripts in the mRNA of harvested cells. Inclusion of a selection step enriched for cells that were co-transfected with a HyPB plasmid and SRT before harvesting RNA and for the sequencing library pre-preparation via the bulk calling card protocol.

A dose-response curve was first established to determine the optimum puromycin concentration to be 2 µg/mL (Appendix Figure 8.4). Using this data, 2 µg/mL was selected as the optimum concentration as it was the lowest concentration that achieved 100% cell death after 3 days.

Following transfection and selection with 2 µg/mL puromycin, MCF7 cells were harvested for bulk calling card processing, and almost all the C-terminal HyPB-ERα and undirected HyPB transfections produced sufficient mRNA for downstream processing (Figure 3.5 and traces in the Appendix Figure 8.5). One of the C-terminal and all the N-terminal HyPB-ERα transfection replicates did not produce sufficient RNA, which was reflected in considerably smaller cell pellets following passage and a failed TapeStation trace.

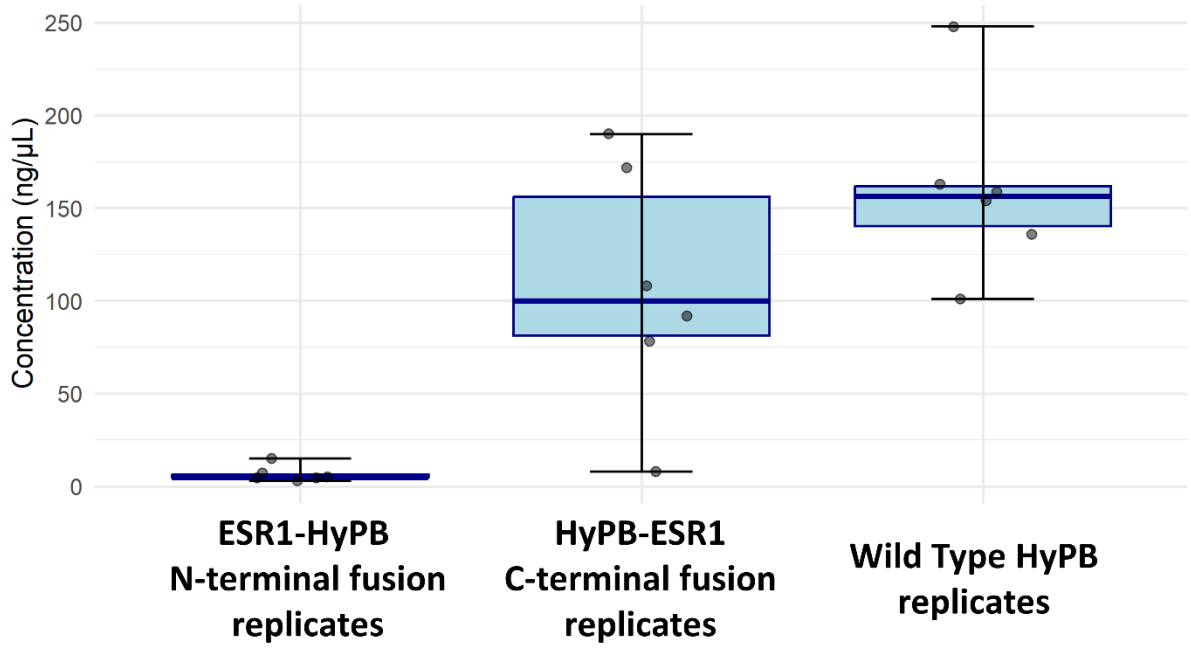


Figure 3.5: Summary of high-sensitivity TapeStation total RNA traces shows good recovery of RNA for C-terminal fusion and undirected HyPB bulk calling cards, whereas N-terminal fusion calling cards fail to produce sufficient quality RNA. Cells transfected with N-terminal fusion protein plasmids showed reduced viability throughout puromycin selection, which is reflected in the low RNA yield shown on the TapeStation traces. The amount of RNA across replicates was not sufficient for at least 50 ng/µL required to carry forward. C-terminal fusion protein and wild-type traces, on the other hand, are sufficient to carry forward, except for the C-terminal fusion replicate 4, which also produced insufficient RNA.

An aliquot of the samples was run on a 1% agarose gel to qualitatively check the amplification of SRT fragments alongside quantitative confirmation of concentration using the Bioanalyzer (Figure 3.6). We visibly observe the range of calling card transcripts amplified by electrophoresis and through bioanalyser data, revealing amplified fragments up to 9000 bp long. These additional rounds of QC confirmed that the calling card processing was proceeding as anticipated.

The concentration in the size region of expected SRT fragments (200 bp to 1000 bp) was within the expected range of 2 – 4 ng/µL. We therefore continued with sequencing and subsequent data analysis of the first ERα bulk calling card dataset.

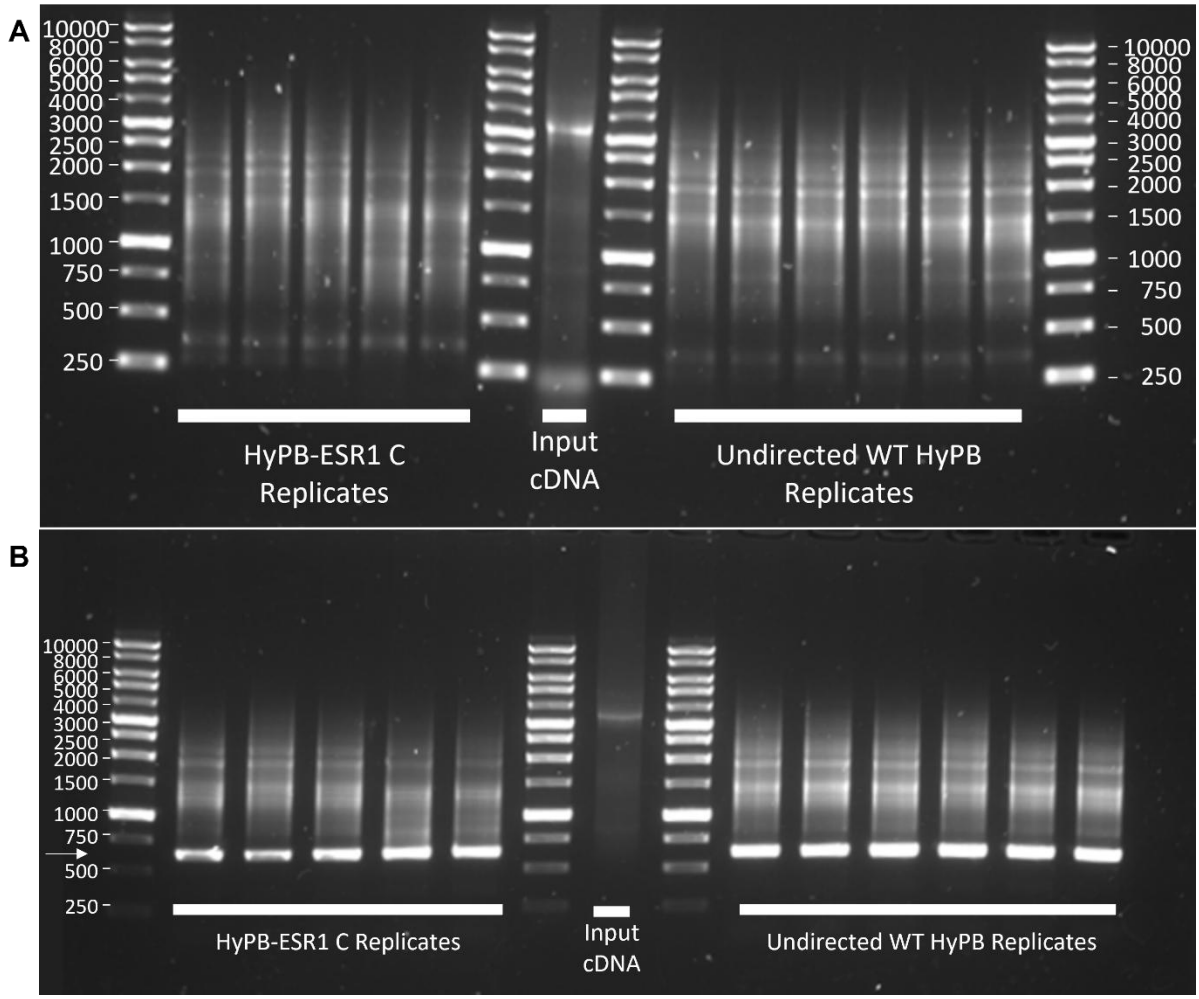


Figure 3.6: PCR amplified SRT libraries run on a 1% agarose gel show smearing of PCR products between the expected range of 500 – 5000 bp, in addition to the β-actin control.

Amplified SRT libraries should produce products between 500 – 5000 bp if the insertions of calling cards into the host cell genome have been successful. Running these amplification products for all C-terminal fusion replicates and undirected HyPB replicates on an agarose gel **(A)** shows a smear of PCR product within this range. Interestingly, the amplification within the undirected transposase appears brighter. As a positive control, amplification of a constitutively active gene such as β-actin is run on the amplified SRT libraries; running these products on the gel **(B)** reveals a product at the expected size of 626 bp in all lanes, as indicated by the white arrow.

3.2.5. HyPB-ERα calling card normalised insertions at called peaks display correlation

Raw fastq files were processed according to the previously published pipeline for calling card data; the workflow for this can be found in an RMarkdown document (See Section 6.1.7). The pipeline generates qBED files (Moudgil et al., 2020), which are BED3+3 files that contain additional information such as the number of times an insertion was detected, the strand it

was detected on and the sample-specific barcode. Finally, sorted qBED files were passed onto the PyCallingCards (Guo et al., 2024) programme for peak calling with MACCs, A calling card-specific peak caller that functions similarly to MACS (Feng et al., 2012; Thomas et al., 2017; Jeon et al., 2020; Guo et al., 2024), a common ChIP-Seq peak caller (Thomas et al., 2017; Nakato and Sakata, 2021).

In this analysis, we examine two sets of calling card peaks derived from the same. qbed input but differ in the p-value cut-offs used in MACCs peak calling to determine the appropriate value for our constructs. Engineering a novel calling card reporter demands the optimisation of the analysis pipeline to ensure genuine HyPB-ERα binding events are properly distinguished from undirected HyPB insertions.

To examine the effects of varying peak calling parameters, a ‘High stringency’ set of peaks was created with PyCallingCards' (Guo et al., 2024) MACCs function, followed by creating a ‘Low Stringency’ set of peaks using looser cut-offs. The difference between these two peak files lies in the ‘pvalue_cutoffbg’ and ‘pvalue_cutoffTTAA’ inputs for low stringency peaks; these parameters were increased by an order of magnitude (0.001 to 0.01 and 0.00005 to 0.0005). The step size was also decreased from 800 to 500 for the high stringency set. The high stringency set contains 506 peaks, whereas the low stringency set contains 935.

The first step in analysis, after calling peaks, was to show whether the insertions across replicates were consistent and reproducible (Moudgil et al., 2020; Roth et al., 2023), we therefore calculated the normalised number of insertions for each replicate for correlation analysis (example Figure 3.7, each replicate combination in the Appendix Figure 8.6). While typically IDR is used in ChIP-Seq analysis to determine reproducibility (Li et al., 2011; Landt et al., 2012), IDR is not used as a standard in calling card analysis due to challenges in generating pseudoreplicates; this is discussed later. Instead, the correlation between normalised insertions at statistically significant calling card peaks is used instead.

The normalised insertion counts were adjusted for sequencing depth but not for potential differences in ERα protein abundance between replicates or cell lines, as this data was not collected. Peaks are recorded as single base pair +1 upstream of the SRT insertion site, and the number of individual binding events detected from that site is used to determine the peak's significance. The number of insertion events at each peak across replicates was normalised and compared with each other replicate to determine the reproducibility of HyPB-ERα insertions at those peaks. This approach captures differences in insertion frequency but does not account for variation in ERα protein levels between replicates or cell lines. More generally,

insertion counts should be interpreted qualitatively, as calling cards indicate that ERα bound at a locus but do not provide a linear or quantitative measure of binding affinity or duration. As a result, some of the observed variability could reflect differences in protein abundance rather than only TF occupancy.

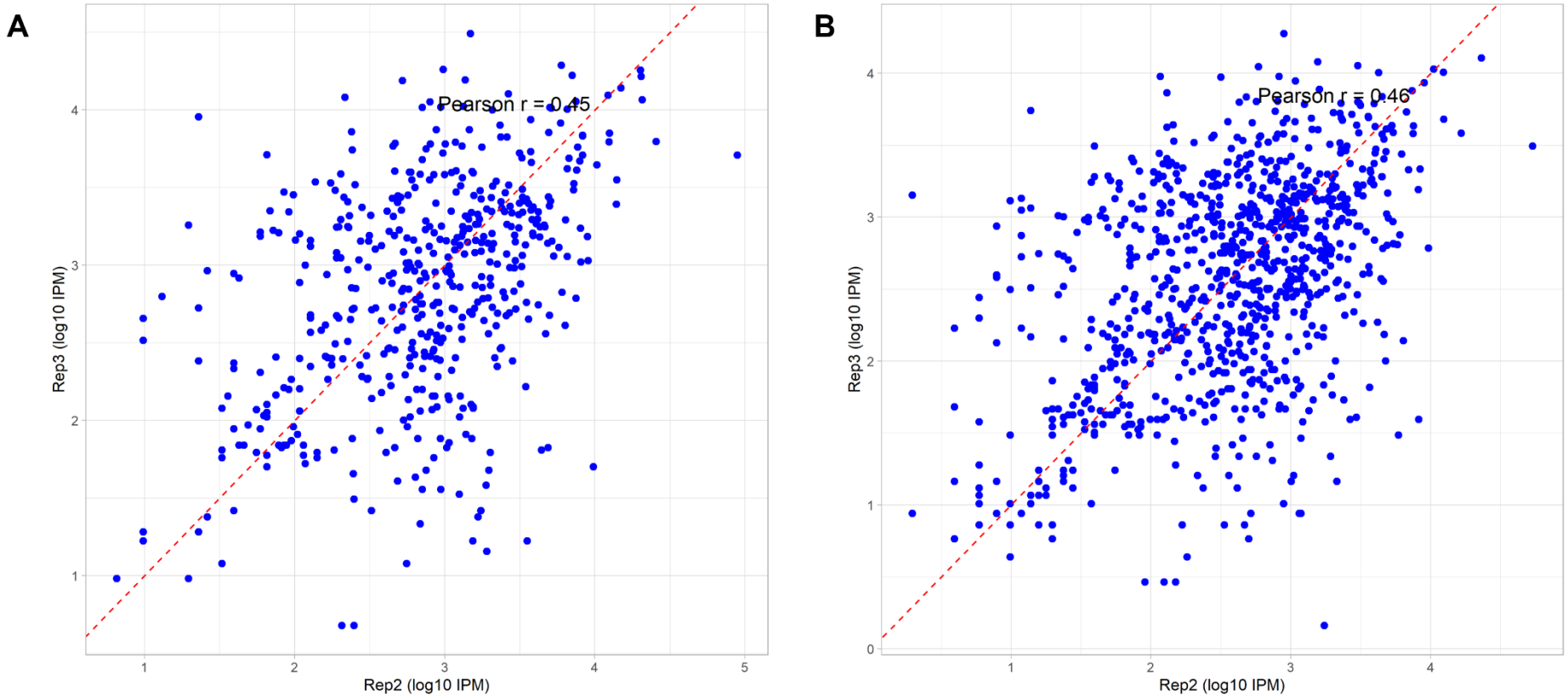


Figure 3.7: HyPB-ERα calling card replicates 2 and 3 show moderate correlation at MACCs peaks.
For both the High stringency, **n** = 506, **(A)** and Low stringency, **n** = 935, **(B)** peaks called with MACCs, there is a moderate correlation of 0.45 or 0.46 between replicates 2 and 3. However, correlations vary from 0.30-0.46. Insertions for each replicate were subset based on overlap with either set of peaks or were normalised to Insertions Per Million (IPM) before plotting. The calculated significance for high stringency peaks correlation ($p = 3.97 \times 10^{-11}$) and low stringency peaks ($p = 7.1 \times 10^{-24}$) for these replicate pairs and all other combinations is $< 1 \times 10^{-10}$.

The correlation across all replicates was moderate or fair (Gust and D'journo, 2015; Mukaka, 2012), with a correlation coefficient of 0.3 to 0.46 across all replicates. For simplicity of analysis, only the two highest-scoring replicates were shown here for each set of calling card peaks; all comparisons and Pearson coefficients can be found in the Appendix Figure 8.6.

The high stringency peak generates a Pearson correlation coefficient of 0.45 in replicates 2 and 3 compared to the 0.46 coefficient shown in low stringency peaks within these replicates. These coefficients show that while there is moderate consistency between replicates, there are high levels of variability at some loci, indicated by the variability between $\log_{10}$ IPM at some peaks across all replicates. Interestingly, the Pearson correlation coefficient is consistently higher in all low-stringency peaks across every replicate pair.

The correlation coefficient may be lower than expected for this data because of the loss of a calling card replicate during harvest, as each replicate is critical in determining genuine HyPB-ERα insertions (Moudgil et al., 2020). Previous publications detail that the number of replicates in a calling card experiment is crucial, as these replicates increase the statistical power of peak calling, allowing genuine peaks to be more easily distinguished from background insertions (Moudgil et al., 2020). Genuine transcription factor binding sites are identified by calculating the likelihood of the number of insertions at a given locus compared to random chance. Therefore, by increasing the number of replicates, it is possible to increase the statistical power of peak calling to differentiate transcription factor-directed insertions from undirected background noise.

The importance of every replicate is demonstrated when attempting to call ERα peaks with only four replicates; no statistically significant peaks are called with any combination of four replicates.

To illustrate the importance of each replicate, the overlap between statistically significant calling card peaks and individual replicates was determined and plotted to show that over 95% of all the peaks are present in each of the replicates (**Figure 3.8**).

These results show that the combined insertions in every replicate alone are not sufficient to call peaks, but when combined, replicates have the statistical power to differentiate HyPB-ERα calling card insertions from undirected insertions.

Together, these data reveal a moderate correlation (Gust and D'journo, 2015; Mukaka, 2012) between the normalised insertions between calling card replicates, and that the vast majority of calling card peaks originate from insertions across all five replicates.

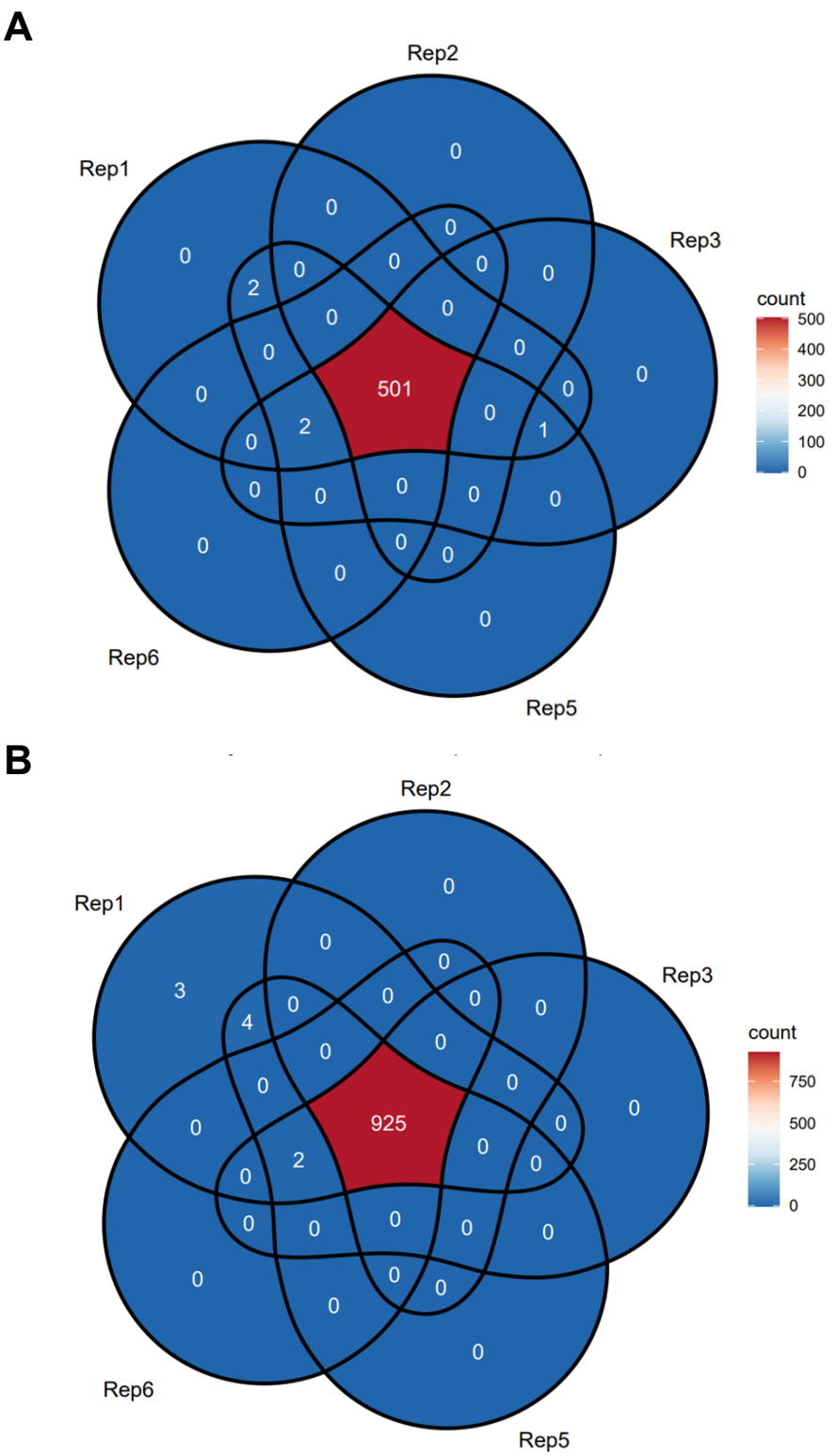


Figure 3.8: Over 95% of calling card peaks called with MACCS overlap with insertions in every ESR1-HyPB replicate.
In both the high stringently called peaks **(A)** and low stringently called peaks **(B)** up to 1% of peaks are found in individual replicates or a group of replicates. However, around 99% of peaks overlap with all the replicates, demonstrating concordance with the significant insertion sites across replicates.

3.2.6. HyPB-ERα calling cards show a significant correlation with ERα ChIP-Seq peak distribution across chromosomes

To characterise the genome-wide distribution of ERα calling card insertions, we first assessed the distribution of calling card peaks across the genome. We visualised the genomic distribution of calling card insertions as karyotype plots, showing calling card insertions across all chromosomes (Appendix Figure 8.7). Regions flagged in the ENCODE blacklist were excluded due to anomalous or repetitive signal in ChIP-Seq data (Amemiya, Kundaje and Boyle, 2019).

To test for non-random distribution, a chi-square (X^2) test comparing the observed number of peaks per chromosome to the expected number based on chromosome size (Figure 3.9) (McHugh, 2013). Both high and low stringency calling card datasets showed significant deviation ($p < 0.001$) from uniform distribution at some chromosomes.

Across both high and low stringency calling cards, the difference in percentage deviation between observed and expected shows chromosomes 17 and 20 were most overrepresented, with observed peaks exceeding 150 to 250% of expected (Figure 3.10). Other chromosomes, like 6, have significant calling card peak over-representation, but not to the same extent ($p < 0.01$) as chromosomes 17 and 20. Some chromosomes had fewer peaks than expected in both datasets, like chromosomes X and 13. Several chromosomes only show significant deviation from the expected number of peaks in the low stringency data; these are chromosomes 4, 8, 12, and 22.

To compare with multiple examples of endogenous ERα binding, we acquired two ChIP-Seq datasets, one from the Gene Expression Omnibus (GEO)(GSE109820) (Holding et al., 2019) and the other from ENCODE (ENCFF063JMY) (ENCODE Project Consortium, 2012). These will be referred to as the GEO and ENCODE datasets from here.

Karyotype plots (Appendix Figure 8.8) and X^2 tests (Figure 3.11) were replicated for ChIP-Seq peaks distribution, which showed significant deviation from uniform distribution; in fact, on the majority of chromosomes, ERα ChIP-Seq peaks deviated from random distribution. For the GEO dataset, chromosomes 3, 6, 7, 8, 9, 11,15 and 22 were the only chromosomes that did not significantly deviate from expected ($p < 0.05$), whereas for the ENCODE data, only 11 and 21 did not significantly deviate. To normalise the difference between the observed and expected distribution of calling card or ChIP-Seq peaks at each chromosome, the percentage difference was calculated for each dataset and was plotted in R.

Interestingly, while the ERα ChIP-Seq peaks show more significant deviation from uniform distribution, the percentage difference between observed and expected is generally lower cross the genome compared to ERα calling cards (Figure 3.10).

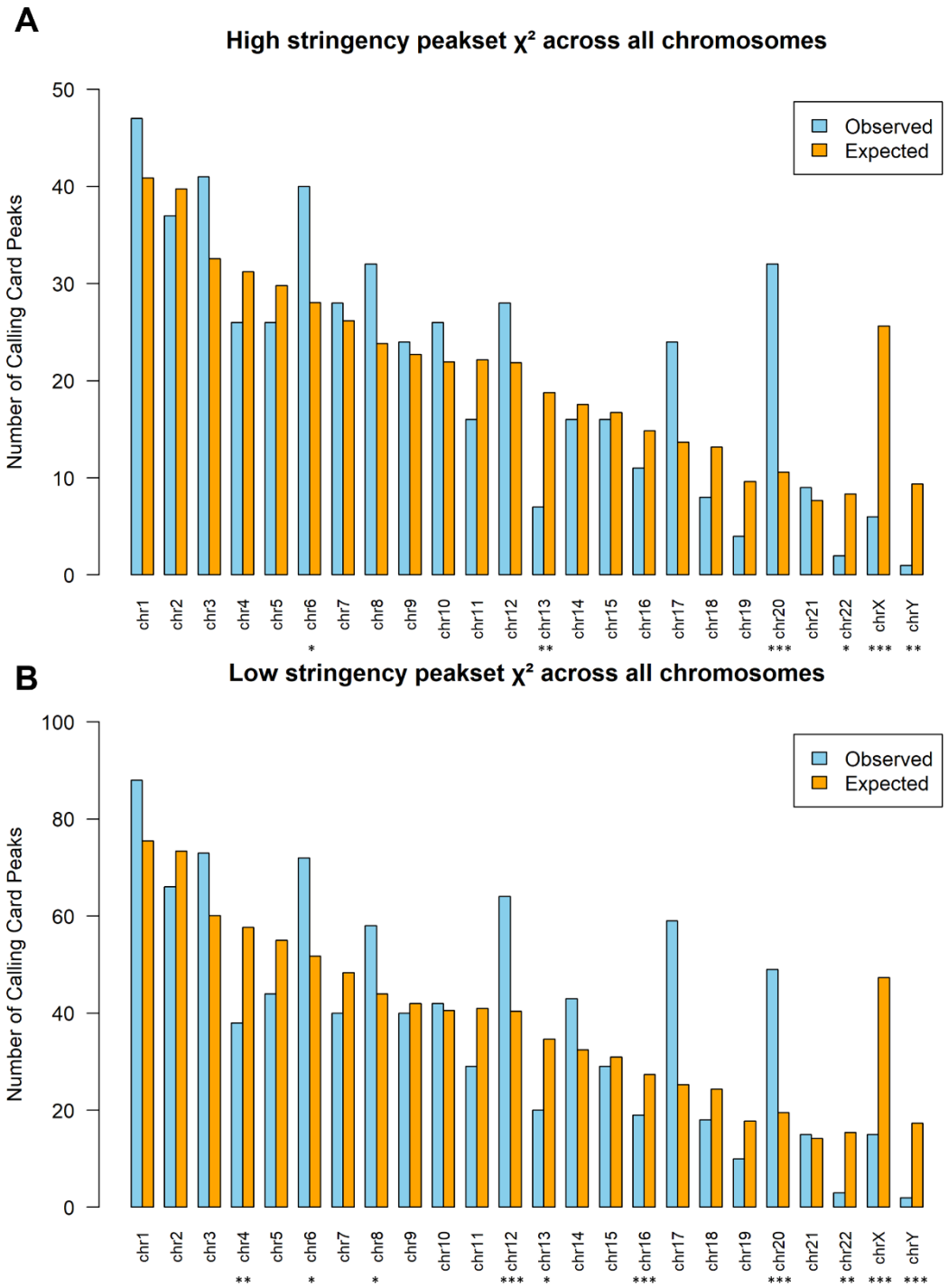


Figure 3.9: Calling card peaks are not equally distributed across chromosomes in the genome.

X^2 tests were carried out on both high **(A)** and low **(B)** stringency calling card peaks, and the observed and expected values for each chromosome were plotted using R. For high stringency peaks, the difference between expected and observed is much higher than in the low stringency peaks. The most similar chromosomes are 7, 9, 15 and 21. In the low stringency peaks, four chromosomes have a similar number of expected and observed peaks: chromosomes 9, 10, 15 and 21. The remainder of the chromosomes have a much larger difference between expected and observed. P-values indicate chromosomes at which observed peaks are significantly different to the expected ($p < 0.05$ = *, < 0.01 = **, < 0.001 = ***).

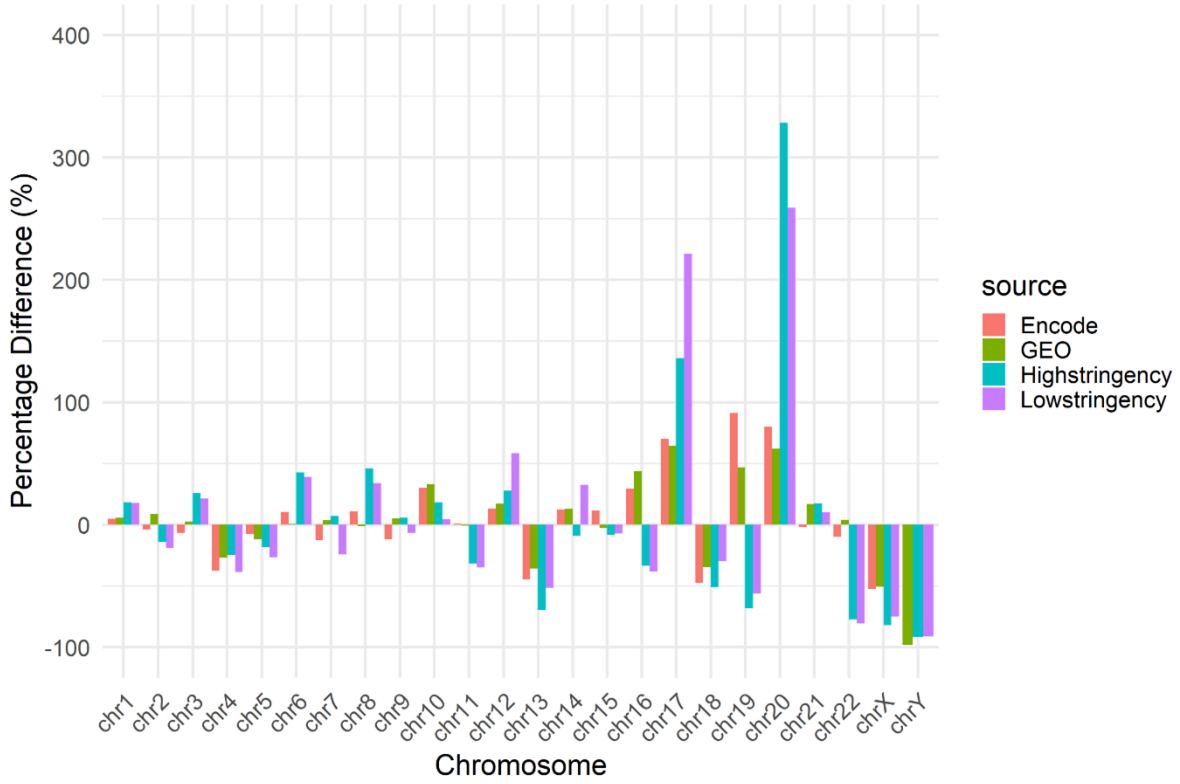


Figure 3.10: Calling card peaks large overrepresentation in chromosomes 17 and 20 compared to ChIP-Seq, but otherwise generally reflect endogenous ERα binding across the genome

The percentage difference between observed and expected peaks on a given chromosome was calculated for each dataset and was plotted compared to each other dataset. For most chromosomes, the percentage difference across datasets is relatively similar. The largest changes between methods occur at the end of the genome, with calling cards showing a large bias towards binding chromosomes 17 and 20.

`

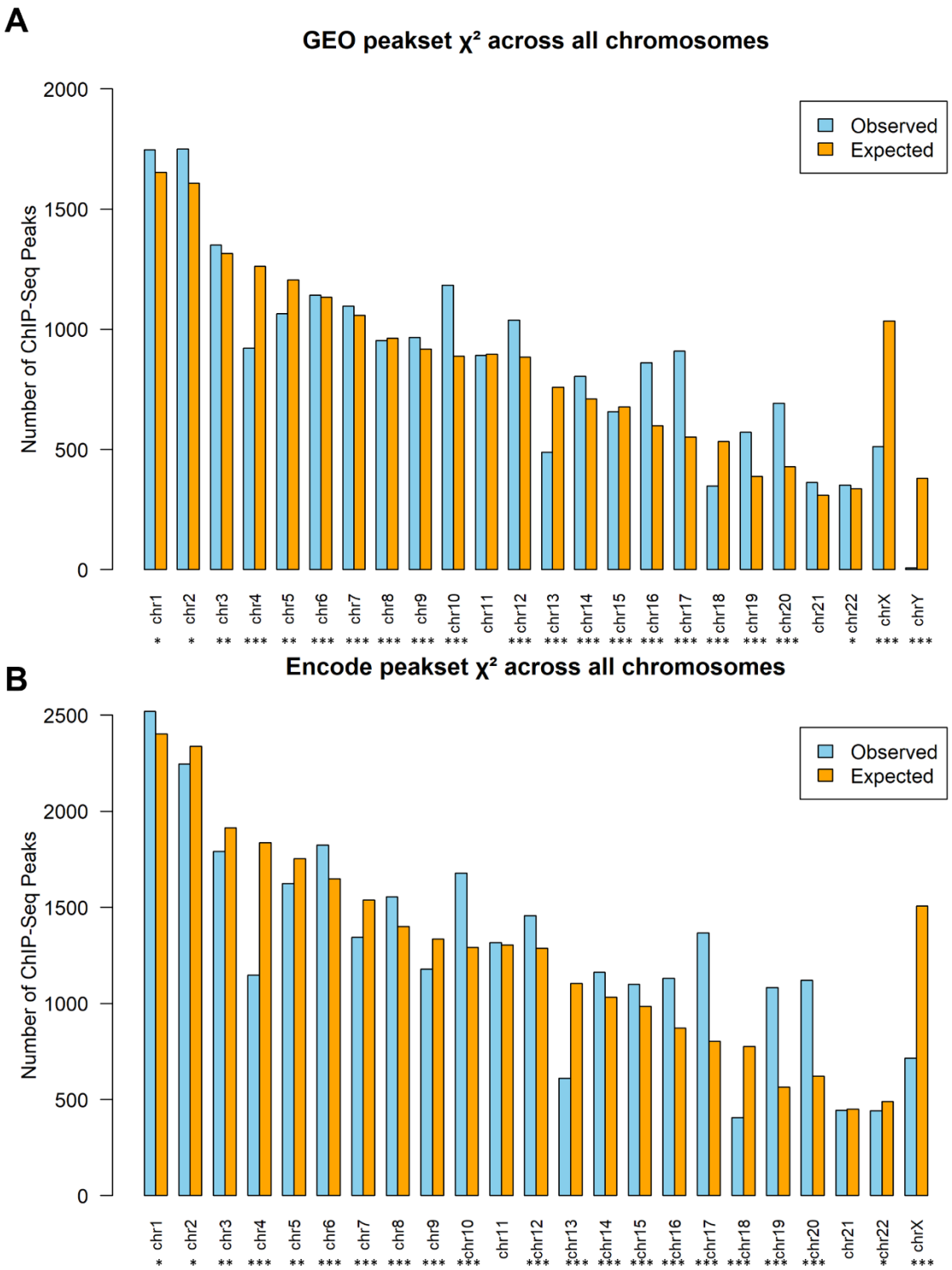


Figure 3.11: ChIP-Seq peaks are not equally distributed across chromosomes in the genome.

X² tests were carried out on both GEO **(A)** and Encode **(B)** ChIP-Seq peaks, and the observed and expected values for each chromosome were plotted using R. For the GEO peaks, nine chromosomes have a similar number of expected and observed peaks: chromosome 3, 6, 7, 8, 9, 11, 21 and 22. For the Encode peaks, 3 chromosomes have a similar number of expected and observed peaks: chromosome 11, 21 and 22. The remainder of the chromosomes have a much larger difference. P-values indicate chromosomes at which observed peaks are significantly different to the expected ($p < 0.05$ = *, < 0.01 = **, < 0.001 = ***).

Pearson correlation analysis between percentage deviations revealed strong within-method consistency ($r > 0.95$, $p < 0.01$) and moderate (Gust and D'journo, 2015; Mukaka, 2012) between-method agreement ($r = 0.56\text{-}0.65$, $p < 0.01$), but crucially, all correlations were statistically significant (Figure 3.12).

Together, these results suggest that HyPB-ERα calling cards and ERα ChIP-Seq peaks share general chromosome targeting patterns, but each method may capture distinct aspects of ERα binding. Given the smaller number of calling card peaks, some over- or under-representation may reflect sampling variation. Furthermore, calling cards record the presence of binding events rather than quantitative occupancy, differences in the magnitude of chromosome-level enrichment are expected, and the key comparison is the shared pattern of targeting rather than absolute signal. Nonetheless, the significant enrichment of chromosomes 17 and 20 warrants further investigation.

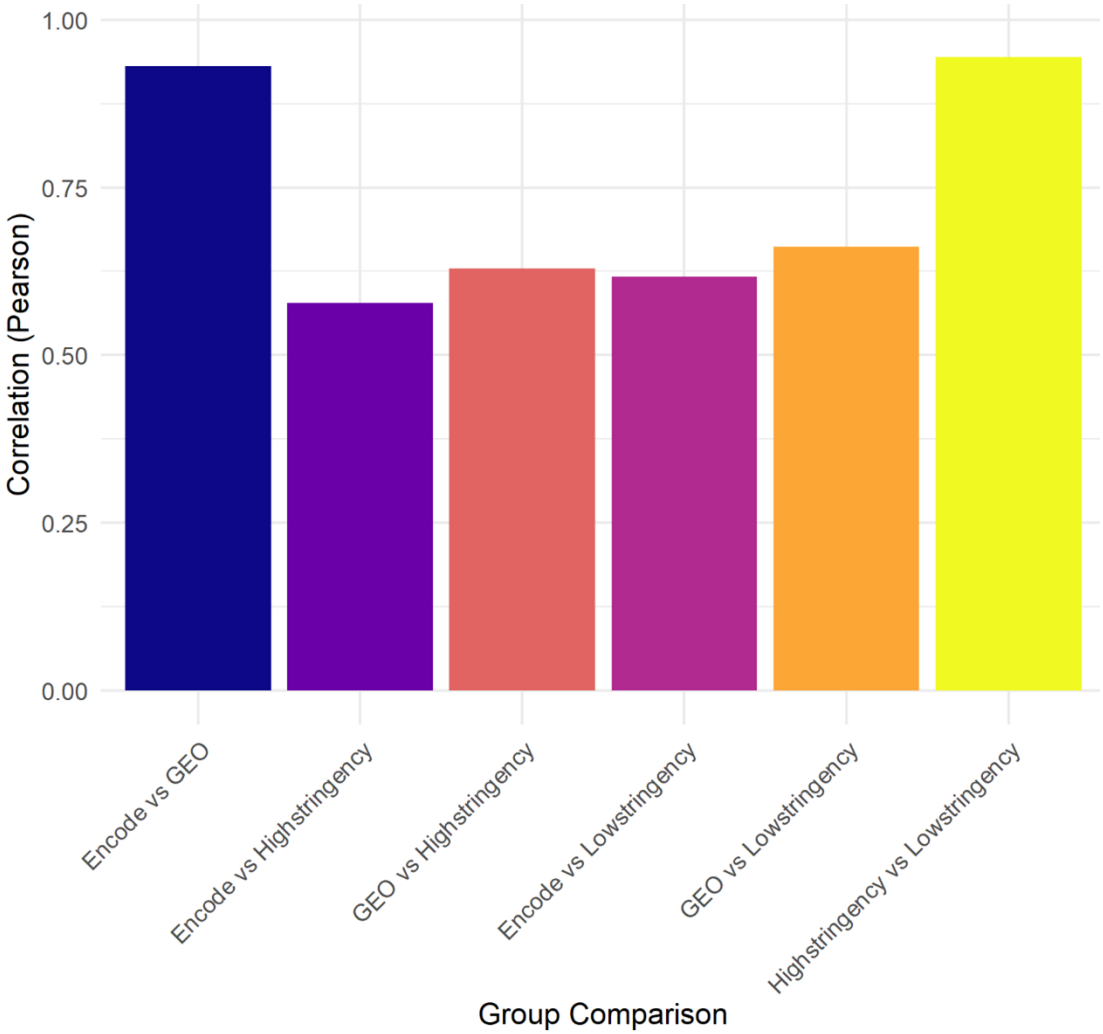


Figure 3.12: The correlation of percentage change between expected and observed peaks located across the genome is very strong within comparisons within ChIP-Seq or calling card datasets, while the correlation between methods is moderate.

Using the calculated percentage difference in each group, the correlation of these changes was calculated to compare the similarity of the distribution of calling card and ChIP-Seq peaks. When comparing peaks from the same method, correlation is very strong (>0.95). In contrast, the correlation of ChIP-Seq and calling card peaks varies between 0.56 – 0.65, a moderate correlation. These data suggest that while the global binding patterns of HyPB-ER are not the same as the endogenous protein, they are quite similar. Statistical significance of correlations was evaluated using rcorr (Hmisc package, R), which computes two-tailed p-values for each pairwise correlation (each $p < 0.001$).

3.2.7. HyPB-ERα Calling Cards are distributed at expected genomic features and distances from gene transcription start sites

To further explore whether HyPB-ERα can reflect genuine ERα binding patterns, we analysed the distribution of ERα calling card peaks across annotated genomic features and relative to transcription start sites (TSS). The distribution of calling card peaks was then compared to the

distribution of ERα GEO and ENCODE ChIP-Seq peaks to identify differences from endogenous behaviour.

Feature distribution analysis was performed using the ChIP Seeker R package, and results were visualised by sample type (Figure 3.13 A) and feature category (Figure 3.13 B). Overall, the pattern of calling card and ChIP-Seq peaks was broadly similar, with consistent representation in intronic, UTR and downstream regions. There were, however, notable differences observed in promoter and distal intergenic regions.

Promoter-associated peaks were less frequent in calling card peaks, with ~7.5 - 10% of calling card peaks located within promoter regions compared to the ~15-20% of ChIP-Seq peaks. In contrast, distal intergenic regions were more represented in the calling card data, with ~25% of calling card peaks within these regions compared to the ~20% of ChIP-Seq peaks. These differences suggest that calling card insertions may preferentially capture distal regulatory element binding or enhancer-associated binding events. ERα is associated with binding to enhancer regions, indicating that our HyPB-ERα may recapitulate ERα binding.

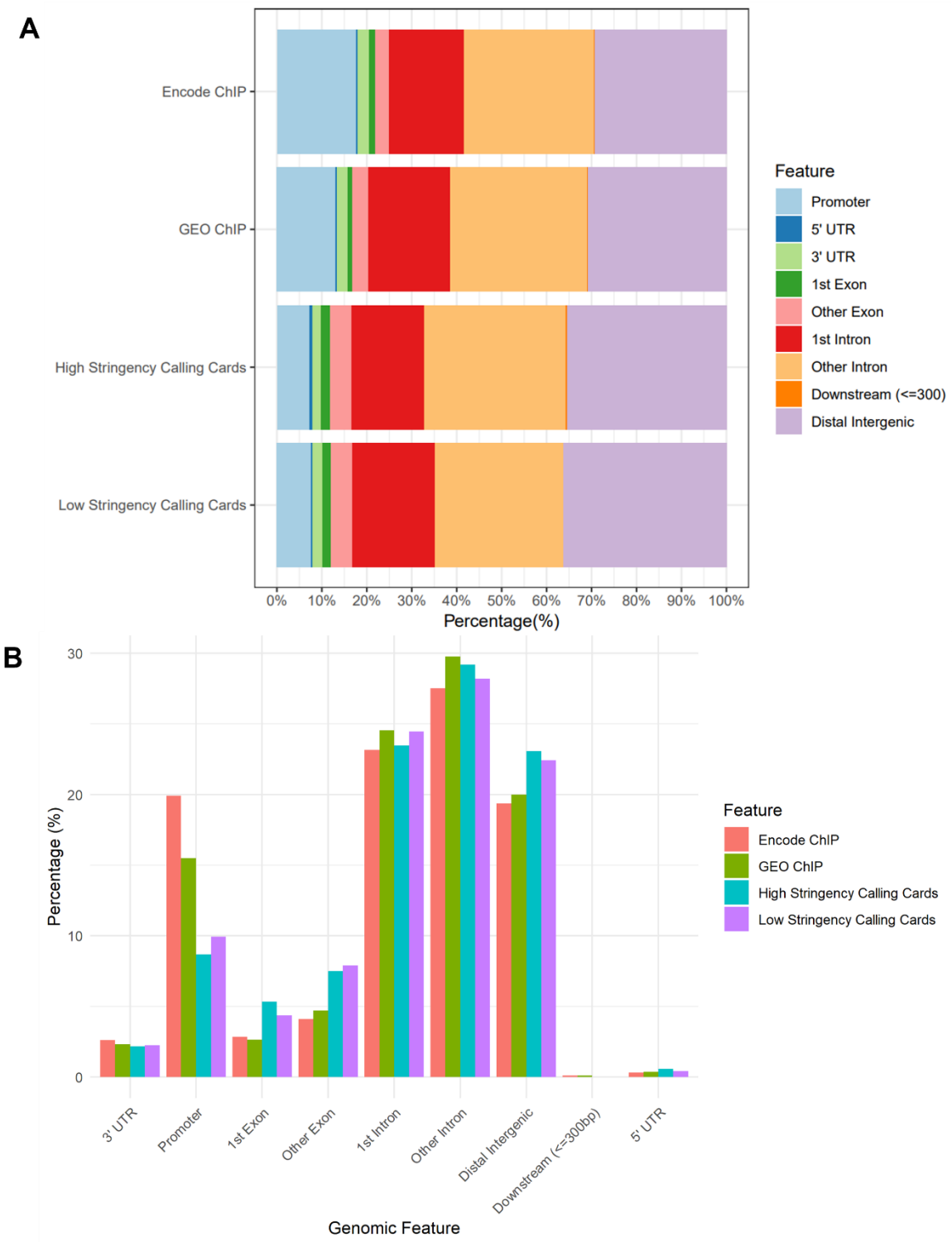


Figure 3.13: 83% of calling Card peaks are located within introns or other distal regions, similar to endogenous ERα binding patterns, with up to 10% reduced binding within promoter regions.

Feature distribution analysis of calling cards in high and low stringent peak sets expressed as a stacked bar chart by sample type **(A)** and a bar chart by genomic feature **(B)** reveals that most peaks are located within a gene intron or at distal intergenic regions, reflecting the binding patterns of ERα acting as a long-range enhancer of gene regulation. However, results from both ChIP-Seq data sets show that peaks from these methods do have a higher proportion of binding sites within the promoter region and a lower proportion of binding at distal intergenic sites. Differences between groups are non-significant.

We next examined the positional distribution of peaks relative to the gene TSS (Figure 3.14) and the calling card and ChIP-Seq datasets showed a bimodal distribution around the TSS. The proportion of peaks within 1 kb of the TSS was higher in ChIP-Seq datasets than calling card datasets, consistent with the reduced promoter expression. Similarly, calling card data showed an increase in peaks of more than 5 kb from TSS compared to ChIP-Seq, concordant with higher representation in distal intergenic regions. These findings are also consistent with ERα's role in participating in distal enhancer regulation (Carroll et al., 2006; Fullwood et al., 2009; Hah et al., 2013).

Together, these results indicate that while ERα calling card peaks broadly recapitulate the genomic feature distribution of ChIP-Seq peaks, they show increased representation at distal intergenic regions and reduced promoter binding. These differences may reflect distinct aspects of HyPB-ERα binding, but overall, the general pattern of binding reflects endogenous ERα binding. Some of this divergence may also arise from the fact that calling cards capture individual insertion events rather than quantitative enrichment, meaning that peak representation does not directly correspond to binding intensity.

Having established that ERα calling card peaks are distributed across expected genomic features and show positional patterns consistent with long-range transcriptional regulation, we next examined direct overlap with ChIP-Seq peaks.

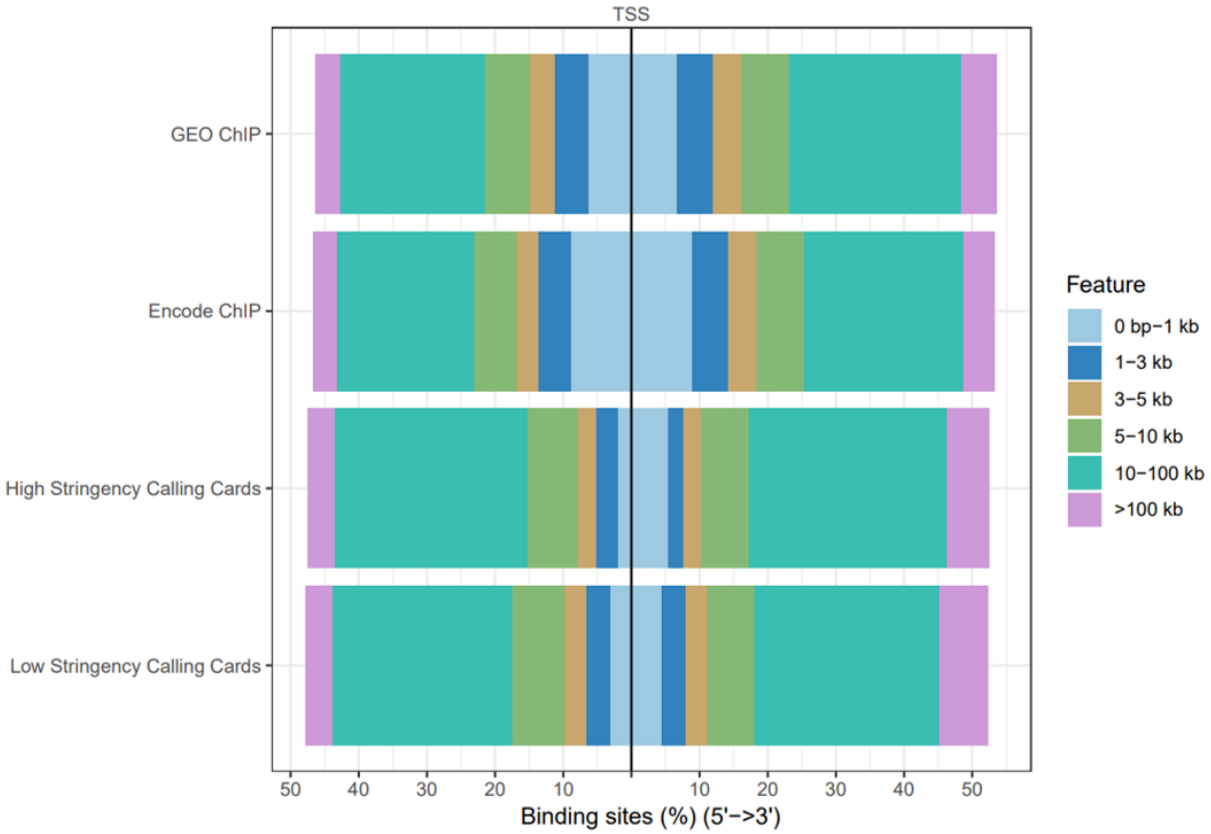


Figure 3.14: Around 80% of calling Card peaks are located more than 5kb from transcription start sites, with less binding within 1 kb than peaks observed in ChIP-Seq datasets.
Analysis of feature distribution relative to the Transcription Start Site (TSS) of the nearest genes reveals that the distribution of peaks up and downstream of the TSS is roughly the same in both sets of calling card peaks, with a small increase in downstream peaks. Around 80% of calling card peaks are found further than 5 kb from the TSS, reflecting ERα's role as a long-range transcription factor. However, this is around 10% higher compared to ChIP-Seq, which shows increased binding within 1kb of the TSS.

3.2.8. HyPB-ERα bulk calling cards peaks display statistically significant overlap with ChIP-Seq peaks

The widely accepted gold standard for determining the binding patterns for transcription factors is ChIP-Seq (Xu et al., 2021; Jaini et al., 2014; Narlikar and Jothi, 2012; Jothi et al., 2008; Johnson et al., 2007). Therefore, to determine if the ERα partner of the fusion protein retains spatial concordance with ChIP-Seq peaks, the percentage overlap of calling card peaks with existing ERα ChIP-Seq datasets was investigated. This is an important step in analysis as because ChIP-Seq provides a quantitative enrichment profile, whereas calling cards capture the genomic locations where ER binding has occurred over time. This difference

offers a complementary and temporally integrative readout that is not intended to measure instantaneous binding strength but instead records binding events with high precision.

Previous calling card reporters have been shown to insert SRTs into the genome within 1000 bp of the initial TF binding site (Wang et al., 2012a; Moudgil et al., 2020; Wang et al., 2011; Cammack et al., 2020). Bedtools window was used to expand the peaks by 1000 bp and intersect was used to count the number of overlaps between calling card and ChIP-Seq peaks. The number of overlapping CC peaks was divided by the total number of CC peaks and multiplied by 100 to calculate the percentage of calling card peaks that overlap with ChIP-Seq.

ChIP-Seq peaks were re-called using MACS2 at a range of p-value cut-offs and using a q-value of 0.05 to determine which was the most appropriate method of peak calling (Figure 3.15 A). Comparison of calling card overlap with ChIP-Seq peaks called at multiple cut-offs is required for new calling card reporters to confirm that peaks are not being called too harshly or loosely. Using stricter cut-offs for peak calling reduced the number of ChIP-Seq peaks for Calling cards to overlap, as decreasing the p-value cut-off also reduced the amount of overlap between peaks. The q-value for all peaks gave the most overlap between the calling card and ChIP-Seq datasets. Given that the original publication utilised q-value instead of p-value cut-offs and MACS2 pipelines recommend using q-value, we continued forward using the FDR corrected peaks (Benjamini and Hochberg, 1995; Feng et al., 2012).

To facilitate clearer data visualisation and reduce complexity in comparative analysis, Encode replicate 2 was used for downstream comparison. This replicate was chosen because it contained the most uniquely mapped reads, which is indicative of a higher signal-to-noise ratio (Landt et al., 2012; Jung et al., 2014; Marinov et al., 2014). While IDR-controlled peak sets are typically preferred for robust reproducibility, the high read depth of replicate 23 made it a practical choice for the initial validation of calling card profiles (Figure 3.15 B).

Hypergeometric testing revealed that TTAA sites within calling card peaks were significantly enriched for overlap with both ERα ChIP-Seq datasets (Appendix Table 8.4). The observed ~1.3 fold-enrichment exceeded expectation compared to random sampling, with highly significant p-values. These results confirm that calling card reporters faithfully capture ER binding events and further highlight their intersection with BRD4 occupancy and ER-mediated chromatin interactions.

Up to 70% of our high-stringency HyPB-ERα calling card peaks were found to overlap with GEO ChIP-Seq peaks, which is comparable to previous SP1-HyPB calling card reporters

(Wang et al., 2012a). To contextualise the overlap percentages, 80% of the SP1 calling card reporter insertion clusters have been found to overlap with their corresponding ChIP-Seq peaks; for called peaks, the percentage was reduced to 72% (Moudgil et al., 2020; Wang et al., 2012a). Low stringency calling card peaks were found to overlap less with GEO ChIP-Seq peaks compared to high stringency peaks, with only 63.3% of low stringency peaks overlapping with GEO ChIP-Seq peaks.

The overlap with ENCODE ChIP-Seq peaks was lower across both high and low stringency calling card peaks, with 64% and 58.4% respectively; although these are lower than previously reported calling cards, they still constitute the majority of calling card peaks.

Comparing the percentage overlap between both calling card datasets, the high stringency dataset has 7% more peaks overlapping with the GEO dataset and 5.4% more overlapping peaks with the ENCODE data compared to low stringency peaks.

Overall, these results demonstrate that utilising stricter calling card cut-offs selects for peaks that are in the immediate vicinity of ERα ChIP-Seq peaks, whereas looser cut-offs result in more peaks that are further than 1000 bp from ChIP-Seq peaks.

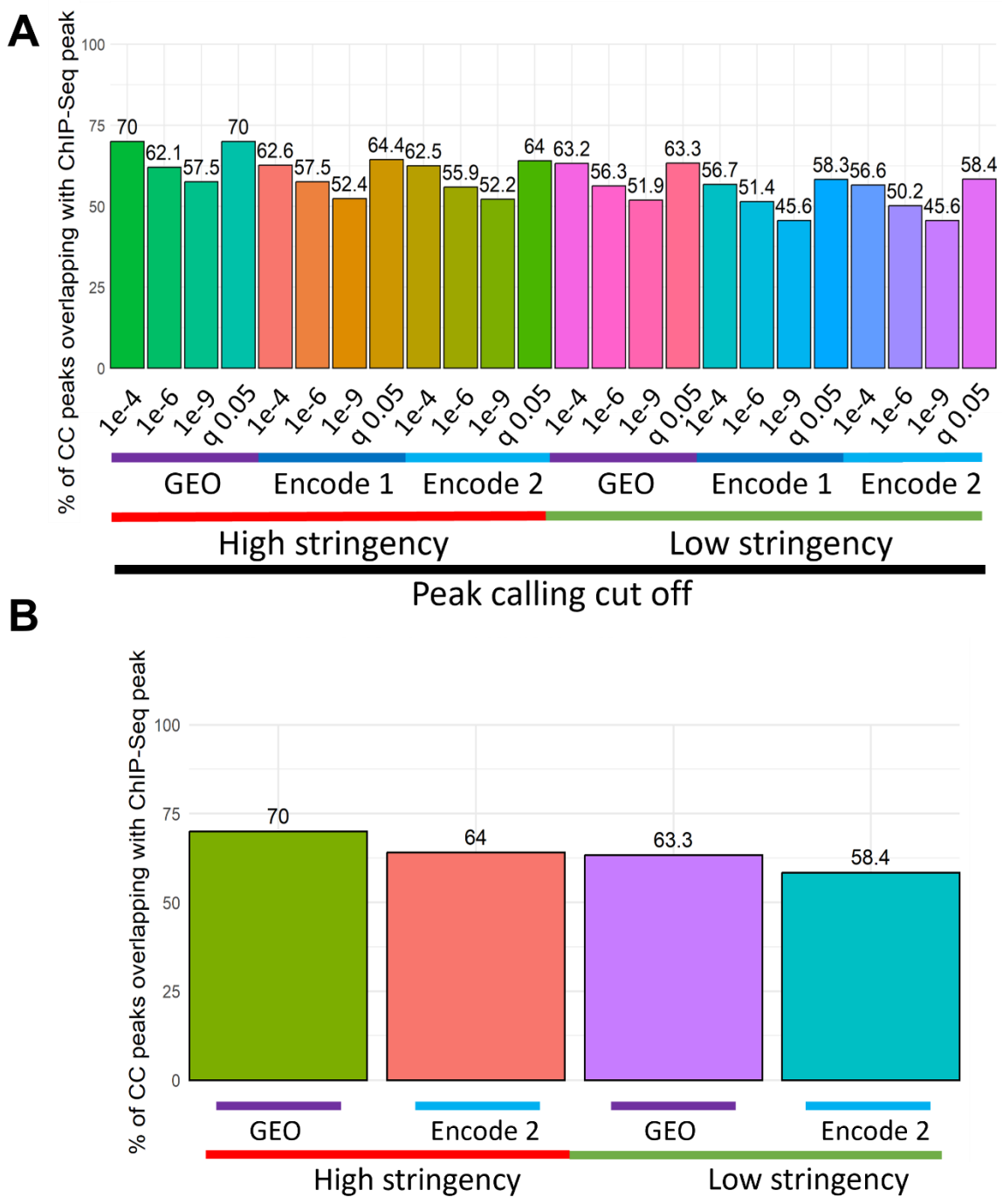


Figure 3.15: Up to 70% of HyPB-ERα directed calling card peaks overlap with ER ChIP-Seq peaks from GSE109820, ENCFF365BIT and ENCFF063JMY datasets.
Bedtools window was used to determine whether peaks called with the PyCallingCards MACCS function overlap with existing ChIP-Seq datasets **(A)** at varying p-value cut-offs. The majority of calling card peaks overlapped with either GSE109820, ENCFF365BIT (Encode 1) or ENCFF063JMY (Encode 2) when called with the more stringent q-value of 0.05 **(B)**. This plot is a subset of only the GEO and Encode 2 ChIP-Seq datasets called using q=0.05. Encode 2 was carried forward as the Encode example, as it was the larger of the two datasets. The overlap of these peaks is lower than the 83% value reported for the model SP1 Calling Card dataset, indicated by the dashed horizontal line. Hypergeometric testing confirms significant overlap between ERα calling card peaks and independent datasets ($p < 1 \times 10^{-300}$).

3.2.9. ChIP-Seq signal is enriched at the centre of HyPB-ESR1 calling card peaks

To further assess concordance between the ERα calling card and ChIP-Seq-defined binding sites, we plot ChIP-Seq signal enrichment centred on calling card peaks. For each dataset, GEO and Encode, we generated average Log_2 fold change signal profiles across ±10 kbp from calling card peak centres (Figure 3.17 and Figure 3.18).

Both GEO and ENCODE ChIP-Seq datasets show clear enrichment at the centres of calling-card peaks, supporting the overlap between insertion sites and previously reported ERα binding sites (Figure 3.16). Both high- and low-stringency peak sets exhibited significantly elevated ChIP-Seq signal compared to shuffled controls (Wilcoxon $p < 0.001$), confirming that calling card reporters capture bona fide ERα binding events. Signal enrichment was observed consistently across GEO and ENCODE datasets

For the GEO ChIP-Seq signal, the average signal peaks at ~3.5 Log_2 fold change at the centre of calling card peaks, with elevated signal extending several kilobase pairs up- and downstream (Figure 3.17 A and C). The heatmap confirms the enrichment extends across many of the peaks in the dataset, particularly in the high stringency peaks (Figure 3.17 B and D).

The Encode ChIP-Seq data shows a higher average Log_2 fold change of ~7 at the centre of calling card peaks, which likely represents differences in sequencing depth of normalisation rather than biological variation (Figure 3.18 A and C). The heatmaps further illustrate that many of the top calling card peaks show strong signal enrichment at the centre of peaks (Figure 3.18 B and D).

Some of the lower-ranked peaks show more variable levels of enrichment across both datasets, which suggests that while many of the calling card peaks align with the ChIP-Seq signal, some represent distal or indirect binding that is not captured by ChIP-Seq.

This is consistent with previously reported models that do not show complete overlap with ChIP-Seq but are found to be consistent with TF occupancy and targeting criteria.

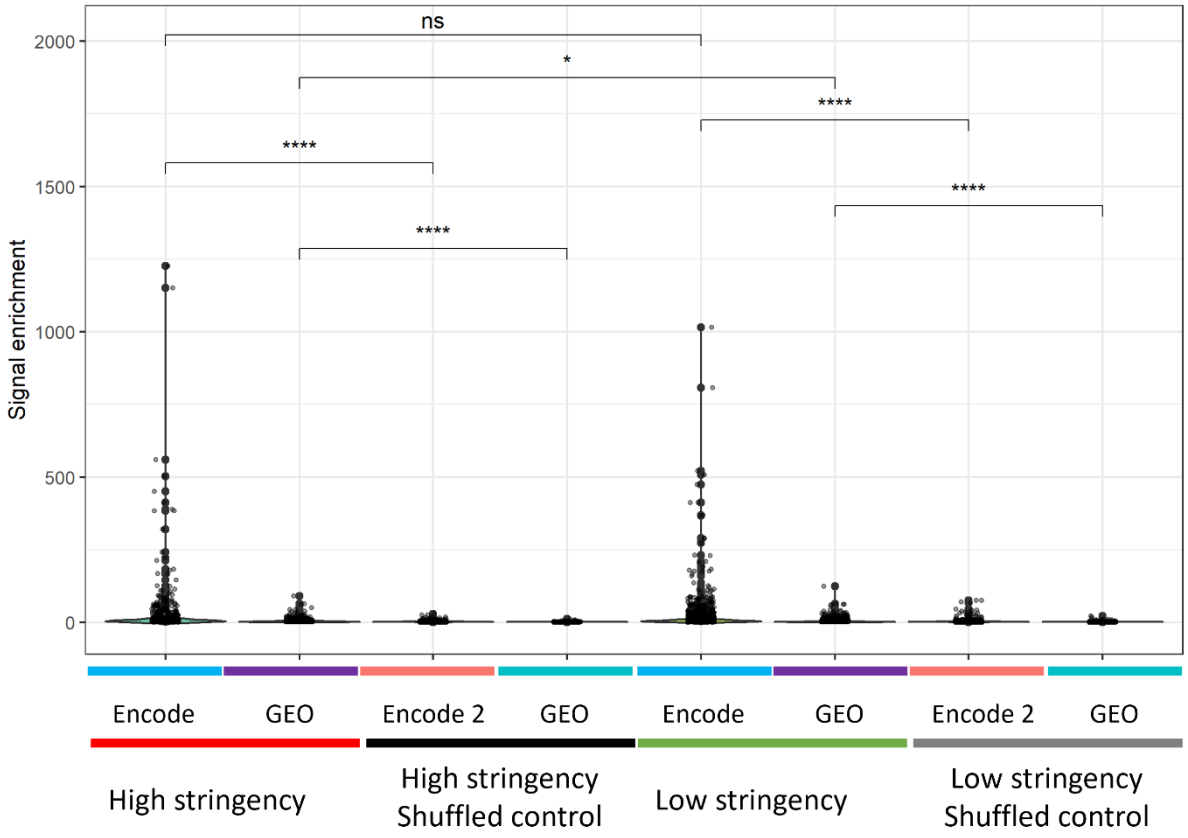


Figure 3.16: ERα ChIP-Seq signal is enriched at calling card peaks. Violin plots show mean ChIP-Seq signal intensity within 1 kb of high and low stringency calling card peaks compared to matched shuffled controls.

Calling card peak bed files were used as inputs for the BedTools shuffle function to generate comparable random controls to compare enrichment with calling card peaks. High and Low-stringency calling card peaks were found to be significantly enriched compared to shuffled controls. P-values from Wilcoxon tests are indicated for each comparison ($p < 0.05$ = *, < 0.01 = **, < 0.001 = ***, < 0.0001 = ****

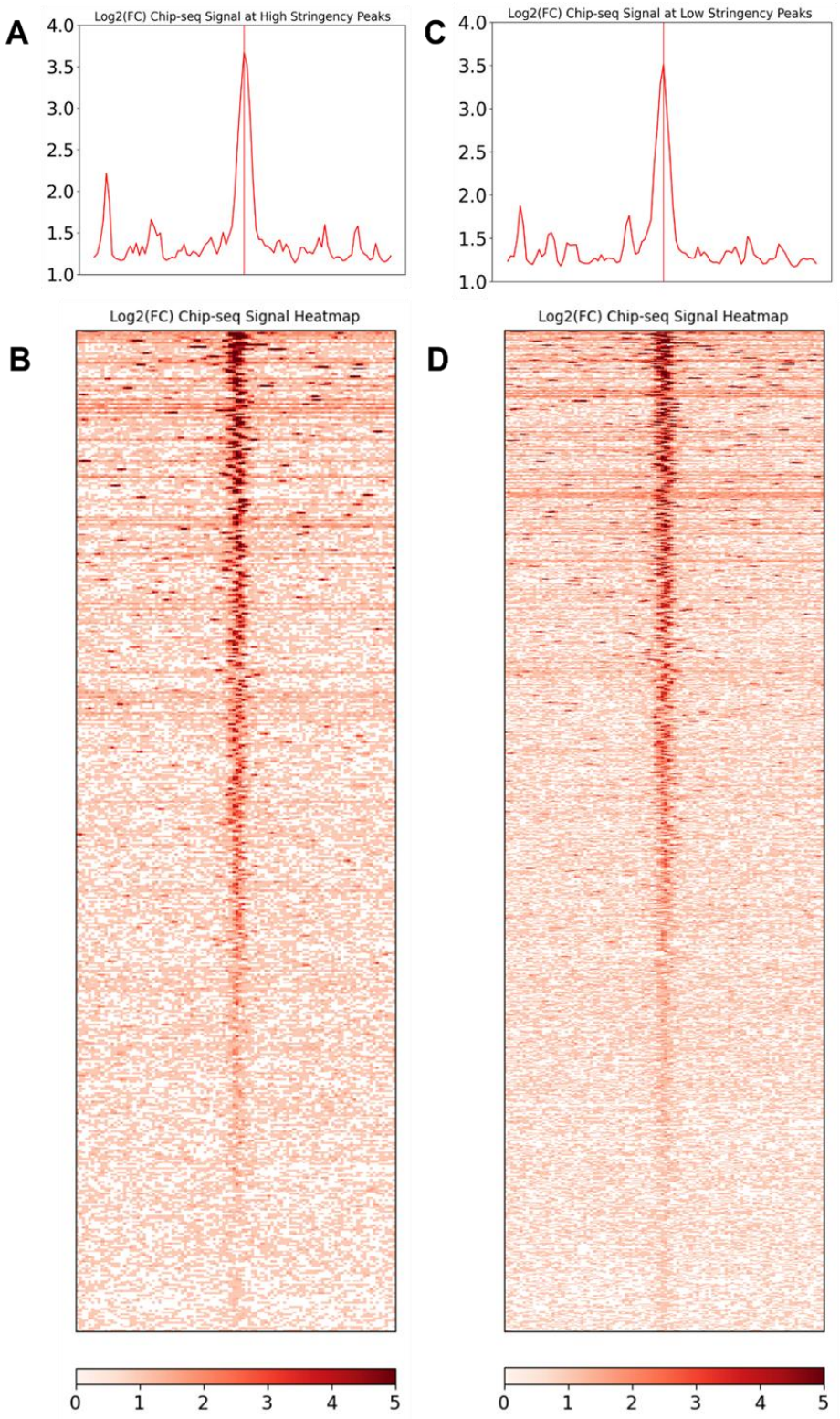


Figure 3.17: GSE109820 ChIP-Seq signal is enriched at the centre of Calling Card peaks.

Analysis of ChIP-Seq signal from GSE109820 shows an average Log_2 fold change of 3.5 when centred on Calling Card peaks for high stringency **(A-B)** and low stringency **(C-D)**, shown as a signal intensity plot **(A), (C)**. The enrichment heatmaps **(B), (D)**, show that at the top peaks, a Log_2 fold change of up to 5 at the centre of peaks and other regions within 10,000 bp of peak centres.

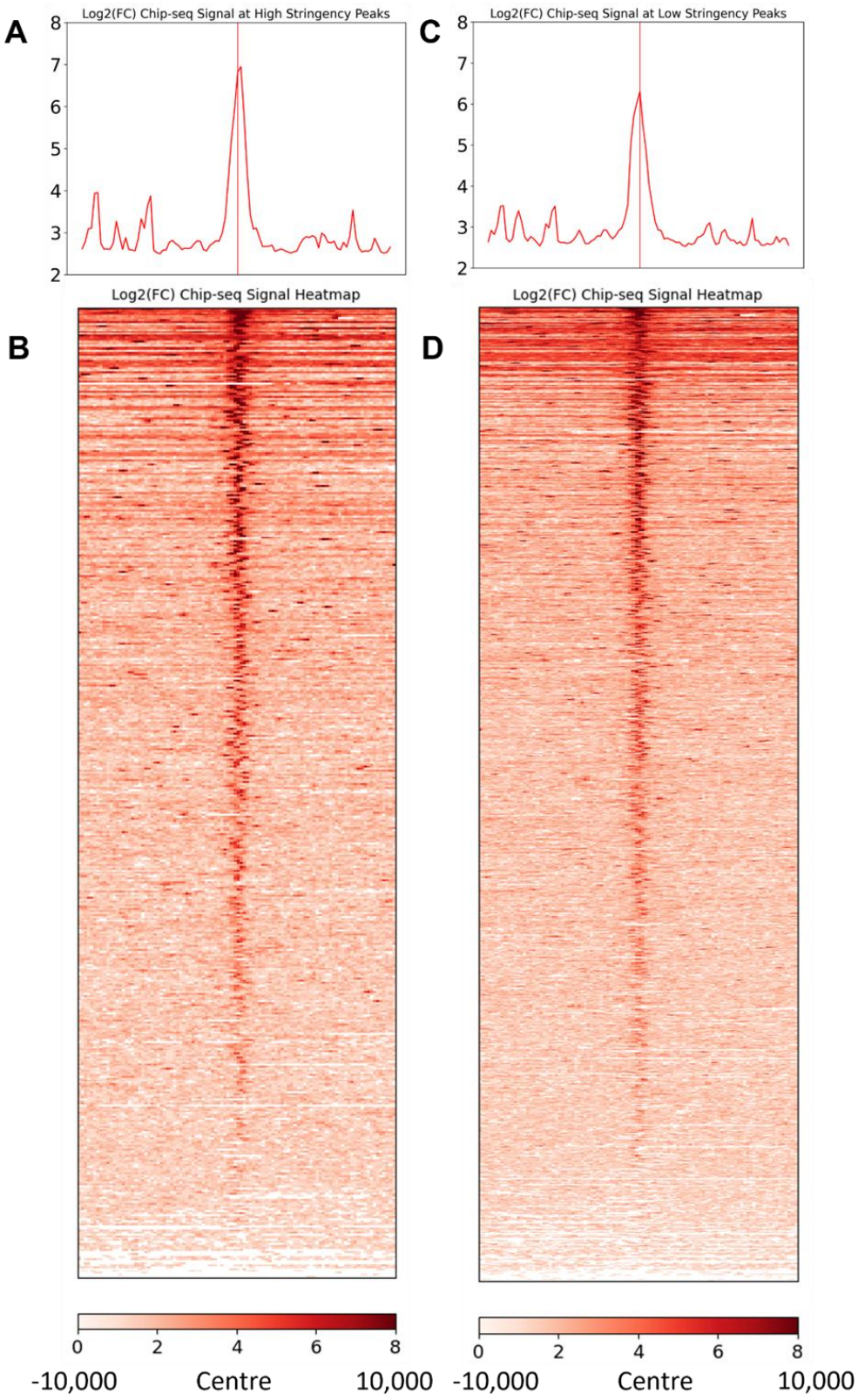


Figure 3.18: ENCFF063JMY ChIP-Seq signal is enriched at the centre of Calling Card peaks.

Analysis of ChIP-Seq signal from ENCFF063JMY shows an average Log_2 fold change of 7 when centred on Calling Card peaks for high stringency **(A-B)** and low stringency **(C-D)**, shown as a signal intensity plot **(A), (C)**. The enrichment heatmaps **(B), (D)**, show that at the top peaks, a Log_2 fold change of up to 8 at the centre of peaks and other regions within 10,000 bp of peak centres.

3.2.10. Sensitivity and Specificity of calling card peaks using an adjusted p-value threshold is not a good indicator of ChIP-Seq overlap

To determine the performance of peak calling using the PyCallingCards (Guo et al., 2024) MACCs program, we applied receiver-operator characteristic (ROC) analysis to high and low stringency calling card peaks. The goal was to assess whether the adjusted p-value assigned to each calling card peak is a reliable threshold for identifying true TF binding sites, like previous calling card reporters (Moudgil et al., 2020). In this analysis, the closer the area under the curve (AUC) is to 1, the more effective the threshold is at correctly classifying peaks based on the ground truth. The binary classifier for this analysis was defined as whether a calling card peak overlapped with a ChIP-Seq peak from the GEO and ENCODE datasets as the ground truth reference.

For the high stringency calling card peaks, the AUC was 0.58 for the GEO dataset and 0.52 for the ENCODE data (Figure 3.19 A-B). This AUC would define adjusted p-value as a poor/non-discriminating model for identifying genuine ERα binding sites, and that this threshold performs only marginally better than chance in distinguishing overlap with ChIP-Seq (Nahm, 2022; Çorbacıoğlu and Aksel, 2023). Notably, the ROC curve for the ENCODE overlap dips below the diagonal line at several points, indicating inconsistency in classification performance across the thresholds tested.

Interestingly, the same analysis using the low stringency calling card peaks yields slightly higher AUCs: with 0.6 for GEO overlap and 0.58 for ENCODE overlap (Figure 3.20 A-B). While this is still classified as a poor model, it appears that relaxing peak calling parameters can improve sensitivity to genuine ERα binding events (Nahm, 2022; Çorbacıoğlu and Aksel, 2023). This could include biologically relevant peaks with higher p-values that were excluded under the more stringent threshold; however, it could also include low p-value peaks that do not overlap, resulting in better discrimination of overlap by adjusted p-value.

Importantly, these results do not contradict the observed ChIP-Seq signal enrichment and overlap with calling card peaks. Rather, they highlight that the adjusted p-value for each peak may not be the optimal discriminator for overlap with ChIP-Seq-defined binding sites.

Previous studies have found that the adjusted p-value of calling card peaks shows much stronger performance in ROC analysis for other TFs, suggesting the discrepancy here could be driven by the ERα fusion protein or other confounding variables that may influence detectability.

Further investigation is needed to understand why the MACCs adjusted p-value performs poorly on HyPB-ERα calling cards. Potential factors could include limited replicate coverage, differences in ERα binding dynamics or the influence of regulatory elements that are not captured by ERα ChIP-Seq. We have already demonstrated the importance of replicates in calling card analysis by failing to generate peaks without each of the five replicates used in this experiment. The fact that our ERα calling cards have used fewer replicates than previously published reporters could contribute to the disparity in ROC performance. However, the issue could also be driven by the difference in ERα binding dynamics compared to other TFs. No other NR has been adapted for calling cards reporting; our model is the first application of calling cards to NRs. Therefore, there could be a mechanistic difference in the binding dynamics of ERα that could confound peak calling in ways that do not affect ChIP-Seq.

While the above factors are plausible influences that require further investigation, the factor with the most obvious driver would be interference by the co-binding of BRD4 (Nagarajan et al., 2015, 2014; Yang, He and Zhou, 2008; Zheng et al., 2022). ERα coordinates the acetylation of histones to recruit BRD4, and the two proteins then work in consort to regulate the transcription of their target genes (Nagarajan et al., 2015, 2014; Yang, He and Zhou, 2008; Zheng et al., 2022). This is a challenge in calling card analysis, given that the undirected HyPB protein has affinity for BRD4 and can be used to map BRD4 super enhancers (Moudgil et al., 2020). The co-localisation of undirected and ERα-directed insertions, therefore, has the potential to confound peak calling as the insertion profile between undirected and ERα-directed calling cards may be far more similar than previous calling card reporters. The lack of resolution between undirected calling card insertions and ERα insertions could therefore increase the number of false negatives and false positives.

Taken together, despite the low AUCs of calling card ROC analysis, the substantial overlap in earlier analysis supports the biological relevance of the majority of calling card peaks and justifies continued exploration of the dataset. The next step was to explore the extent of BRD4 co-binding with high and low stringency calling card peaks.

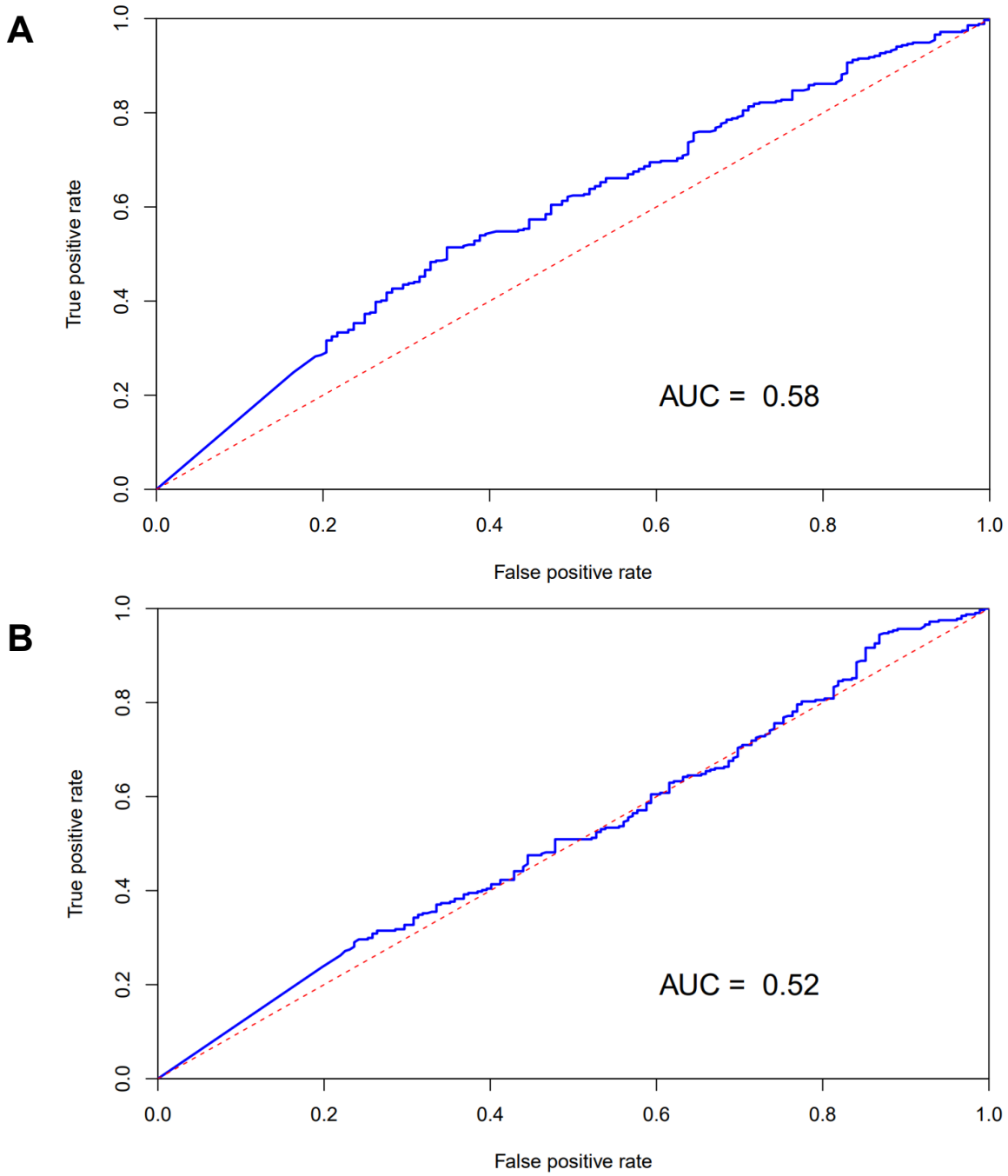


Figure 3.19: Area Under Curve (AUC) of 0.52 and 0.58 for the high stringency peak set overlap between GSE109820 and ENCFF063JMY indicates that successful peak calling using MACCs adjusted p-value is slightly higher than random chance.
The ROC plot for overlap of the high stringency peaks with GSE109820 **(A)** and ENCFF063JMY **(B)** ChIP-Seq datasets using increasing adjusted p-value cut-offs to threshold positive and negative peak calls shows that the True Positive Rate (TPR) and False Positive Rate (FPR) are very close to random when using the adjusted p-value as the threshold, as indicated by the red dotted diagonal line. Overlap with GSE109820 indicates a 58% chance of correctly classifying an unknown peak as a true or false ER binding site. Whereas when using the ENCFF063JMY ChIP-Seq dataset as the ground truth, the chance of correct identification is almost 50%, with the ROC curve dipping underneath the random chance line.

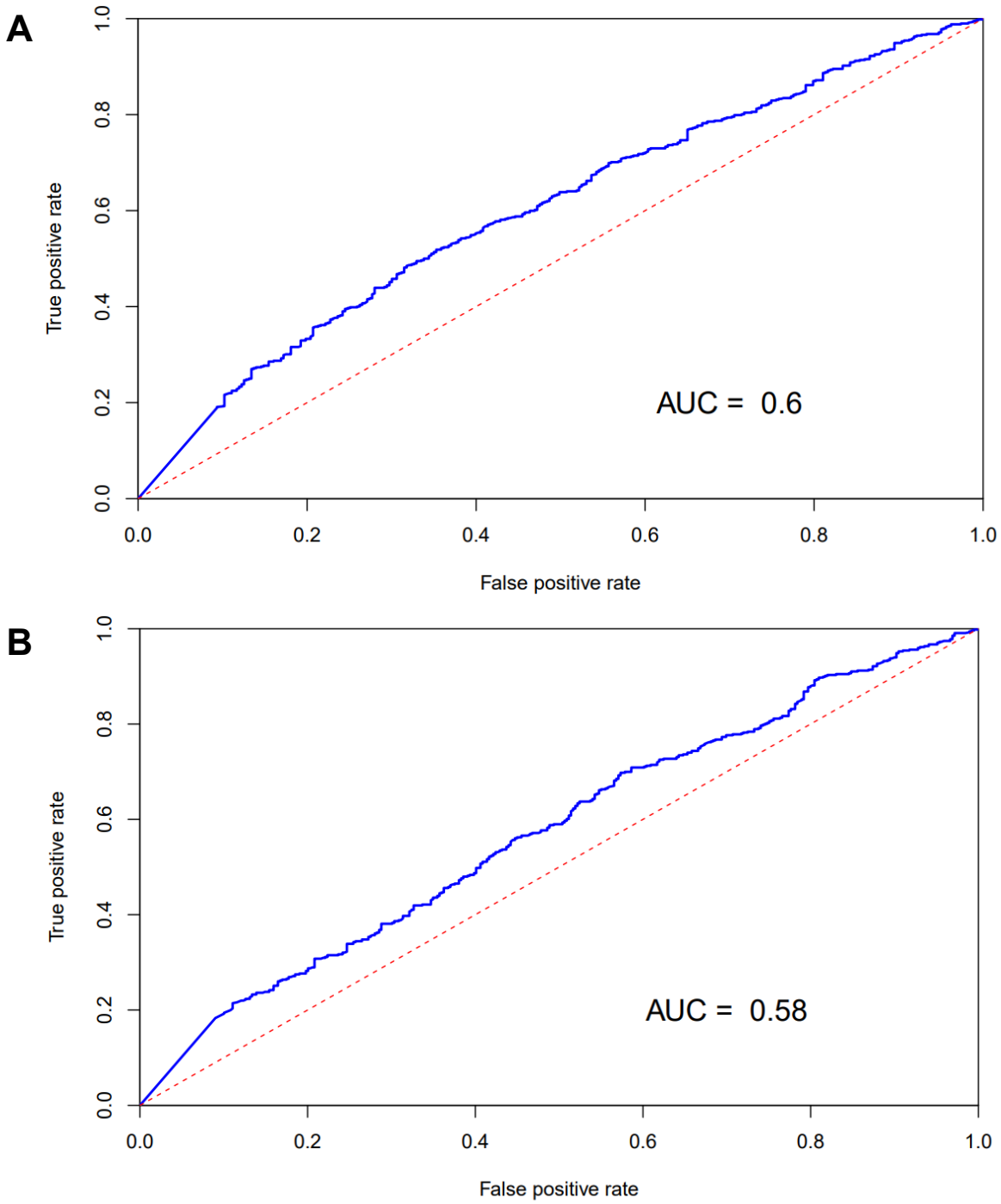


Figure 3.20: AUC of 0.6 and 0.58 for the low stringency peak set overlap between GSE109820 and ENCFF063JMY indicates that successful peak calling is slightly higher than random chance.

The ROC plot for overlap of the high stringency peaks with GSE109820 **(A)** and ENCFF063JMY **(B)** ChIP-Seq datasets using increasing adjusted p-value cut-offs to threshold positive and negative peak calls shows that the TPR and FPR using MACCs adjusted p-value as the threshold are very close to random, as indicated by the red dotted diagonal line. Overlap with GSE109820 indicates a 60% chance of correctly classifying an unknown peak as a true or false ER binding site. When using the ENCFF063JMY ChIP-Seq dataset as the ground truth, the chance of correct identification is almost 58%, improved over the high stringency peaks called, but still close to random chance.

3.2.11. BRD4 peaks overlap highly with ERα Calling Card and ChIP-Seq peaks

It's possible that the close relationship between BRD4 and ERα (Nagarajan et al., 2015, 2014; Yang, He and Zhou, 2008; Zheng et al., 2022) is confounding MACCs peak calling; to explore this, the extent of overlap between BRD4 ChIP-Seq peaks and calling card or ERα ChIP-Seq peaks was determined (Figure 3.21).

We analysed two BRD4 binding datasets from Liu et al. (2022) (Liu et al., 2022a) and the other from Zheng et al. (2022) (Zheng et al., 2022), both were obtained from GEO. Across both datasets, ERα calling card peaks show consistently higher overlap with BRD4 ChIP-Seq peaks than ERα ChIP-Seq peaks. Both high and low stringency peak sets exhibited significantly elevated ChIP-Seq signal compared to shuffled controls (Wilcoxon $p < 1 \times 10^{-300}$, Appendix Table 8.5) confirming that calling card insertions are significantly enriched for BRD4 binding sites.

This trend was more pronounced in the Liu dataset, where the number of overlapping calling card peaks ranged 7-12% higher than ERα ChIP-Seq peaks. The difference between the calling card and ERα ChIP-Seq percentage overlap with the Zheng peaks narrowed, but still, 5-8% more calling card peaks showed overlap with BRD4 peaks. Notably, the overall overlap across both ERα calling cards and ChIP-Seq was lower in the Zheng dataset.

Like ERα ChIP-Seq, enrichment analysis shows that ERα calling card peaks are significantly enriched for BRD4 ChIP-Seq signal compared to random controls (Figure 3.22).

These findings suggest that ERα calling card peaks may be enriched for ERα**:**BRD4 co-binding sites either due to genuine biological interactions or technical bias introduced by HyPB affinity. The piggyBac transposase used in calling cards has been reported to preferentially integrate at TTAA motifs and to show enrichment in accessible chromatin and active regulatory regions (Wang et al., 2012a; Moudgil et al., 2020). BRD4 is frequently associated with active enhancers and super-enhancers, this chromatin accessibility bias may contribute to the elevated overlap observed between calling card peaks and BRD4 ChIP-Seq peaks. It is also possible that BRD4-associated regions are overrepresented in calling card peaks due to the peak calling parameters; however, because the pool of peaks is small due to the number of replicates, this overrepresentation may be the result of random sampling chance.

Taken together, these results support the idea that although HyPB-ERα does capture genuine binding, there could be an influence of BRD4 co-binding with ERα on peak calling, resulting in enriched co-binding sites in calling card peaks. Further refinement of peak calling strategies and additional controls may help to distinguish between genuine ERα binding and BRD4-associated background.

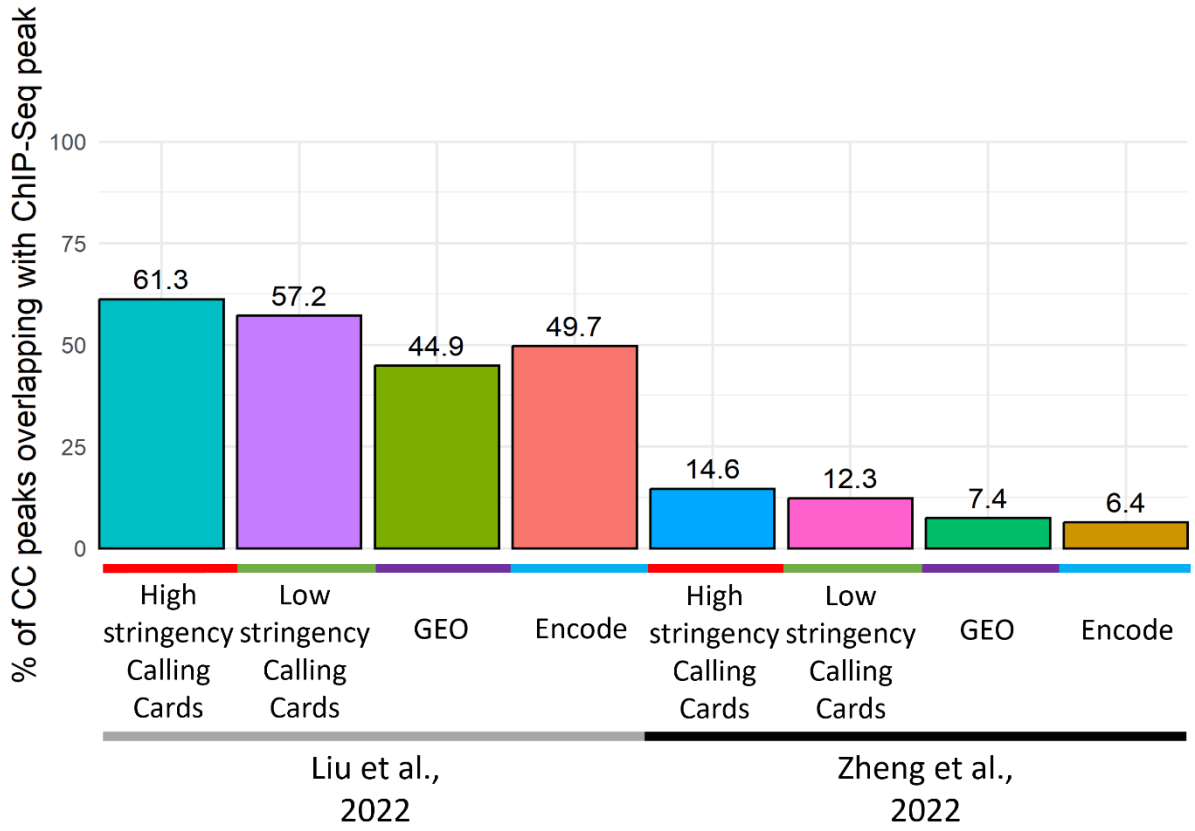


Figure 3.21: Calling card peaks overlap with BRD4 ChIP-Seq peaks more than ERα ChIP-Seq peaks by up to 17%.

Bedtools window was used to determine whether the calling card and ChIP-Seq overlap with BRD4 ChIP-Seq datasets. The majority of calling card peaks overlapped with the Liu et al., 2022 dataset (GSE171908), whereas the ChIP-Seq datasets both had less than 50% of peaks overlapping. In comparison to the Zheng et al., (2022) dataset, overlap with Liu et al. peaks was much higher, with ~30-40% less overlap between Zheng et al. peaks (GSE186645) and either ERα calling card or ChIP-Seq peaks. Calling cards still overlap the Zheng et al. peaks more than ChIP-Seq, with around double the percentage of overlapping peaks. These data show that calling card peaks appear to be located at sites that are either more likely to have BRD4 binding or are false positive binding sites identified due to undirected background insertions. Calling card enrichment of BRD4 was determined to be statistically significant through hypergeometric testing ($p > 1 \times 10^{-300}$).

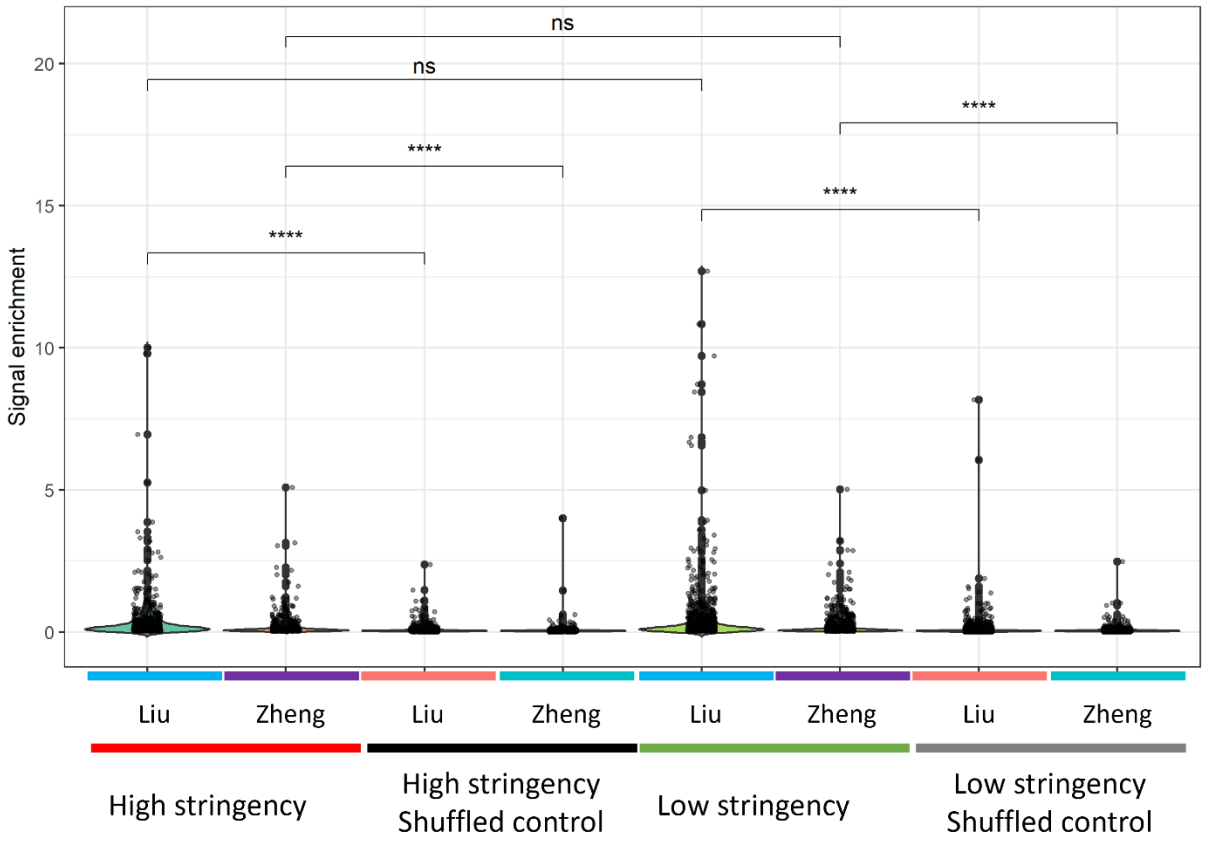


Figure 3.22: The BRD4 ChIP-Seq signal is enriched at calling card peaks. Violin plots display the mean BRD4 ChIP-Seq signal intensity within 1 kb of both high and low stringency HyPB-ERα calling card peaks compared to matched shuffled controls.

Calling card peak BED files were used as inputs for the BedTools shuffle function to generate random control regions of equivalent size and distribution. In both peak sets, BRD4 signal was significantly enriched relative to shuffled controls, indicating preferential overlap of calling card peaks with BRD4 binding sites. Statistical significance was assessed using Wilcoxon rank-sum tests, with p-value thresholds denoted as follows: p < 0.05 = *, < 0.01 = **, < 0.001 = ***, < 0.0001 = ****.

3.2.12. Calling Card peaks overlap with ERα ChIA-PET 3D interaction anchors more than ERα ChIP-Seq peaks

The ERα has the propensity to create long-range loops and contacts in DNA across chromosomes, much larger than 1000 bp (Fullwood et al., 2009; Pan et al., 2008; Liu and Cheung, 2014), and it's therefore possible that the flexibility of our HyPB-ERα fusion protein could enable insertion across chromatin loops. Given that the linker domain in the HyPB-ERα fusion protein allows deposition of SRT insertions at sites within 1000 bp, we hypothesise that our fusion can deposit calling cards across long-range chromosomal loops in proximity to ERα.

To investigate whether ERα calling card peaks can reflect long-range chromatin interactions, we compared the genomic overlap with ERα ChIA-PET anchors and ChIP-Seq peaks. ERα ChIA-PET data were obtained from the UCSC ENCODE/GIS-Ruan data from Fullwood et al. (2009), which mapped ERα-mediated chromatin loops in MCF7 cells (Fullwood et al., 2009). The method utilises ChIP against ERα on MCF7 chromatin, followed by an additional proximity ligation step and paired-end tag (PET) addition for sequencing (Fullwood et al., 2009). The ChIA-PET data were in a bigBED12 format, in which every two rows mark an interaction between two distant chromosomal regions that were brought into proximity by ERα.

We split rows 1 and 2 of the ChIA-PET interactions into separate bed files, referred to as anchor 1 and 2, respectively, to calculate the percentage overlap (within 1000 bp) for both ERα calling card and ChIP-Seq peaks (Figure 3.23 A). Both high and low stringency peak sets displayed significantly elevated ChIA-PET signal compared to shuffled controls (Wilcoxon $p < 1 \times 10^{-190}$ for high stringency peaks and $p < 1 \times 10^{-300}$ for remaining samples, Appendix Table 8.6), confirming that calling card insertions are significantly enriched for ChIA-PET anchors.

For anchor 1, 71.8% of calling card peaks overlapped compared to ~64% for ChIP-Seq peaks. For anchor 2, ~69% of calling card peaks overlapped versus ~61% of ChIP-Seq peaks. Therefore, calling card peaks showed greater overlap with both ChIA-PET anchors compared to ChIP-Seq Peaks.

When combining the overlap across either anchor (peaks overlap anchor 1 or 2), the percentage overlap for calling card peaks increases to ~81% compared to the 72% overlap for ChIP-Seq peaks.

Notably, an additional ~10% of ERα calling card peaks overlapped either anchor compared to anchor 1 alone, the anchor with the highest amount of overlap, indicating these peaks independently bound a genomic locus more than 1000 bp away. These findings suggest that HyPB-ERα calling card insertions are distributed across both ends of ERα-mediated chromatin loops.

The ChIA-PET signal was found to be significantly enriched at calling card peaks compared to random shuffled controls

This pattern supports the idea that ERα calling cards can capture spatial proximity through insertion within distal looped regions. Given that calling cards record the cumulative history of ERα binding over a period of time rather than a single snapshot of occupancy, they are well

suited to capturing transient or sporadic interactions that may contribute to chromatin looping events, which could help explain the strong concordance observed with ChIA-PET anchors. While ChIP-Seq shows a similar increase in combined anchor overlap, our calling card method may provide complementary insight into 3D chromatin architecture through insertions at looped loci that lack direct ERE motifs.

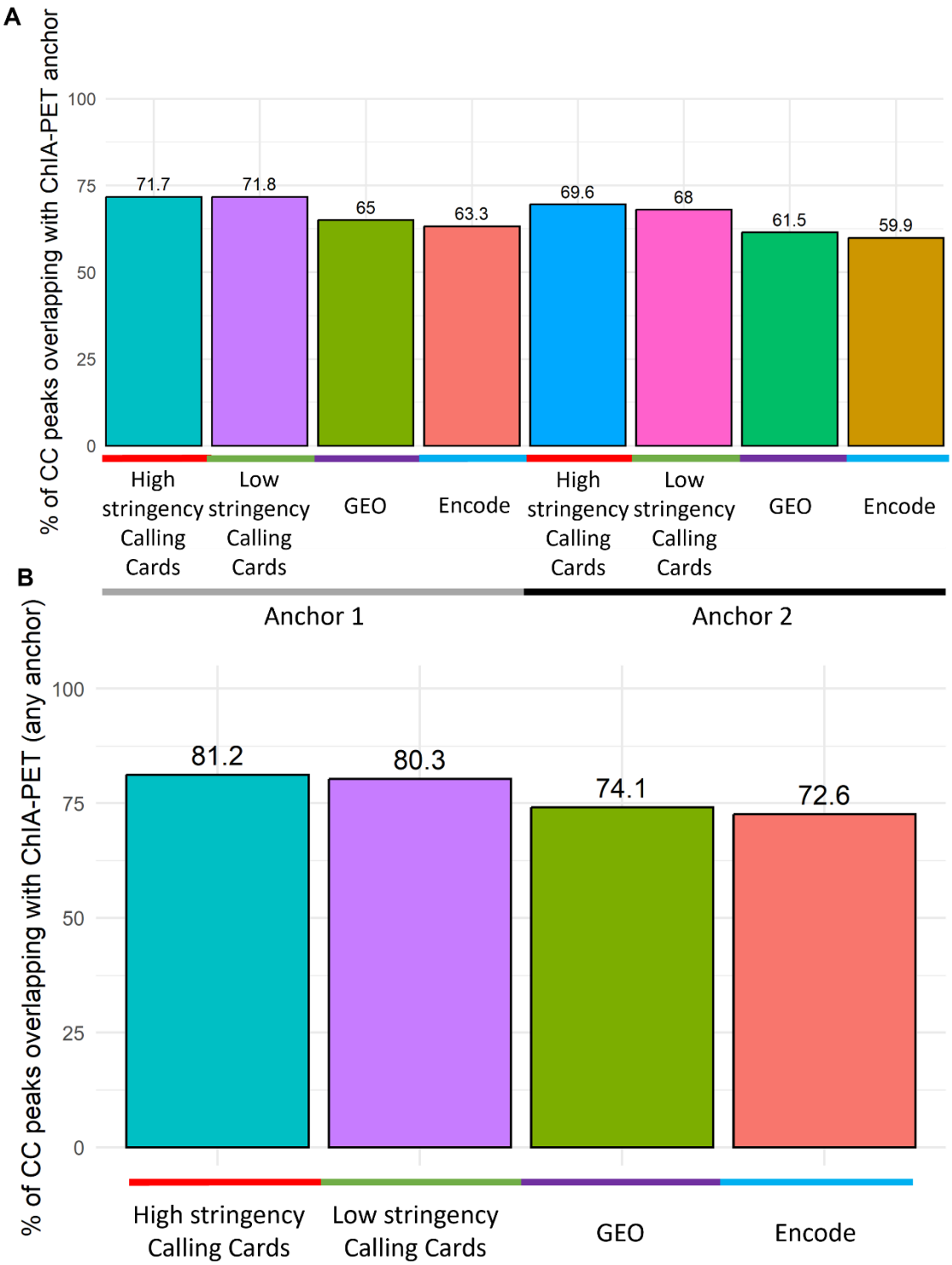


Figure 3.23: Up to 81.2% of HyPB-ERα directed calling card peaks overlap with ERα ChIA-PET anchors compared to up to 74.1% overlap from ChIP-Seq datasets.
Bedtools window was used to determine whether the calling card and ChIP-Seq overlap with existing ChIA-PET datasets. **(A)** The majority of both sets of peaks overlapped with either end of a ChIA-PET interaction, with calling cards showing ~6-10% more overlap than ChIP-Seq. **(B)** Overlap with either of the ChIA-PET anchors increases overlap to up to 81.2% for calling cards and 74.1% in ChIP-Seq peaks. This amount of overlap brings calling cards more in line with previous reporters that show up to 83% overlap with ChIP-Seq peaks. Calling card enrichment of ChIA-PET anchors was determined to be statistically significant through hypergeometric testing ($p > 1 \times 10^{-190}$ for high stringency and 1×10^{-300} for others).

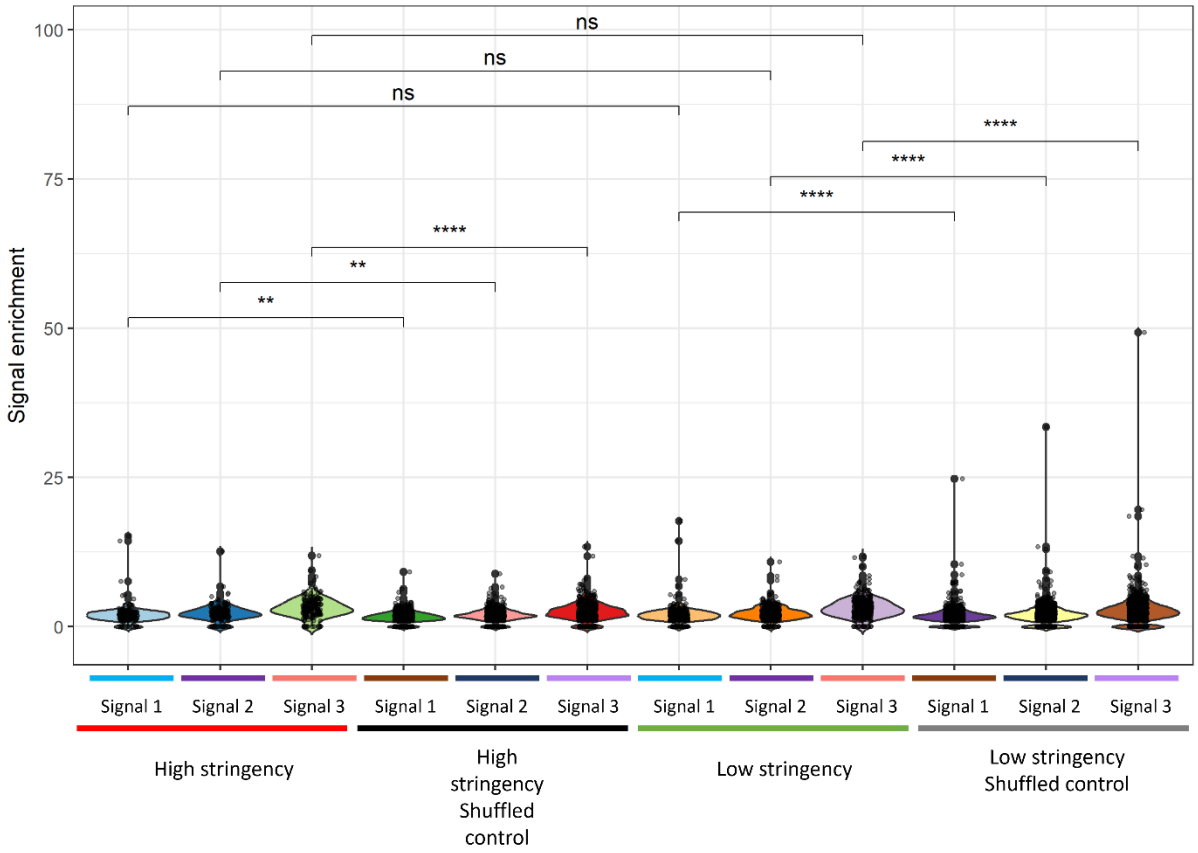


Figure 3.24: ERα ChIA-PET signal is enriched at calling card peaks.

Violin plots show mean ChIA-PET signal intensity within 1 kb of high and low stringency calling card peaks compared to matched shuffled controls. Calling card peak BED files were used as inputs for the BedTools shuffle function to generate random control regions of equivalent size and distribution. Both peak sets were significantly enriched relative to shuffled controls, indicating preferential overlap of calling card peaks with ChIA-PET anchors. P-values from Wilcoxon rank-sum tests are indicated for each comparison (p< 0.05 = *, < 0.01 = **, < 0.001 = ***, < 0.0001 = ****).

3.2.13. Significant HyPB-ERα Calling Card peaks are located at the genomic loci of putative ERα targets

To assess whether ERα-directed calling card insertions occur at biologically relevant loci, peak regions were visualised with the draw.area function of PyCallingCards (Guo et al., 2024) to determine the proximity of known ERα targets. Several key ERα target genes were examined in detail.

A prominent peak was identified downstream of the GATA3 gene (Figure 3.25 A), which encodes an essential ERα target gene for mammary gland development and a well-established ERα co-regulator (Sandström et al., 2024; Lee et al., 2017a; Hoch et al., 1999; Fang, Chen and Weigel, 2009). The density of ERα-directed insertions within this region

exceeded background levels, supporting the specificity of the signal. Notably, the called peak extended beyond the cluster of experimental insertions.

A second peak was located downstream of FOXA1 (Figure 3.25 B), a pioneer factor required for ERα binding and transcriptional activity (Robinson and Carroll, 2012; Fu et al., 2023; Bernardo and Keri, 2012; Hurtado et al., 2011; Lee et al., 2017a). This locus exhibited a higher density of undirected insertions compared to GATA3, consistent with known BRD4 co-binding at this site (Nagarajan et al., 2014; Angus et al., 2021). Despite the elevated background signal, the number of ERα-directed insertions was sufficient to each statistical significance, indicating that genuine ERα binding can be detected even in regions of high transposase activity.

Upstream of the GREB1 gene, which is involved in E2-driven proliferation and enhancer activation (Hodgkinson et al., 2018; Haines et al., 2020; Sun, Nawaz and Slingerland, 2007; Cheng, Michalski and Kommagani, 2018), a significant peak was observed in both high and low stringency datasets (Figure 3.25 C). Although the number of background insertions was more frequent at this locus and was closer to the ERα-directed insertions than other peaks, experimental insertions were significantly enriched, supporting the validity of the peak call.

An additional peak was identified downstream of the GREB1 gene only in the low stringency dataset (Figure 3.25 D). The presence of this peak exclusively in the more relaxed dataset highlights the utility of lower stringency parameters in capturing additional ERα binding events.

Collectively, these examples demonstrate that the HyPB-ERα calling card can identify insertion sites at known ERα target loci, including within regions that display elevated levels of undirected insertions. The detection of HyPB-ERα calling card peaks at GATA3, FOXA1 and GREB1 genes supports the biological relevance of the dataset and underscores the value of relaxed peak calling thresholds, uncovering additional binding sites.

Figure 3.25: Calling Card peaks are located at genomic loci of putative ER targets GAT3, GREB1 and FOXA1.
Peaks found adjacent to GATA3 **(A)**, FOXA1 **(B)**, and upstream of GREB1 **(C)** were shared across both sets of peaks. Whereas the peak downstream of GREB1 **(D)** was only identified in the low stringency peaks. Identification of peaks adjacent to bona fide ER binding locations indicates that the fusion protein can redirect wild-type HyPB insertions to genuine ER binding sites.

3.2.14. HyPB-ERα fusion calling cards are enriched for the ERα binding partner motifs.

To assess whether ERα calling card peaks reflect known transcriptional networks, we performed motif enrichment analysis using HOMER via the PyCallingCards package. HOMER identifies overrepresented DNA motifs by comparing peak sequences to matched genomic background for calculating statistical significance; it is often used to assess data quality and biological plausibility (Heinz et al., 2010; Jeon et al., 2020; Devailly et al., 2015).

Across the high and low stringency peak sets, HOMER identified 109 and 157 significantly enriched motifs ($p < 0.01$), respectively. Notably, the ERE motif – The ERα's binding site – derived from ERα ChIP-Seq in MCF7 cells was highly ranked in both datasets (Figure 3.26).

The motif with the smallest p-value in the high stringency dataset was the ERE motif from ovary-derived ChIP-Seq, while the most significant motif in the low stringency dataset was JUN-AP1, identified from c-Jun ChIP-Seq in K562 cells.

Within the top 10 motifs ranked by adj-p-value, 5 proteins were known to colocalise with ERα in the active transcription complex, indicating that these sites are highly enriched within calling card peaks, supporting the validity of HyPB-ERα calling card peaks.
including FOXA1 (Carroll et al., 2005; Hurtado et al., 2011; Swinstead et al., 2016), GATA3 (Eeckhoute et al., 2007; Wilson and Giguère, 2008; Porras, Ismail and Mader, 2021), RXR (Ross-Innes et al., 2010; Shao et al., 1994; Johansson et al., 2013), MYB (Drabsch et al., 2007; Mitra et al., 2012; Miao et al., 2011). We have already shown that the pioneer factors FOXA1 and GATA3 closely cooperate with ERα to establish binding (Carroll et al., 2005; Hurtado et al., 2011; Swinstead et al., 2016; Eeckhoute et al., 2007; Wilson and Giguère, 2008; Porras, Ismail and Mader, 2021), Whereas RXR and MYB function to enhance ERα target gene transcription.

RXR is a nuclear receptor that heterodimerises with RAR and other NRs, which has been shown to co-localise with ERα to enhance target gene expression (Ross-Innes et al., 2010; Shao et al., 1994; Johansson et al., 2013). E2 stimulates transcriptional cooperativity between the two proteins, and RXR/RAR are thought to enhance transcription through stabilising co-factor binding.

The other key factor we identified was MYB, an essential effector of the ERα transcriptional programme (Drabsch et al., 2007; Mitra et al., 2012; Miao et al., 2011). While not a direct ERα target gene, the receptor does recruit factors to overcome pausing of MYB transcription, and in turn, MYB functions as a critical coordinator of estrogen-stimulated cell proliferation.

These results confirm that calling card insertions occur at canonical ERα binding sites due to the significant enrichment of key ERα binding partner motifs across peaks. These motifs reflect known co-regulators of ER and suggest that the calling card system captures broader transcriptional complexes. Motif p-values were generally more significant in the low stringency dataset, likely due to the larger number of peaks increasing motif representation.

These results demonstrate that ER calling card peaks are enriched for motifs associated with ER and its transcriptional partners, reinforcing the biological relevance of the insertion sites.

Motif name	Motif	High stringency peak set p-value	Low Stringency peak set p-value
ERE		1 x10^{-39}	1 x10^{-46}
FOXA 1		1 x10^{-3}	1 x10^{-8}
GATA 3		1 x10^{-3}	1 x10^{-7}
RXR		1 x10^{-3}	1 x10^{-6}
MYB		1 x10^{-3}	1 x10^{-6}

Figure 3.26: Motifs for known ER binding partners were discovered in the high stringency and low stringency peak sets with significant p-values.

Motifs reported here correspond to known ER binding partners, indicating that the peaks called in both datasets are genuine. The peaks called with less stringent parameters contain lower p-values for motifs discovered at these sites, indicating increased significance.

3.3. Discussion

3.3.1. Summary of results

The generation of ERα–HyPB fusion proteins in both N- and C-terminal orientations, together with validation of full-length expression, and confirmation of transposase activity, establishes the technical feasibility of this approach. Importantly, the pilot bulk calling card experiment using the puromycin-resistant SRT revealed that, despite variability in insertion numbers and relatively low sensitivity and specificity, the majority of calling card loci overlapped with ERα ChIP-Seq peaks. This overlap, reinforced by motif analysis identifying bona fide ERα-binding sites through the presence of the ERE and motifs of known members of the ERα complex, confirms that the fusion protein can recapitulate endogenous ERα binding.

Beyond simple overlap with ChIP-Seq, the calling card data revealed notable enrichment of peak insertions at distal intronic regions, consistent with endogenous ERα function (Theodorou and Carroll, 2010; Liu and Cheung, 2014; Pan et al., 2008; Fullwood et al., 2009). When ERα calling card peak loci were compared with ERα ChIA-PET anchors, both relaxed- and strict- threshold calling card peaks showed increased overlap relative to ChIP-Seq. This suggests that calling card insertions may capture aspects of ERα chromatin interactions that are not fully resolved by ChIP-Seq alone, particularly long-range looping events.

Taken together, these findings demonstrate that the HyPB-ERα fusion protein can map the endogenous ERα binding patterns through calling card insertions and downstream analysis, providing a complementary perspective to existing ChIP-based approaches.

3.3.2. Fusion protein design and cloning implications on function

The design of HyPB-ERα fusion proteins required careful consideration of partner orientation and linker composition, as both factors directly influence functional integrity (Wang et al., 2012a; Moudgil et al., 2020; Chen, Zaro and Shen, 2013; Arai et al., 2001; Klein et al., 2014). While restriction digest cloning proved challenging, NEBuilder HiFi assembly offered a reliable alternative, yielding integrated clones on the first attempt (Figure 3.1 and Figure 3.2).

Constructs were generated with the HyPB partner at either the C- and N-terminus of the ERα coding sequence to determine whether domain positioning would affect activity of either partner. Previous studies report the use of an N-terminal fusion protein of ERα. (Zhou et al., 2014; Larson et al., 2021; Powell and Xu, 2008) as well as utilising C-terminal fusions (Hayashi and McMahon, 2002; Kalaitzidis et al., 2004; Kurata et al., 1999; Powell and Xu, 2008; Wen

et al., 1999; Shim et al., 1997; Kolligs et al., 2002). Notably, it does appear more common to create C-terminal fusions of the ERα in which the ERα partner is at the end of the protein sequence. The prevalence of C-terminal ERα fusion proteins in the literature is notable for the results of this chapter, given the improved performance of our C-terminal fusion over the N-terminal (Figure 3.5). Further work is required to determine the cause of the N-terminal fusion protein's impaired function compared to the C-terminal fusion, but several key aspects of fusion protein design could impair function (Chen, Zaro and Shen, 2013; Arai et al., 2001; Klein et al., 2014).

The linker used to bridge the ERα and HyPB domains was previously validated for other TFs (Moudgil et al., 2020), but differences in structure and biology between the TFs used in that publication and steroid hormone receptors like ERα may necessitate testing alternative linkers. This thesis is the first application of calling card reporting to a NR, as the other mammalian reporters are SP1 (O'Connor, Gilmour and Bonifer, 2016; Beishline and Azizkhan-Clifford, 2015), FOXA2 (Iwafuchi-Doi et al., 2016; Liu et al., 2024a) and BAP1 (Masclef et al., 2021; Kwon, Lee and Lee, 2023; Habara and Shimada, 2022). SP1 is a classical sequence-specific TF that is ubiquitously expressed as a housekeeping TF with some context-dependent regulatory mechanisms (O'Connor, Gilmour and Bonifer, 2016; Beishline and Azizkhan-Clifford, 2015). FOXA2, like FOXA1, is a pioneer factor that, unlike FOXA1, can act as a tumour suppressor through the stimulation of E-cadherin expression, downregulating the EMT (Iwafuchi-Doi and Zaret, 2014; Zhang et al., 2015; Liu et al., 2021). Finally, the BRCA1-associated protein (BAP1) does not bind DNA directly and is instead recruited to specific regions of the genome through association with other proteins (Sahtoe et al., 2016; Campagne et al., 2019). Crucially, none of these proteins are ligand-activated steroid receptors that undergo dimerisation, meaning the linker may not function as well with ERα.

The reduced mRNA levels of the N-terminal fusion may reflect steric hindrance or suboptimal linker properties. Such interference is a well-recognised issue in fusion protein design, arising from inappropriate length, charge or composition that disrupts fusion partner function (Monterrey et al., 2022; Chen, Zaro and Shen, 2013).

The protein linker may not be the only contributing factor to steric hindrance, however, as orientation effects further compound these challenges. The C-terminal fusion retained its ERα activity, consistent with prior kinase fusion protein designs in which ERα was placed at the C-terminal end without impairing LBD function (Pritchard et al., 1995; Scherrer et al., 1993; Kohn et al., 1998). By contrast, the N-terminal fusion protein positioned HyPB adjacent to the LBD, where an additional protein sequence could obstruct ligand binding or subsequent

conformational change (Kumar et al., 2011). The steric hindrance may also extend to protein-protein interactions, as ERα functions within the multi-protein estrogen-responsive complex (Lebedeva et al., 2012; Métivier et al., 2008, 2003; Shang et al., 2000; Naughton et al., 2007; Lee et al., 2017a). Although most co-regulatory interactions are coordinated by the AF1 domain in the N-terminus (Yaşar et al., 2017; Liu and Cheung, 2014; Safe, 2001), some occur within the AF2 and F domains at the C-terminal end (Yaşar et al., 2017; Montano et al., 1995; Koide et al., 2007; Arao et al., 2013). Thus, while linker engineering could potentially alleviate interference, the consistent functionality of the C-terminal fusion makes it a more practical construct for future experiments.

However, given the evidence of successful integration of calling cards into the MCF7 genome during the functional validation (Figure 3.4 A and B), the N-terminal fusion protein is at least partially functional, so perhaps the steric hindrance prevents homodimerization (Panter and Jerala, 2011; Chen, Zaro and Shen, 2013). Previous fusion protein studies have reported the inhibition of dimerisation due to the addition of large globular protein partners in some fusion proteins (Panter and Jerala, 2011; Chen, Zaro and Shen, 2013; Reuten et al., 2016). The steric hindrance in HyPB-ERα could be weak enough to allow heterodimers between the endogenous ERα and our fusion protein to form, allowing some calling card depositions. However, if two N-terminal fusion proteins attempt to dimerise, the position of the additional protein domain may harm the function of the ERα domain's ability to form a functional transcription complex. This interpretation supports the findings of initial calling card insertions coordinated by the N-terminal HyPB-ERα reporter, but failure to sustain long-term cell survival.

Additional construct design choices shaped performance, including site-directed mutagenesis of the HyPB coding sequence and promoter selection. The HyPB coding sequence gifted by the Mitra lab (Moudgil et al., 2020) originally contained a Myc tag sequence, which was removed to avoid unnecessary epitopes that could interfere with function, given the high quality of ERα antibodies. Similarly, promoter selection was optimised for expression in MCF7 cells; the gifted SRTs already contained EF1a promoters for reporter expression, which data from our lab indicates is the second-best performing promoter for MCF7s (Appendix Figure 8.9). We selected the CMV promoter for HyPB-ERα expression as it was shown to drive stronger exogenous expression, which was considered advantageous due to ERα's function as a dimer. Elevated protein levels would therefore increase the likelihood that dimers contain at least one fusion protein containing

Together, these design considerations highlight why the C-terminal fusion emerged as the more robust HyPB-ERα construct. Its orientation preserved ligand binding and protein

interactions, while promoter-driven expression maximised functional incorporation into dimers. In contrast, the N-terminal fusion faced structural and functional barriers that limited its utility, positioning the C-terminal construct for use in subsequent experiments.

3.3.3. Western blot analysis confirms full-length HyPB-ERα expression

Western blot analysis confirmed expression of the full-length HyPB-ERα fusion proteins, with a prominent band at the expected molecular weight of 136 kDa (Figure 3.3 A and B). This provides direct evidence that both constructs were successfully expressed as complete proteins.

In the C-terminal transfections, however, an additional strong band appeared just above 100 kDa, comparable in intensity to the expected band. One explanation is the alternative transcription initiation due to the absence of a Kozak sequence in the expression plasmid, which can improve start site fidelity (Valášek, 2013; Hernández, Osnaya and Pérez-Martínez, 2019). Such downstream initiation would generate a truncated mRNA, and, consequently, a shorter fusion protein. Yet, this cannot be the sole factor, as the N-terminal construct also lacked a Kozak sequence but did not produce the same band. Other possibilities include proteolytic digestion resulting from protein overexpression; however, this seems unlikely as comparable fragments were not observed in N-terminal transfections.

Smaller bands at ~50 kDa and below are only detected by the application of antibodies targeting the respective termini probed (Abcam N-terminal, Santa Cruz C-terminal), raising questions about their origin. The literature reports intracellular cleavage of ERα in uterine and breast cancer, producing fragments at 50, 37 and 28 kDa (Maaroufi et al., 2000; Horigome et al., 1988). Some of the faint bands in whole blots (Appendix Figure 8.2 and Figure 8.3) may correspond to these isoforms, though their detection with N-terminal antibodies would be inconsistent with published findings.

Taken together, the previously reported proteolytic cleavage of ERα is inconsistent with the 100 kDa band, suggesting that alternative transcription initiation is a more plausible explanation. It would therefore be possible to improve full-length protein expression through the addition of a Kozak sequence, which will be investigated in the next chapter. We have demonstrated that both fusion proteins are fully expressed following transfection into MCF7 cells.

3.3.4. BrokenHeart and TdTomato Functional validation confirms the function of the HyPB domain in both fusion proteins

Functional validation using BrokenHeart and TdTomato SRT plasmids demonstrated that the HyPB domain retained activity in both fusion orientations, as evidenced by the red fluorescent cells and colonies (Figure 3.4). These results confirm that the transposase partner in both fusions could mediate genomic integration of SRTs.

Notably, some cells and colonies from C-terminal transfections appeared brighter than those from N-terminal transfections, with a wide range of signal intensity heterogeneity (Appendix Figure 8.10). One interpretation of this finding is that the N-terminal fusion was less efficient, consistent with reduced mRNA levels. Alternatively, the observation of brighter cells in C-terminal transfections could indicate the function of the ERα, consistent with the high levels of ERα activity reported in MCF7 cells (Liao et al., 2014; Brooks, Locke and Soule, 1973; Ş. Comşa, Cîmpean and Raica, 2015). This result aligns with later findings that only the C-terminal fusion generated sufficient mRNA for analysis. At the same time, heterogeneity in colony intensity raises questions about the variability of HyPB-ERα binding and insertional capacity across cells that require further investigation.

Overall, these results confirmed that the HyPB domain was functional in both constructs but highlighted the orientation-dependent differences in activity. To fully assess whether the ERα domain retained its function, we proceeded to bulk calling card experiments using puromycin selection, enabling validation of the ERα partner of the fusion protein.

3.3.5. HyPB-ERα bulk calling cards were optimised to achieve sufficient material for library preparation

Establishing bulk ERα calling cards required optimisation of selection conditions with the puromycin resistance SRT, to determine the concentration to achieve complete selection. We determined 2 µg/mL to be the optimal concentration to achieve 100% death of non-transfected cells after three days (Appendix Figure 8.4). Initial bulk CC attempts failed due to insufficient material, as cells subjected to the published method of Lipofectamine transfection, passage, and puromycin treatment within 24 hours exhibited excessive cytotoxicity and cell death. To overcome this, we modified the protocol from the published method (Moudgil et al., 2020), allowing cells to reach 70-80% confluence before puromycin selection. These adjustments yielded sufficient material for analysis (Figure 3.5 and Figure 3.6), with further refinements described in the following chapter. This is when the N-terminal fusion was found to still

generate insufficient material after the method improvement, and the C-terminal fusion was carried forward as previously discussed.

Replicates are critical for calling card experiments, as more replicates increase the statistical power to distinguish genuine ERα-directed insertions from background. In this pilot, six replicates per condition were sequenced in line with recommendations, although future experiments would ideally utilise 8 to 12 replicates in bulk experiments.

3.3.6. Bulk Calling Card data analysis

3.3.6.1. Assessing reproducibility of HyPB-ERα bulk calling cards through normalised insertions

A critical step in validating the HyPB-ERα reporter was to assess whether insertions were reproducible across replicates. Given that each insertion represents a biologically meaningful integration event, reproducibility provides confidence that peaks reflect genuine ERα-directed activity rather than undirected HyPB insertions. Correlation analysis of normalised insertions demonstrated a moderate pattern across replicates (Gust and D'journo, 2015; Mukaka, 2012), supporting the robustness of the dataset (Figure 3.7).

In ChIP-Seq analysis, reproducibility is typically tested through Irreproducible Discover Rate (IDR) analysis to distinguish true biological signal from noise (Landt et al., 2012; Bailey et al., 2013; Nakato and Shirahige, 2017); however, IDR is not standard in calling card pipelines due to fundamental differences in data structure. Calling card peak calling relies on clusters of insertions per locus with unique sample barcodes (Moudgil et al., 2020; Wang et al., 2012a), making pseudoreplicate generation problematic. Each insertion carries biological meaning in distinguishing ERα-directed from undirected transposase activity, unlike ChIP-Seq reads that can be resampled. Consequently, previous studies have demonstrated reproducibility through correlation of normalised insertions rather than IDR

Interestingly, reproducibility was higher in the low-stringency set of peaks compared to the high-stringency set, a pattern also reflected in the ROC/AUC comparisons with ChIP-Seq datasets (Figure 3.19 and Figure 3.20). This likely reflects the greater number of low-stringency peaks (935 vs. 506), which reduces sensitivity to variability across replicates (Landt et al., 2012; Li et al., 2011). While the stringent cut-offs increase confidence in individual peaks, they may exclude reproducible, lower signal events. The correlation patterns, therefore,

suggest that low-stringency parameters capture a broader set of consistent ERα binding events, albeit with reduced statistical certainty.

The bulk ERα calling card experiment yielded fewer peaks than expected, particularly compared to the ~25,000 peaks reported for other calling card reporters (Moudgil et al., 2020). Several factors may explain this reduction, including insufficient replicates, elevated undirected background signal and impaired fusion protein function. We demonstrate that every replicate is needed for HyPB-ERα calling cards, as over 95% of peaks are present in all replicates and if any fewer than all replicates are used in peak calling, no significant peaks are identified (Figure 3.8). Regardless of the cause, the limited peak pool increases sensitivity to variability and complicates reproducibility assessments.

Taken together, these results indicate that the reproducibility of HyPB-ERα calling cards can be demonstrated through correlation analysis and that MACCs parameter choice influences the balance between stringency and consistency. For future experiments, increasing the replicate numbers will be essential to strengthen reproducibility assessments and allow more stringent thresholds without loss of genuine ERα peaks. The reproducibility trends here provided sufficient confidence to proceed with overlap analysis against ChIP-Seq datasets, enabling direct evaluation of whether peaks identified correspond to genuine ERα binding sites in MCF7 cells.

3.3.6.2. HyPB-ERα calling card peaks display significant correlation with the distribution of ERα ChIP-Seq peaks across chromosomes

Analysis of genome-wide distribution revealed that both calling cards and ChIP-Seq peaks deviate significantly from uniform expectation, confirming that ERα binding is not random but biased towards specific chromosomes.

Chromosomes 17 and 20 show a striking over-representation in calling card data, with peak counts exceeding 150-250% of expected values. This enrichment was more pronounced than in ChIP-Seq datasets, which could suggest that calling cards can capture aspects of ERα binding within these regions that are not captured by ChIP-Seq. This is particularly notable given that both chromosomes 17 and 20 have been previously highlighted as ERα binding hotspots with many ERE sites that are frequently amplified in ERα+ tumours (Hsu et al., 2013; Lee et al., 2023a; Palaniappan et al., 2019; Wang et al., 2013). The increased representation of calling card peaks within these regions from the expected values and the ERα ChIP-Seq observed peaks could suggest an increased sensitivity for HyPB-ERα calling cards in

detecting binding within these regions. However, given the smaller pool of peaks than the previous calling card reporters (Moudgil et al., 2020), it is unclear if this is a genuine feature of HyPB-ERα calling cards or an artefact of under sampling peaks; therefore, this pattern should be investigated in the future. Other regions, like chromosome 6, also showed significant enrichment, while X and 13 were under-represented in HyPB-ERα peaks.

Despite these differences, overall distribution patterns were broadly similar between methods. Pearson correlation analysis demonstrated strong within-method consistency ($r > 0.95$, $p < 0.001$) and moderate (Gust and D'journo, 2015; Mukaka, 2012) but significant agreement between calling cards and ChIP-Seq ($r > 0.56$-65, $p < 0.001$). This concordance supports the conclusion that HyPB-ERα calling cards reflect endogenous ERα binding, even though variability at individual chromosomes remains.

Artefactual peaks were also observed, most notably on the Y chromosome. MCF7 cells lack a Y chromosome as they arise from a pleural effusion from a 69-year-old woman (Brooks, Locke and Soule, 1973; Russo et al., 1976), but these signals originate from mis-mapping repetitive sequences, a recognised issue in ChIP-Seq data analysis (Amemiya, Kundaje and Boyle, 2019; Webster et al., 2019; Teytelman et al., 2013; Marinov et al., 2014). These sequences are also to be expected in calling cards due to the insertional nature of the HyPB transposase, making it possible to create repetitive sequences that could be mis-mapped to the Y chromosome (Galvan et al., 2009; Li et al., 2013a, 2013c).

Taken together, these findings show that HyPB-ERα calling cards reproduce the broad chromosomal distribution of ERα binding observed in ChIP-Seq, while also revealing distinctive biases that may reflect co-regulator interactions or methodological differences. The enrichment of chromosomes 17 and 20 in calling card peaks is particularly notable and provides a clear direction for future work to investigate and explore these functional hotspots of ERα activity in MCF7 cells.

3.3.6.3. The feature distribution patterns are broadly similar between HyPB-ERα calling card peaks and ERα ChIP-Seq peaks.

The distribution of HyPB-ERα calling card peaks across genomic features closely resembled that of ERα ChIP-Seq, supporting the conclusion that the fusion protein recapitulates endogenous binding behaviour (Figure 3.13 and Figure 3.14). Consistent representation in intronic, UTR, and downstream regions provides confidence that the ERα fusion partner remains functional within the construct.

Notable differences emerged at promoters and distal intergenic regions. Promoter-associated peaks were less frequent in calling card data compared to ChIP-Seq, while distal intergenic peaks were more abundant. These shifts suggest that calling card may preferentially capture enhancer-associated binding events, consistent with ERα's established role in long-range transcriptional regulation (Carroll et al., 2005, 2006; Fullwood et al., 2009; Pan et al., 2008). The positional distribution relative to transcription start sites reinforces this interpretation: calling cards showed fewer peaks within 1 kb of the TSS but more peaks beyond 5 kb, mirroring the increased representation of ERα binding in distal intergenic regions (Carroll et al., 2005, 2006; Fullwood et al., 2009; Pan et al., 2008).

Taken together, these findings indicate that while HyPB-ERα calling cards broadly reproduce the genomic feature distribution of ChIP-Seq peaks, they also highlight enhancer-associated binding patterns more strongly. This bias may reflect methodological differences in how transposase insertions capture binding events, but it also underscores the potential of calling cards to provide complementary insights into ERα's regulation of distal elements. Future work should test whether this enhancer-associated bias is consistent across cell types and experimental conditions, as this would strengthen the case for calling cards as a tool to study long-range ERα activity.

3.3.6.4. The majority of HyPB-ERα calling card peaks show overlap with ChIP-Seq peaks, comparable with other calling card reporters.

Comparison of HyPB-ERα calling card peaks with ChIP-Seq datasets demonstrated the substantial overlap, providing strong evidence that the ERα fusion partner retains its binding specificity (Figure 3.15). Up to 70% of high stringency calling card peaks overlap with GEO ChIP-Seq peaks, a level comparable to previously reported SP1-HyPB fusion (Wang et al., 2012a; Moudgil et al., 2020). Even under lower stringency parameters, most peaks overlapped (63.3%), confirming that calling cards capture genuine ERα binding events.

Overlap with ENCODE ChIP-Seq peaks was slightly lower across both sets of calling card peaks (64% for high, and 58.4% for low stringency), but still represented most peaks. These values, while modestly reduced compared to GEO, remain consistent with the expectation of calling card depositions within ~1 kb of TF binding sites (Wang et al., 2012a; Moudgil et al., 2020; Wang et al., 2011; Cammack et al., 2020). Importantly, stricter cut-offs enriched for peaks within the immediate vicinity of ChIP-Seq sites, whereas looser cut-offs captured

additional peaks further away, reflecting the balance needed between sensitivity and specificity in peak calling.

The fact that 30-40% of calling card peaks did not overlap with ChIP-Seq does not necessarily indicate reduced reporter function. Instead, peaks may represent genuine ERα binding events that are lost or too weak in ChIP-Seq data, particularly at distal regulatory elements, which is consistent with earlier observations of enhanced representation and distal intergenic regions.

Signal enrichment further reinforced the concordance as both GEO and ENCODE ChIP-Seq datasets showed clear enrichment centred on calling card peaks with an average $\log_2$ fold change value of ~3.5 and ~7, respectively. Heatmaps confirmed that many of the top-ranked calling card peaks aligned with a strong, central ChIP-Seq signal, particularly within the high stringency dataset. While the lower-ranked peaks showed more variable enrichment, this pattern is consistent with the idea that some calling card insertions represent distal or indirect binding events that are not fully captured by ChIP-Seq.

ROC analysis of the adjusted p-value threshold highlighted the limitation of current peak-calling approaches' application to HyPB-ERα calling cards (Moudgil et al., 2020; Guo et al., 2024). For both ENCODE and GEO datasets, the AUC values were low (0.52-0.6), indicating that the adjusted p-value produced by MACCs is a poor discriminator of genuine ERα binding sites. Relaxing the stringency modestly improved sensitivity, but overall performance remained weak compared to other TF reporters (Moudgil et al., 2020). This discrepancy may reflect the limited replicate coverage, unique binding dynamics of NRs, or confounding by BRD4 co-binding (Nagarajan et al., 2015, 2014; Yang, He and Zhou, 2008; Zheng et al., 2022). These factors could reduce resolution between ERα-directed and undirected insertions, increasing false positives and negatives.

Taken together, despite modest ROC performance, the substantial overlap with ChIP-Seq peaks and strong signal enrichment at peak centres confirm the biological relevance of HyPB-ERα calling cards. These results demonstrate that the ERα fusion partner of the construct is functional and capable of mapping endogenous binding patterns, while methodological differences highlight areas for optimisation. Future work should increase replicate numbers and refine peak-calling strategies to improve discrimination, while also investigating the contribution of BRD4 co-binding to insertion profiles. These steps will strengthen the utility of calling cards as a complementary approach to ChIP-Seq for mapping ERα binding across the genome.

3.3.6.5. HyPB-ERα calling card peaks are enriched for BRD4 co-binding sites compared to ERα ChIP-Seq peaks

A unique challenge in the analysis of ERα calling cards using the HyPB transposase is the affinity of undirected HyPB for BRD4 (Moudgil et al., 2020), a known co-regulator of ERα binding (Zheng et al., 2022; Nagarajan et al., 2014). Because BRD4 frequently co-localises with ERα, undirected HyPB controls show a strong signal at ERα binding sites, as evident in the FOXA1 and GREB1 calling card peaks (Figure 3.25 B and C). It is unclear whether this elevated background has contributed to the relatively low number of peaks called, but calling card peaks do show an increased percentage of peaks overlapping BRD4 ChIP-Seq peaks compared to ERα ChIP-Seq peaks (Figure 3.21).

Initial attempts to process the dataset using the pipeline from Moudgil et al. (2020) (Moudgil et al., 2020) were hindered by this issue, but the release of PyCallingCards (Guo et al., 2024)Incorporating the MACCs peak caller provided a more compatible framework. Of the two available peak calling algorithms, MACCs and CCcaller (Guo et al., 2024), MACCs were selected for further analysis due to their similarity to MACS2 and comparable performance with CCcaller. Parameter optimisation was necessary to maximise genuine peak recovery, and given the limited number of peaks, generating two datasets with different stringency thresholds allowed exploration of ERα calling card performance under varying conditions.

Analysis of BRD4 ChIP-Seq datasets confirmed that ERα calling card peaks may be biased towards ERα:BRD4 co-binding sites, either reflecting genuine biology or technical bias introduced by HyPBs affinity for BRD4. One possibility is that co-binding sites accumulate both ERα-directed and undirected insertions, inflating their representation in the final set of peaks. This overrepresentation could also contribute to false positives, where background insertions are grouped with a nearby co-binding regions to be misclassified as a genuine ERα binding site. At the same time, the consistent enrichment across independent BRD4 datasets supports the idea that HyPB-ERα calling cards are capturing biologically relevant ERα:BRD4 co-binding events.

Taken together, these results reinforce that while HyPB-ERα calling cards reproduce genuine ERα binding, they are also enriched for BRD4 co-binding. The dual signal highlights both a strength and a limitation of this approach: calling cards may be particularly sensitive to ERαs enhancer-associated interactions with BRD4, but distinguishing these from background requires careful optimisation of peak calling and appropriate controls. Future work should incorporate additional replicates and undirected HyPB datasets to better resolve ERα-specific binding from BRD4-associated background.

3.3.6.6. HyPB-ERα calling card peaks show superior overlap with ChIP-PET anchors compared to ChIP-Seq peaks

ERα is known for conducting long-range contacts with enhancers (Theodorou and Carroll, 2010; Liu and Cheung, 2014; Pan et al., 2008; Fullwood et al., 2009), these interactions are not fully captured by ChIP-Seq, which sequences only DNA directly bound to the protein of interest. By contrast, calling cards are deposited into adjacent DNA within ~1000 bp of the binding site, meaning that insertions can occur at distal loci brought into proximity by ERα. If compared only to ChIP-Seq, such insertions might appear as off-target binding, even though they represent genuine ERα-directed events. This highlights both the potential and the challenge of bulk calling cards: while they may not exclusively map ERα binding at EREs, they could provide complementary insight into 3D chromatin interactions.

To test this, we compared ERα calling card peaks with ChIA-PET anchors from Fullwood et al. (2009) (Fullwood et al., 2009); the establishing publication for the method. was acquired and used to determine the overlap with calling card and ChIP-Seq peaks (Figure 3.23). Across both anchors, more calling card peaks overlapped more than ChIP-Seq peaks, with ~72% overlap at anchor 1 and ~69% at anchor 2, compared to ~64% and ~61% for ChIP-Seq. When overlap across either anchor was combined, calling card peaks reached ~81% overlap, higher than the ~72% for ChIP-Seq. This enrichment is stronger than the SP1-HyPB calling card peaks overlap for ChIP-Seq (72%), which strongly supports the function of our HyPB-ERα fusion protein. Furthermore, this high degree of ChIA-PET enrichment at calling card peaks, combined with the flexibility of reporter fusion proteins, opens the potential for applications in studying ERα-directed changes in 3D chromatin architecture.

At the same time, ~20% of calling card peaks did not overlap with either ChIP-Seq or ChIA-PET anchors. These may represent genuine ERα binding sites or long-range contacts that neither ChIP-Seq nor ChIA-PET can capture, reflecting shared limitations across chromatin capture approaches (Wang et al., 2012a; Moudgil et al., 2020; Fullwood et al., 2009; Gogol-Doring et al., 2016; Landt et al., 2012; Dekker, Marti-Renom and Mirny, 2013). Without an independent benchmark, it is not currently possible to validate these loci, but their presence highlights the need for further methodological refinement.

Taken together, these findings demonstrate that HyPB-ERα calling cards not only reproduce ERα conventional binding patterns but also enrich for long-range contacts captured by ChIA-PET. The superior overlap with ChIA-PET anchors compared to ChIP-Seq underscores the potential of calling cards to provide complementary information about ERαs 3D regulatory landscape. Future work should focus on refining peak calling strategies and exploring whether

non-overlapping peaks represent novel ERα interactions beyond the reach of current methods.

3.3.6.7. Top HyPB-ERα calling card peaks are located at known ERα binding sites, and relaxed peak calling reveals additional biologically relevant loci.

Annotation of calling cards peaks to nearby genes confirmed that HyPB-ERα fusion protein identifies biologically relevant ERα targets (Figure 3.25). Prominent peaks were detected at GATA3, FOXA1 and GREB1 – All well-established ERα-regulated target genes critical for breast cancer biology(Robinson and Carroll, 2012; Fu et al., 2023; Bernardo and Keri, 2012; Hurtado et al., 2011; Lee et al., 2017a; Sandström et al., 2024; Hoch et al., 1999; Fang, Chen and Weigel, 2009; Hodgkinson et al., 2018; Haines et al., 2020; Sun, Nawaz and Slingerland, 2007; Cheng, Michalski and Kommagani, 2018). GATA3 functions in a positive feedback loop with ERα expression (Lee et al., 2017a; Sandström et al., 2024; Hoch et al., 1999; Fang, Chen and Weigel, 2009), and FOXA1 acts as a pioneer factor required for ERα transcriptional activity (Robinson and Carroll, 2012; Fu et al., 2023; Bernardo and Keri, 2012; Hurtado et al., 2011; Lee et al., 2017a) and GREB1 mediates E2-driven proliferation and enhancer activation (Hodgkinson et al., 2018; Haines et al., 2020; Sun, Nawaz and Slingerland, 2007; Cheng, Michalski and Kommagani, 2018). The presence of significant calling card peaks at these loci strongly supports the functional integrity of the ERα fusion partner in the construct.

Closer inspection of individual loci highlights both the strengths and challenges of the HyPB-ERα calling cards method. At GATA3, the called peak extended beyond the cluster of experimental insertions, likely reflecting the merging of consecutive enriched regions where insertions exceed background. At FOXA1, elevated undirected insertions were observed, consistent with known BRD4 co-binding at this site (Nagarajan et al., 2014; Angus et al., 2021). Despite this background, ERα-directed insertions were sufficiently enriched to each statistical significance, demonstrating that genuine binding can be detected even in regions with high levels of transposase activity. At GREB1, peaks were observed in both high- and low-stringency datasets, with an additional downstream peak detected only under relaxed calling parameters. This additional binding event highlights the utility of more relaxed thresholds in capturing additional biologically relevant binding events that could represent distal enhancer activity.

Taken together, these examples demonstrate that HyPB-ERα calling cards faithfully identify ERα binding at key target genes, even in the presence of high levels of background insertions. The detection of GATA3, FOXA1 and GREB1 validates the biological relevance of our dataset and underscores the complementary value of relaxed peak calling thresholds. While high-

stringency parameters provide confidence in core binding sites, lower stringency analysis can reveal additional loci, expanding the scope of ERα binding events captured. This dual approach strengthens the case for calling cards as a functional tool to map ERα activity across both canonical and enhancer-associated regions.

3.3.6.8. Motif analysis reveals significant enrichment of ERα binding partners

Motif enrichment analysis provides the final line of evidence confirming that the HyPB-ERα fusion protein faithfully reproduces endogenous ERα binding behaviour (Figure 3.26). Across both high- and low-stringency datasets, canonical motifs of the estrogen-responsive complex were recovered, including the ERE itself and motifs of key co-regulators such as FOXA1, GATA3, RXR and MyB. The presence of these motifs within calling card peaks demonstrates that insertions occur at biologically relevant loci and that the fusion construct retains the ability to recruit the transcriptional machinery normally associated with ERα.

FOXA1 (Carroll et al., 2005; Hurtado et al., 2011; Swinstead et al., 2016) and GATA3 (Eeckhoute et al., 2007; Wilson and Giguère, 2008; Porras, Ismail and Mader, 2021) identification is crucial as these pioneer factors are required for ERα binding. RXR motifs highlight the nuclear receptor cooperativity, consistent with evidence that RXR/RAR complexes enhance ERα target gene transcription through stabilising co-factor binding (Ross-Innes et al., 2010; Shao et al., 1994; Johansson et al., 2013). MYB motifs further reinforce the biological relevance of calling cards, reflecting the ERαs role in overcoming transcriptional pausing and driving estrogen-stimulated proliferation (Drabsch et al., 2007; Mitra et al., 2012; Miao et al., 2011).

Interestingly, motif p-values were generally more significant in the low-stringency dataset, which likely reflects the larger number of peaks and increasing representation of co-regulator motifs. While high-stringency peaks provide confidence in core binding sites, relaxed thresholds expand the scope of analysis to include additional motifs, offering complementary insight into ERαs regulatory landscape.

Taken together, motif analysis confirms that HyPB-ERα calling cards identify canonical ERα binding sites and recover the motifs of its transcriptional partners. This enrichment validates the functional integrity of the fusion protein and underscores the biological relevance of the dataset, reinforcing the conclusion that calling cards can map direct ERα binding that underpins estrogen-responsive transcription.

3.3.6.9. Rationale for progression to the next chapter: Single-cell ERα Calling Cards

The initial bulk calling card dataset provided sufficient evidence that the HyPB-ERα fusion protein can replicate endogenous ERα binding patterns. Peaks overlapped extensively with ChIP-Seq, showed strong ChIP-Seq signal enrichment at their centres, were located at known ERα target genes and were enriched for motifs of ERα co-regulators. Together, these findings validate the biological relevance of the dataset and confirm that the fusion protein is functional.

At the same time, this work highlighted areas for improvement. The relatively small number of replicates limited peak recovery and statistical power, while variability in peak distribution suggested that certain chromosomes represent hotspots for ERα activity requiring further investigation. In addition, some peaks did not overlap with ChIP-Seq or ChIA-PET anchors, raising the possibility of novel binding events that warrant closer scrutiny.

For those reasons, the next stage of the project aimed to expand upon bulk calling card experiments with optimised culture conditions, followed by single-cell calling cards. Single-cell experiments offer the advantage that each cell acts as an independent replicate, greatly increasing statistical power and enabling finer resolution of ERα binding variability across the population. The following chapter details the steps taken to optimise bulk experiments and establish a single-cell pipeline.

3.3.6.10. Future directions for bulk ERα calling cards

3.3.6.10.1. Increasing replicates and refining peak calling

Future experiments should prioritise increasing the number of replicates in bulk calling card assays in line with the recommended 8-12 (Moudgil et al., 2020). A richer dataset will improve reproducibility, allow more stringent statistical thresholds and reduce sensitivity to variability. Refinement of peak calling strategies is also essential, particularly to distinguish genuine ERα binding from BRD4-associated background. Novel peak calling strategies could simplify exploring non-overlapping peaks systematically to reveal novel binding events of methodological artefacts, guiding further optimisation of analysis pipelines.

3.3.6.10.2. Investigating chromosomal ERα binding hotspots

The initial bulk calling card data revealed striking enrichment on chromosomes 17 and 20, suggesting a potential hotspot of ERα activity that is supported by the literature (Hsu et al., 2013; Lee et al., 2023a; Palaniappan et al., 2019; Wang et al., 2013). These regions should be investigated in greater detail to determine whether they represent functional hubs of ERα

regulation or reflect technical bias. Integration with additional datasets such as Hi-C or ATAC-Seq (Buenrostro et al., 2015; Corces et al., 2018; Hah et al., 2013; Fullwood et al., 2009; Rao et al., 2014; Mumbach et al., 2016) could aid in clarifying whether these hotspots correspond to accessible chromatin regions or long-range enhancer interactions.

3.3.6.10.3. Improving fusion protein design

The N-terminal fusion protein showed impaired function, likely due to the linker interference or steric hindrance caused by the linker of the HyPB fusion partner. Future work should explore high-throughput screening of alternative linkers to identify designs that preserve ERα activity (Norris and Hughes, 2018; Cao et al., 2025). Platforms such as microarrays with immobilised ERE sites could enable multiplexed testing of linker variations by releasing bound ERE sites for capture with ChIP. Such a platform would provide a systemic approach to optimise fusion protein design, as has been done previously for GFP fusion proteins (Norris and Hughes, 2018; Cao et al., 2025). The promoter choice for both fusion protein and SRT could also be revisited to balance expression levels and insertion coverage; furthermore, the fusion protein could be mutated into the endogenous locus to ensure physiological expression levels.

3.3.6.10.4. Optimising RNA yield and cell selection

Low RNA yield in N-terminal fusion experiments may reflect the sensitivity of MCF7 cells to puromycin selection. Alternative selection strategies, such as blasticidin, have shown promise in reducing stress and improving viability by other students in our lab (Data not shown). Adopting a less harsh selection method could increase RNA yield and enable a more robust assessment of both fusion constructs.

4. Engineering Long-read HyPB-ERα Calling Cards using Oxford Nanopore Technology

4.1. Introduction

ERα binding is highly dynamic, shifting in response to hormonal and environmental stimuli (Mohammed et al., 2015; Ross-Innes et al., 2012; Cho et al., 2005; Rana, Singh and Koch, 2019). For example, Progesterone receptor signalling can redirect ERα occupancy, altering transcriptional programmes (Mohammed et al., 2015), while Ross-Innes et al. identified distinct ERα binding sites associated with poor patient outcomes (Ross-Innes et al., 2012),

demonstrating the clinical relevance of cistrome variation. Environmental stressors such as hypoxia also reshape ERα binding: long-term $CoCl_2$ exposure has been shown to reprogram the ERα cistrome, highlighting how microenvironmental conditions can drive transcriptional plasticity (Cho et al., 2005; Rana, Singh and Koch, 2019).

Bulk assays such as ChIP-Seq have provided important insights into ERα biology (Carroll et al., 2006; Welboren et al., 2009; Hurtado et al., 2011; Magnani et al., 2011; Hah et al., 2011), but they average signals across populations and cannot resolve subpopulation differences (Rotem et al., 2015; Kaya-Okur et al., 2019; Park, 2009; Ludwig and Bintu, 2019). Low sample input adaptations (e.g. MOW- and DROP-ChIP-Seq) have shown promise for histone modifications, but limited applications to TFs, particularly NRs like ERα (Zhu et al., 2019; Grosselin et al., 2019). Some ChIP-Seq alternatives like Cut&RUN (Patty and Hainer, 2021; Lardo and Hainer, 2022) have improved upon previous iterations with some success, as scCut&TAG have demonstrated application to single-cell TF binding analysis (Bartosovic, Kabbe and Castelo-Branco, 2021). However, these methods are limited to probing ERα genomic occupancy only, whereas the calling card method simultaneously captures the transcriptome of cells, granting additional insights into the consequences of the ERα binding profile (Moudgil et al., 2020). There is currently no robust single-cell method to simultaneously dissect ERα binding heterogeneity and the transcriptomic landscape it gives rise to.

However, a practical consideration when using our HyPB-ERα fusion protein is that the current method introduces an additional ERα species into cells already expressing the endogenous ERα. The overexpression of our fusion could therefore alter receptor stoichiometry, influence co-factor availability or shift the balance between ERα dimers, which would all affect the native binding profile of ERα. Previous calling card studies have shown that reporters can function in the presence of endogenous TFs, but it remains important to recognise that the fusion may not fully recapitulate endogenous ERα binding, particularly as this reporter has not been validated previously. This could be mitigated in the future through using low-expression promoters or, ideally, knocking the HyPB domain in frame with the endogenous ERα locus, but this limitation should still be acknowledged when beginning this work. Despite these considerations, the CC system still provides a robust and highly informative readout of ERα binding, and its compatibility with single-cell workflows offers a level of resolution that is not achievable with other methods. Furthermore, by overexpressing the HyPB CC fusion proteins, researchers can first confirm their fusion is functional and not subject to steric hinderance before undertaking a more costly knock-in to the endogenous locus. Together, single-cell HyPB-ERα calling cards a particularly valuable

In the previous chapter, we validated the ERα calling card in bulk, demonstrating overlap with ChIP-Seq, enrichment of ERα motifs, and localisation to known target genes. To meet the overall aim of this thesis, the next step is to apply the validated HyPB-ERα fusion protein to a single-cell workflow capable of resolving subpopulation-specific ERα binding. This chapter, therefore, focuses on two objectives: First, further optimising the use of HyPB-ERα fusion proteins to ensure robust performance in single-cell contexts; and second, the evaluation of their integration with the 10X platform to enable simultaneous profiling of ERα binding and transcriptomes at single-cell resolution

A further objective of this chapter is to evaluate our novel application of long-read sequencing to single-cell calling cards. While previous implementations of calling cards have relied on the circularisation of SRT transcripts for short-read sequencing, our pipeline leverages long-read technology to capture full insertion sites and associated transcript information in a single read. This approach not only resolves technical challenges encountered during library preparation but also provides a unique opportunity to increase the accessibility of single-cell calling cards, as some long-read sequencers are cheap and portable (Yasir et al., 2022; Zheng et al., 2023; Smits and Faulkner, 2023). By integrating long-read sequencing into the single-cell workflow, we aim to demonstrate how this adaptation enhances the resolution and interpretive power of ERα calling cards compared to other methods

This chapter aims to adapt the validated HyPB-ERα fusion proteins to a single-cell workflow, integrating long-read sequencing to simultaneously resolve ERα binding heterogeneity and transcriptomic consequences at single-cell resolution.

Objectives to achieve this aim are:

- Optimise HyPB-ERα fusion protein performance in single-cell contexts to ensure robust insertion capture.
- Integrate the ERα calling card system with the 10X Genomics platform to enable simultaneous profiling of ER binding and transcriptomes.
- Develop and apply a long-read sequencing pipeline to capture full insertion sites and associated transcript information in single reads.
- Evaluate the added value of long-read sequencing compared to short-read approaches, particularly in resolving technical challenges and enhancing accessibility.

- Establish the foundation for analysing ERα binding heterogeneity across subpopulations, providing tools to investigate how ERα variation contributes to endocrine therapy resistance.

By addressing these objectives, we establish the foundation required to analyse ERα binding heterogeneity across subpopulations. This chapter, therefore, represents a critical transition from bulk validation to single-cell long-read application of HyPB-ERα calling cards, providing the tools necessary to evaluate how ERα binding variation contributes to endocrine therapy resistance in heterogeneous tumour microenvironments.

4.2. Results

4.2.1. Addition of a Kozak sequence by site-directed mutagenesis improves full-length fusion protein expression

In Chapter 3, we observed that the C-terminal ERα fusion protein consistently showed lower full-length protein expression on Western blots compared to N-terminal constructs, indicated by an equally intense, smaller protein band ~110 kDa (Figure 3.3). This observation is accompanied by a lower number of peaks than expected (Moudgil et al., 2020), and in the case of one sample, insufficient RNA was recovered for sequencing. While these findings do not establish a direct causal link between protein expression and downstream assay performance, they highlight the need to optimise the C-terminal construct to improve its reliability for bulk and single-cell calling card experiments.

We hypothesised that the absence of a Kozak sequence in the expression vector permits translation initiation at an alternative start site (Kozak, 2002; Benitez-Cantos et al., 2020; Andreev et al., 2022), producing a truncated protein product. These truncated forms will reduce the relative abundance of the full-length fusion protein, and their presence complicates the interpretation of expression data. Therefore, increasing the proportion of full-length fusion protein is an important step towards improving the efficiency of the fusion construct.

To address this, primers were designed to introduce a Kozak sequence upstream of the C-terminal fusion coding sequence. Engineering a Kozak sequence into the N-terminal construct was not feasible without compromising the integrity of the ERα protein composition; a silent mutation in the second ERα codon is not possible. To preserve the integrity of the ERα sequence, only the C-terminal construct was mutated. The mutation was successfully

incorporated and validated by matching Sanger sequencing results to the expected snapgene template (Figure 4.1).

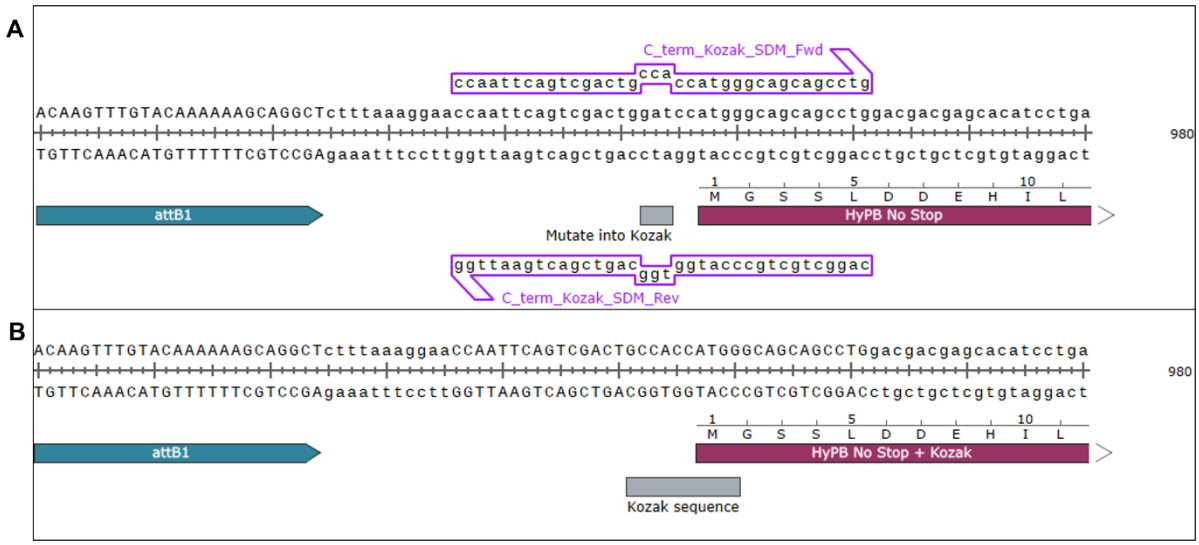


Figure 4.1: Site-directed mutagenesis successfully mutated three base pairs preceding the HyPB-ERα fusion construct to create a Kozak Sequence.
Primers were designed to mutate a three-base pair stretch of DNA preceding the C-terminal HyPB-ERα fusion protein **(A)**. Site-directed mutagenesis was carried out using these primers, which resulted in the successful engineering of a Kozak Sequence preceding the fusion protein gene **(B)**. The successful integration was confirmed with Sanger sequencing that matched the predicted sequence.

Following sequencing validation, the Kozak-modified plasmid was transfected into MCF7 cells for Western Blot analysis. Blotting revealed a clear increase in the intensity of the full-length band compared to the original construct, accompanied by a marked reduction in the smaller band intensity (Figure 4.2). In fact, the smaller band was almost completely absent in the Kozak-modified sample. Although this experiment was performed once (n=1), the disappearance of the smaller band is consistent with improved translation fidelity, and the stronger full-length band suggests enhanced expression of the intended fusion protein.

These results demonstrate that the addition of a Kozak sequence upstream of the C-terminal HyPB-ERα fusion coding sequence increased full-length protein expression and reduced truncated products. This optimisation step strengthens the efficiency of the HyPB-ERα protein by increasing the proportion of fully expressed fusions. The improved construct will provide a more robust basis for downstream integration with bulk sequencing, single-cell 10X analysis and long-read calling card pipelines.

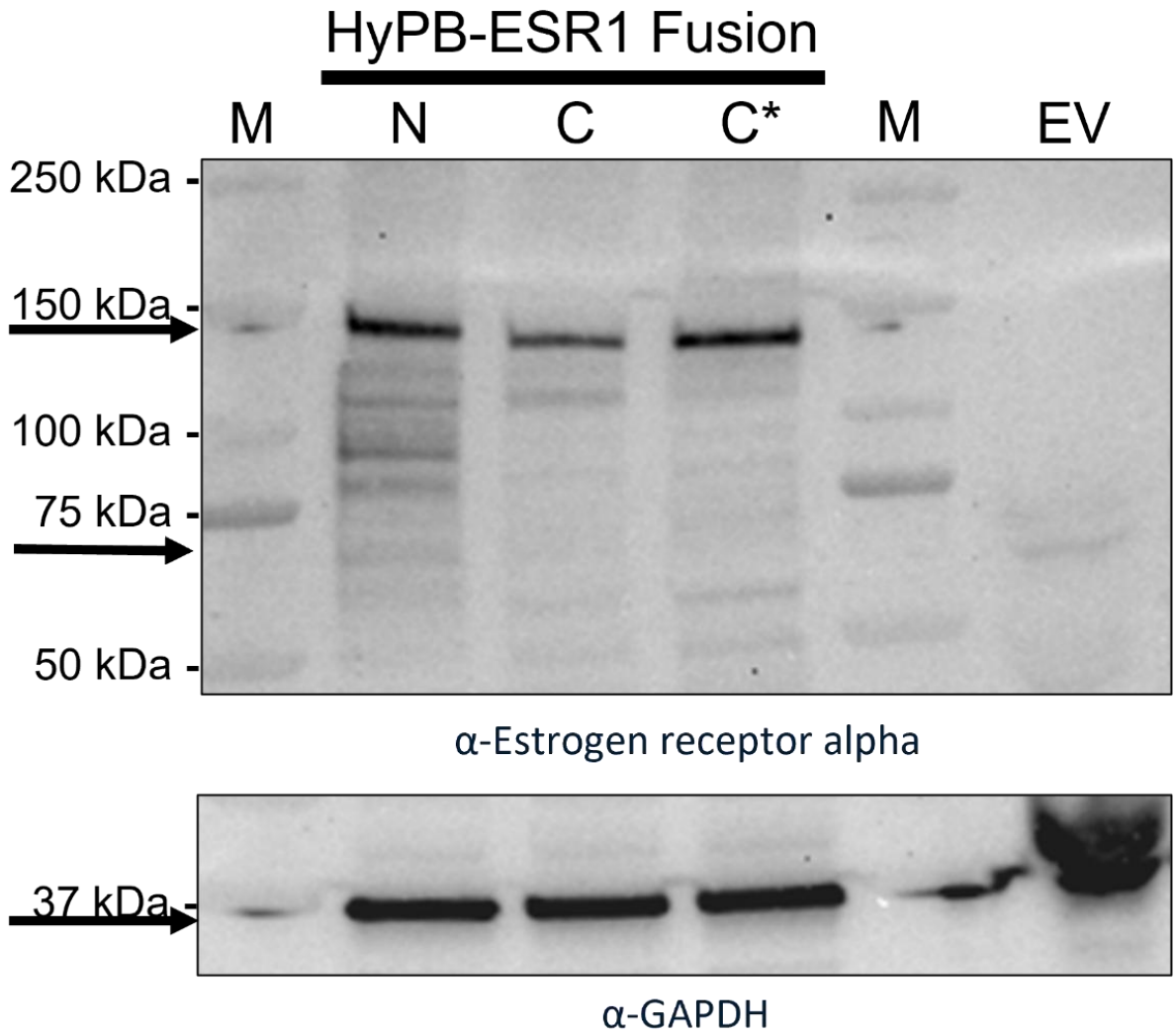


Figure 4.2: The addition of the Kozak sequence visibly increases the intensity of the full-length fusion protein band and almost eliminates the second smaller band in C-terminal fusion protein-transfected lysates.
Following site-directed mutagenesis of the Kozak sequence into the C-terminal fusion protein plasmid, the new mutant was transfected into MCF7 cells and cell lysates were collected after 72 hours for Western blotting. Lysates from previous blots were run alongside the lysate from the new mutant transfection. **M** = Marker, **N** = N-terminal fusion protein transfection, **C** = C-terminal fusion protein transfection, **C*** = C-terminal fusion protein with Kozak sequence transfection, **EV** = Empty vector transfection. The molecular weight of the full fusion protein is 136 kDa, as indicated by the first arrow; each transfection shows a band at this molecular weight, whereas the empty vector does not. The C* condition shows a reduction of these smaller bands and increased intensity of the full-length band. The added Kozak sequence could reduce the presence of smaller bands by aiding the identification of the correct translation initiation site. The second arrow indicates the endogenous ERα band at 66 kDa, while expression can be seen in the EV and N conditions, but it is not apparent in either the C condition. The membrane was re-blotted with an α-GAPDH antibody as a loading control for this blot (band found at 37 kDa), demonstrating equal protein loading in each lane.

4.2.2. Seeding density and puromycin concentration optimisation to improve RNA yield

The objective of this optimisation was to improve the viability of co-transfected MCF7 cells, thereby enabling recovery of sufficient RNA for downstream calling card assays. Viability improvement is critical for further use of the N-terminal ERα-HyPB fusion protein to generate sufficient RNA for sequencing.

4.2.2.1. Seeding density optimisation in 6-well plates

The effect of cell seeding density on transfection efficiency was tested using the BrokenHeart SRT, which produces dsRed fluorescence when successfully co-transfected alongside a functional HyPB protein (Cammack et al., 2020). The expression of red fluorescence enabled quantification of the co-transfection efficiency as a percentage of dsRed-positive cells at 48 hours post-transfection across a range of seeding densities (Figure 4.3).

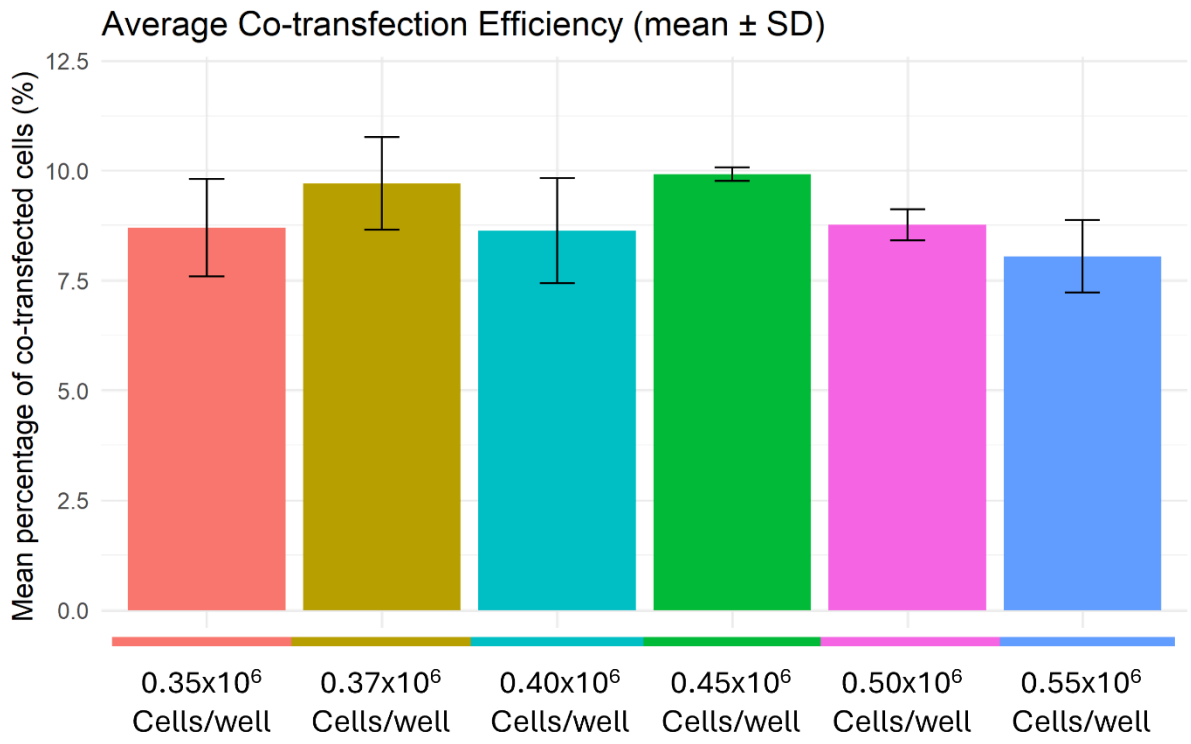


Figure 4.3: The highest co-transfection efficiency of calling card plasmids is achieved when seeding MCF7 cells at 0.45 x10^6 cells per well in a 6-well.
Co-transfection of WT-HyPB and BrokenHeart transposon plasmids in cells seeded at different densities revealed that seeding at 0.45 x10^6 cells per well in a 6-well plate achieved the highest efficiency, with 9.7% of cells showing dsRed expression after 48 hours.

Only small variations in co-transfection efficiency were observed across densities. The highest average efficiency obtained was at 0.45 x10^6 cells per well, with 9.7% of cells seeded at this density expressing dsRed. Increasing seeding density to 0.5 x10^6 cells per well did not result in a significant change. Based on these results, 0.45 x10^6 cells per well was adopted as the working density, balancing cell number and co-transfection efficiency.

.

4.2.2.2. Puromycin concentration optimisation

Initial bulk calling card experiments used 2 µg/mL, a concentration based on the kill curve generated on untransfected MCF7 cells (Appendix Figure 8.4). However, the application of this treatment within 48 hours of transfection, as per previously published methods (Moudgil et al., 2020) led to premature cell death in MCF7 cells, confounding RNA recovery. To reduce this issue, the protocol was altered to allow cells to grow to 70-80% confluence before selection with puromycin. This change improved the viability of C-terminal fusion transfections to generate sufficient material for sequencing, but it was not enough to improve N-terminal transfection viability.

To refine selection conditions specifically for transfected MCF7 cells, puromycin concentrations below 2 µg/mL, guided by the previous kill curve data. This data shows the largest difference in killing was between 1 and 2 µg/mL. Therefore, concentrations of 1.25, 1.5 and 1.75 µg/mL were applied to cells transfected with the N-terminal fusion plasmid and puromycin SRT

4.2.2.3. Influence on passage number

Differences in viability were also noted between high and low-passage MCF7 cells. Pilot bulk experiments used cells at passage 15 (P#15), which showed reduced viability compared to passage 27 (P#27) cells used in earlier optimisation. To test this difference systematically, puromycin concentration experiments were repeated in parallel in P#15 and P#31 MCF7 cells.

4.2.2.4. RNA yield increases with higher-passage number and lower puromycin concentration

The cells were cultured in the same way as N-terminal bulk calling card assays, up to RNA harvest, and both cell counts and RNA concentrations were recorded (Figure 4.4)

MCF7 Passage and Puromycin concentration used	**Concentration (ng/µl)**	**Purity A260/280**	**Purity A260/230**	**Total cells harvested (x10⁴ cells)**
Passage 15 1.75 µg/ml Puromycin	18.0	1.86	0.17	2.0
Passage 15 1.5 µg/ml Puromycin	84.0	2.02	0.94	17.5
Passage 15 1.25 µg/ml Puromycin	48.2	1.9	1.52	8.5
Passage 31 1.75 µg/ml Puromycin	257.8	2.07	1.17	59.5
Passage 31 1.5 µg/ml Puromycin	456.2	2.05	2	110.0
Passage 31 1.25 µg/ml Puromycin	490.6	2.04	2.01	158.0

Figure 4.4: Both cell passage and puromycin concentration greatly affect RNA yield and total cells harvested.

The treatment of passage 15 and 31 MCF7 cells with a range of puromycin concentrations reveals that reducing the concentration below 1.75 µg/ml can increase the total number of cells and RNA concentrations harvested from them, ~2-12 fold depending on condition. As puromycin concentration decreased, cell cultures appeared healthier and had fewer dead cells accumulating.

Across conditions, higher-passage cells (P#31) yield more cells and higher RNA concentrations than low-passage cells (P#15), replicating the findings of previous bulk calling cards. Depending on the condition, RNA yield was approximately 2 to 12-fold higher in P#31 cells.

The condition that was most similar to bulk ERα calling cards conditions (P#15 cells treated with 1.75 µg/mL) again failed to produce sufficient RNA for downstream analysis. A minimum of 2 µg of RNA is required for bulk calling card preparation, corresponding to 40 ng/µL when eluted in 50 µL of water. All conditions were treated with 1.5 µL/mL puromycin or less to achieve this threshold. Lower concentrations were also associated with healthier cultures and reduced cell death, as reflected by the number of cells harvested.

Taken together with the kill curve data, 0.5 µg/mL was adopted as the working concentration, as this was the lowest dose to achieve complete cell death in controls after 7 days and increases the viability of co-transfected cells compared to bulk calling cards in Chapter 3. High passage aliquots of MCF7 cells were avoided and the passage number of cells was kept below 20 to avoid clonal selection driven by successive passage.

4.2.2.5. Evaluation of the optimised protocol

To evaluate the optimised protocol, the bulk HyPB-ERα calling card workflow was repeated with both N-and C-terminal constructs. RNA was harvested in the same way to allow comparison of the C-terminal fusion constructs with and without Kozak sequence modification. While the C-terminal fusion transfection grew more robustly and yielded sufficient RNA, the N-terminal fusion transfections remained less viable and again failed to generate sufficient RNA for library preparation (Appendix Table 8.9).

While the N-terminal fusion insertions could not be recovered, we do demonstrate that method optimisations increase C-terminal transfection viability and RNA yield. As a result, the additional bulk RNA was stored for future use, and we proceeded to develop single-cell ERα calling cards using the C-terminal HyPB-ERα construct.

4.2.3. Long-read ERα calling cards overlap with ChIA-PET anchors more than ChIP-Seq peaks

4.2.3.1. The development of long-read calling cards

Despite the successful optimisation of bulk ERα calling card protocols, attempts to adapt the single-cell calling card workflow for short-read sequencing proved challenging. The genomics team trialled the circularisation of calling card transcripts using an unpublished, updated method from the Mitra Lab, but the efficiency of this step and subsequent reactions was too

low to recover sufficient insertions for analysis. As the eluate aliquot designated for calling card processing was expended in these trials, we instead leveraged aliquots from RNA already prepared for 10X transcriptome profiling. From these samples, we developed a novel long-read sequencing strategy in partnership with the genomics core, amplifying calling card insertions directly from the PCR-amplified RNA fraction. This approach provided a practical solution to the limitations of short-read circularisation and opened the opportunity to evaluate long-read calling cards as a complementary method for profiling ERα binding.

To understand how long-read calling cards perform relative to established ERα binding datasets, we compared insertions processed through two different pipelines: trimmed and untrimmed. Trimmed data follows the conventional approach, where reads are aligned to the TTAA tetranucleotide motif at HyPB terminal repeats and the adaptor sequences are removed to facilitate genomic alignment. While this is the standard method, in our dataset it proved overly stringent: most reads were discarded, preventing effective peak calling. To address this, the main analysis of this chapter will focus on the untrimmed dataset, in which adaptor sequences are retained during alignment. The Epi2Me pipeline can accommodate these non-genomic sequences, allowing substantially more reads to pass through the calling-card workflow, improving the recovery of usable signal. A full comparison of the trimmed versus untrimmed datasets can be found in the appendix (Appendix 8.2.4).

4.2.3.2. The long-read calling card pipeline approaches the target of captured cells and is comparable to the bulk ERα calling card in the number of peaks called

The Epi2Me single-cell report shows that from the untrimmed dataset, we recovered 4,326 cells, much closer to the 6,000 cells expected from a 10,000-cell target (Zheng et al., 2017). Median values remained low (16 UMIs and 14 loci per cell), but the increase in recovered cells provided sufficient material to continue developing the pipeline. From this dataset, 101,099 insertions were identified, and after applying a UMI filter that removes insertions with only a single UMI (reads coming from only one cell), the number of insertions was reduced to 4,259. Peak calling with low-stringency MACCs parameters, 502 peaks were produced from the unfiltered data and 10 peaks from the filtered, numbers that are comparable with bulk calling card experiments.

4.2.3.3. Long-read ERα calling card peaks overlap with ChIA-PET anchors more than ChIP-Seq, reflecting overlap patterns of

Overlap analysis revealed distinct patterns, with long-read calling card peak overlap with ChIP-Seq peaks being lower than that with ChIA-PET anchors (Figure 4.5**Error! Reference source not found.**), reflecting the pattern observed in bulk calling cards.

When compared to ChIP-Seq peaks (Figure 4.5**Error! Reference source not found.**), 9.2% of long-read calling card peaks overlapped with known ERα peaks in the GEO dataset, while 22.2% of long-read peaks overlapped with ENCODE peaks. While these values are modest, they demonstrate that long-read calling card peaks do recover loci coinciding with established ERα binding sites. Hypergeometric testing confirms that untrimmed long-read calling card peaks are significantly overlapping with ChIP-Seq peaks ($p = < 1x10^{-7}$).

In contrast to the modest overlap of calling card peaks with ChIP-Seq, the number of long-read peaks overlapping with ChIA-PET anchors was higher for long-read data as 63.1% of long-read peaks overlapped with ChIA-PET anchors (Figure 4.5**Error! Reference source not found.**). Hypergeometric testing confirms that all untrimmed long-read calling card conditions showed statistically significant enrichment with ChIA-PET anchors ($p < 0.005$).(Appendix Table 8.10). This degree of concordance highlights the consistency of long-read ERα calling cards with chromatin interaction data.

These results show that long-read ERα calling cards recover a set of peaks that modestly overlap with ChIP-Seq but show much stronger concordance with ChIA-PET anchors, mirroring the pattern observed in bulk calling card experiments. The high overlap with ChIA-PET, together with significant enrichment by hypergeometric testing, indicates that long-read calling cards capture ERα binding events that could be involved in long-range chromatin interactions rather than simply recapitulating all ChIP-Seq peaks.

Taken together, these findings support long-read sequencing as a viable calling card strategy that directly recovers insertion sites and reproduces the characteristic overlap profile of bulk ERα calling cards. In the remainder of this chapter, analyses therefore focus on the untrimmed long-read dataset, with alternative processing strategies, including UMI filtering and trimming, presented in the Appendix for comparison (Appendix 8.2.4).

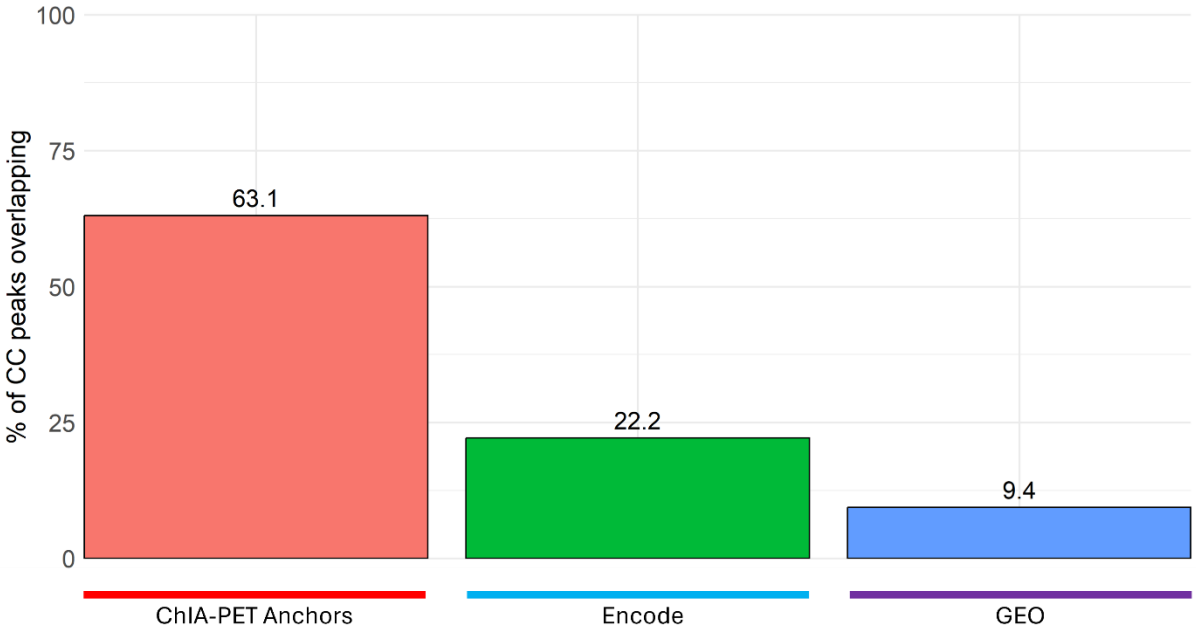


Figure 4.5: Long-read calling card peaks show more than double the amount of overlap with ChIA-PET anchors compared to overlap with ChIP-Seq peaks.
Overlap with ChIP-Seq datasets remains limited with only 9.4% of long-read calling card peaks overlapping with GEO peaks, and 22.2% overlapping with ENCODE peaks. However, the number of long-read peaks overlapping with ChIA-PET anchors was much higher, with 63.1% of peaks overlapping with either anchor in a Erα directed 3D interaction. Untrimmed calling card overlap with both ChIP-Seq peaks and ChIA-PET anchors was statistically significant ($p < 0.005$).

4.2.4. Signal analysis at long-read calling card insertions confirms the overlap patterns with ChIP-Seq and ChIA-PET

As in bulk calling card experiments, overlap analysis alone does not fully capture whether insertions coincide with genuine ERα binding sites; we therefore examined ChIP-Seq signal intensity at long-read insertion loci. This approach allowed us to determine whether regions overlapping ChIP-Seq or ChIA-PET datasets also show enrichment for binding signal, providing an additional layer of validation.

For the GEO dataset, which showed the lowest overlap, signal intensity at insertion sites was modest. The mean $\log_2$ fold change at TTAA motifs did not exceed ~1.5, and heatmaps showed little visible enrichment across loci (Figure 4.6 A and B); however, enrichment analysis shows that the long-read calling card peaks do exhibit statistically significant enrichment (Appendix Figure 8.15)(Wilcoxon $p < 0.05$), confirming that long-read calling cards show ERα enrichment in poorly overlapping datasets. These results indicate that GEO overlaps are not representative of strong ERα binding, consistent with the lower overlap percentages reported above.

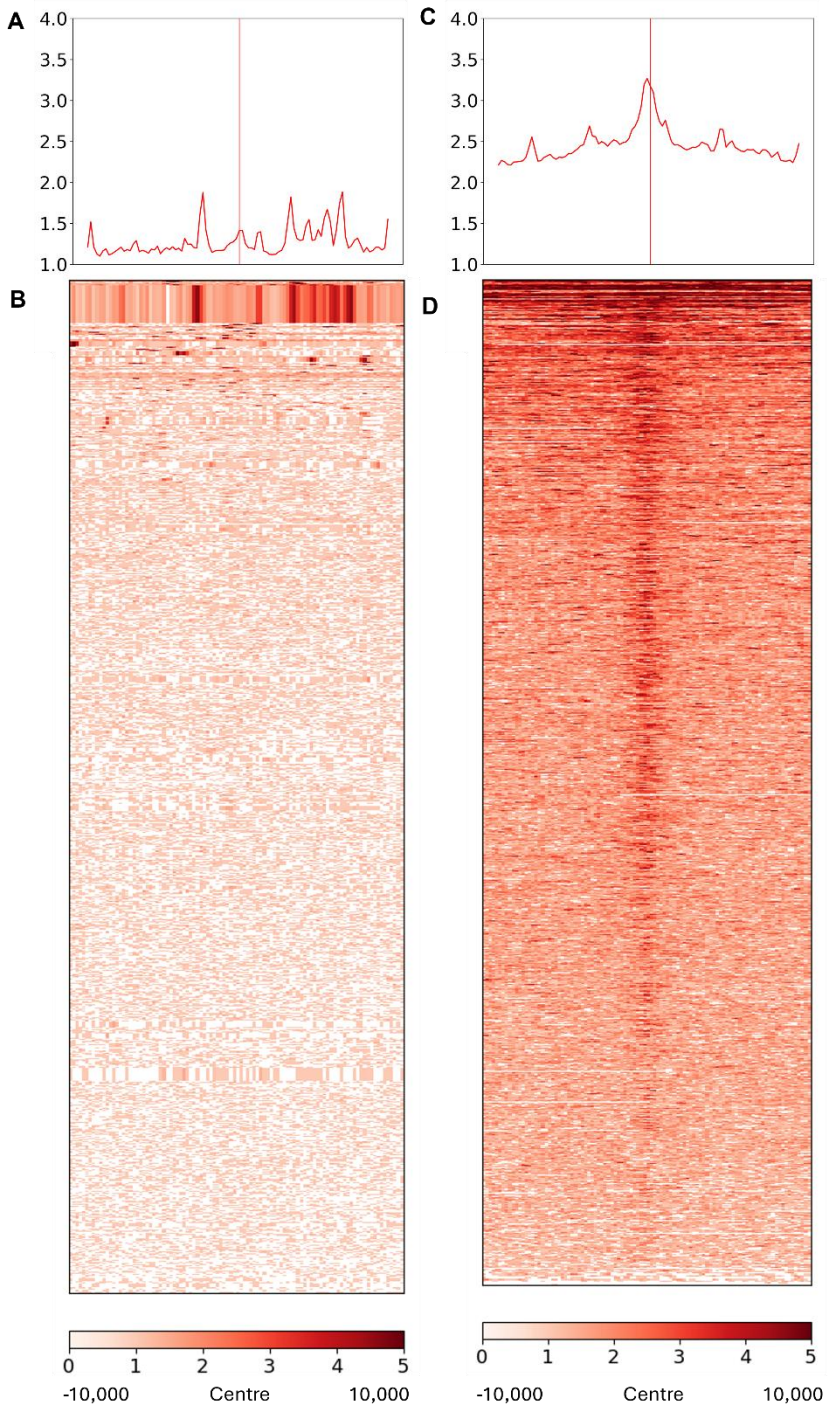


Figure 4.6: ENCFF063JMY ChIP-Seq signal is enriched at the centre of full long-read Calling Card insertions, but GSE109820 ChIP-Seq signal is not.
ChIP-Seq signal was plotted ±10 kb around long-read Calling Card peak centres for non-filtered and UMI-filtered insertions. **(A, B)** GSE109820 ChIP-Seq shows low enrichment at Calling Card peak centres, with average $\log_2$ fold-change values below 2. Signal intensity plots (A) and heatmaps (C) indicate that while some individual regions show elevated signal, there is no consistent enrichment at insertion centres, particularly after UMI filtering. **(C, D)** ENCFF063JMY ChIP-Seq shows clear enrichment at Calling Card peak centres, with average $\log_2$ fold-change values of 3.25–4 and peak intensities reaching $\log_2$ fold-change ~5 in the strongest regions. Signal intensity plots (B) and heatmaps (D) demonstrate robust central enrichment across both non-filtered and UMI-filtered insertions.

In contrast, the ENCODE dataset showed clear enrichment at insertion sites. Long-read insertions had a mean log_2 fold change of ~3.25 at TTAA motifs, with a consistent signal across loci Figure 4.6 C and D). Signal intensity across flanking regions also remained high, mirroring the enrichment seen in bulk calling card experiments. Encode signal enrichment was observed consistently across long-read calling card peaks compared to shuffled controls (Wilcoxon $p < 0.0001$); demonstrating consistency with both the gold standard for ERα binding analysis and previous bulk ERα calling cards.

Like bulk calling cards, long-read calling cards were enriched for the signal from ERα ChIP-Seq, ChIA-PET anchors and BRD4 ChIP-Seq are all significantly enriched in long-read calling card peaks. Violin plot analyses revealed that long-read peaks were significantly enriched for ERα ChIP-Seq signal from ENCODE and GEO. In contrast, BRD4 ChIP-Seq datasets demonstrated robust enrichment across both Liu et al. (2022) and Zheng et al. (2022), with stronger concordance observed for the Liu dataset. Importantly, overlap with ERα ChIA-PET anchors was consistently higher than with ChIP-Seq peaks, with long-read peaks reaching up to 63.1% for ChIA-PET anchors, almost double the highest ChIP-Seq overlap of 33.3%. Hypergeometric testing confirmed that these enrichments were highly significant ($p < 1 \times 10^{-300}$ for most comparisons), underscoring that long-read calling cards preferentially capture chromatin interaction anchors and BRD4 binding sites, while showing more modest overlap with ERα ChIP-Seq peaks.

Together, these analyses show that long-read calling cards reproduce some patterns observed in bulk calling card experiments: Strong overlap with ENCODE and ChIA-PET datasets. This consistency across overlap and signal analysis highlights the benefits of our long-read sequencing approach: Even with limited recovery, it enables direct validation of insertion sites against genome-wide binding signal, strengthening confidence that recovered loci represent genuine ERα binding activity.

4.2.5. Long-read peaks are located at known ERα binding sites that overlap closely with ChIP-Seq peaks

To further validate the long-read calling card dataset, we examined individual peaks and the genes adjacent to their insertion sites. Showing insertions that occur near known ERα target genes provides confidence that these represent genuine binding events. Peaks were visualised in IGV alongside qbed files and ChIP-Seq peaks, and representative examples are shown in Figure 4.7.

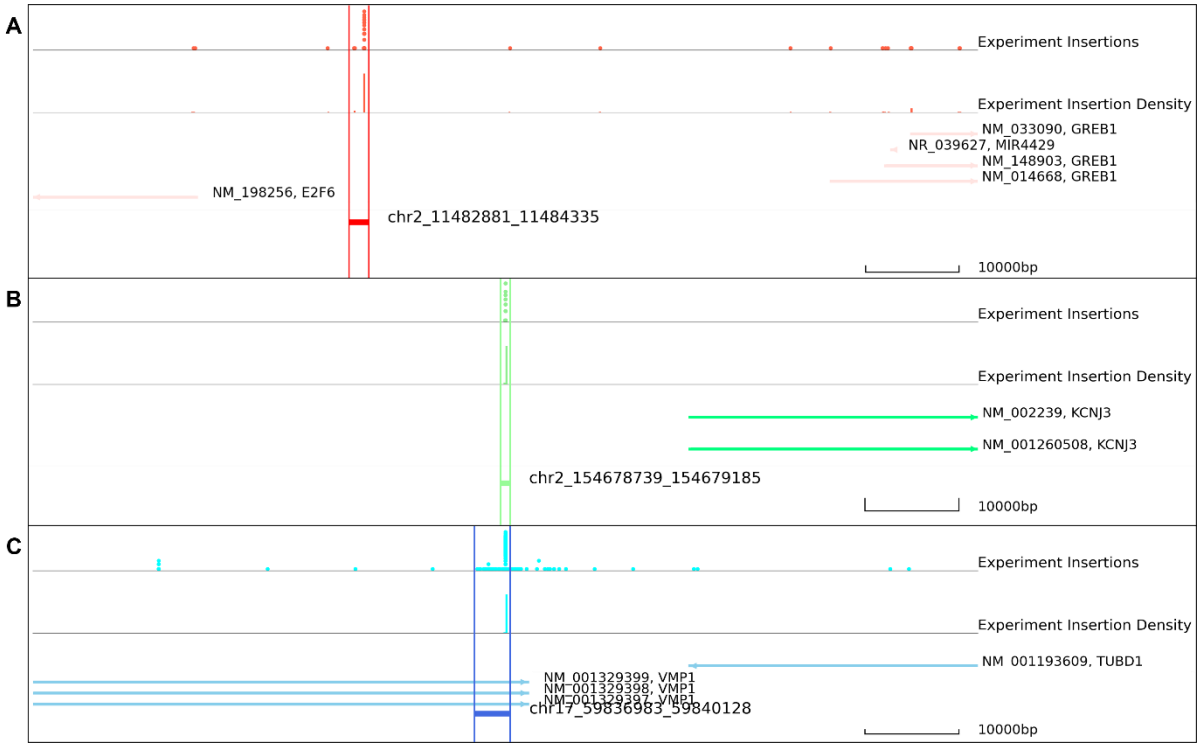


Figure 4.7: Long-read calling card peaks are located at genomic loci of putative ER targets GREB1, KCNJ3 and VMP1.
Peaks found adjacent to GREB1 **(A)**, KCNJ3 **(B)**, and VMP1 **(C)**. All these peaks overlap with ERα ChIP-Seq peaks. Identification of peaks adjacent to bona fide ERα binding locations indicates that the fusion protein can redirect wild-type HyPB insertions to genuine ERα binding sites. Each of these peaks appears genuine, as there are several insertions stacked on top of each other, indicating the amplification of several calling card insertions.

Several well-characterised ERα target genes were recovered. For example, a peak upstream of the GREB1 gene was identified in the long-read dataset, close to the E2F6 gene and overlapping with a narrow ChIP-Seq peak (Figure 4.7 A). Interestingly, the exact locus differs from the bulk ERα calling card peaks (Figure 3.25); given the overlap with ChIP-Seq and proximity to GREB1, this data strongly supports this as a genuine insertion.

Similarly, a peak adjacent to KCNJ3 mapped to an untranscribed genomic locus that overlapped with a ChIP-Seq peak (Figure 4.7 B). KCNJ3 encodes subunit 1 of the G-protein-activated inwardly rectifying K+ channel (GIRK1), and high levels of expression have been linked to poor patient prognosis, especially in ERα-positive breast cancer (Rezania et al., 2016; Kammerer et al., 2016), making this a biologically relevant peak.

A broader peak was also detected within the VMP1 gene, overlapping ChIP-Seq peaks in both datasets and located near a bulk calling card peak (Figure 4.7 C). VMP1 has been associated with poor prognosis in HER2+ and triple-negative breast cancers (Amirfallah et al., 2019; Inaki

et al., 2011; Persson et al., 2017), while not directly associated with ERα, insertion within cancer-associated genes does support the relevance of this peak.

We also investigated peaks that initially appeared less convincing in visualisation plots but proved genuine on closer inspection (Figure 4.8). For example, the TFF1 peak appeared equally flat as several non-significant insertions upstream on the experimental insertion track, and yet the density track revealed 118 insertions across different cell barcodes, contributing to the peak (Figure 4.8 A). Each of these insertions was only detected once across the 118 cells, giving a 'score' of one per barcode; therefore, because the height of peaks is determined by the score, the samples on the experiment insertions track appear flat. The accumulation of insertions enabled the peak to be called significant over the expected number of insertions within that region. TFF1, similarly to GREB1, is a well-defined transcriptional target gene of the ERα gene expression programme (Sun et al., 2005; Laganière et al., 2005; Baron et al., 2007), which in combination with ChIP-Seq peak overlap, indicates this is a genuine ERα-directed peak.

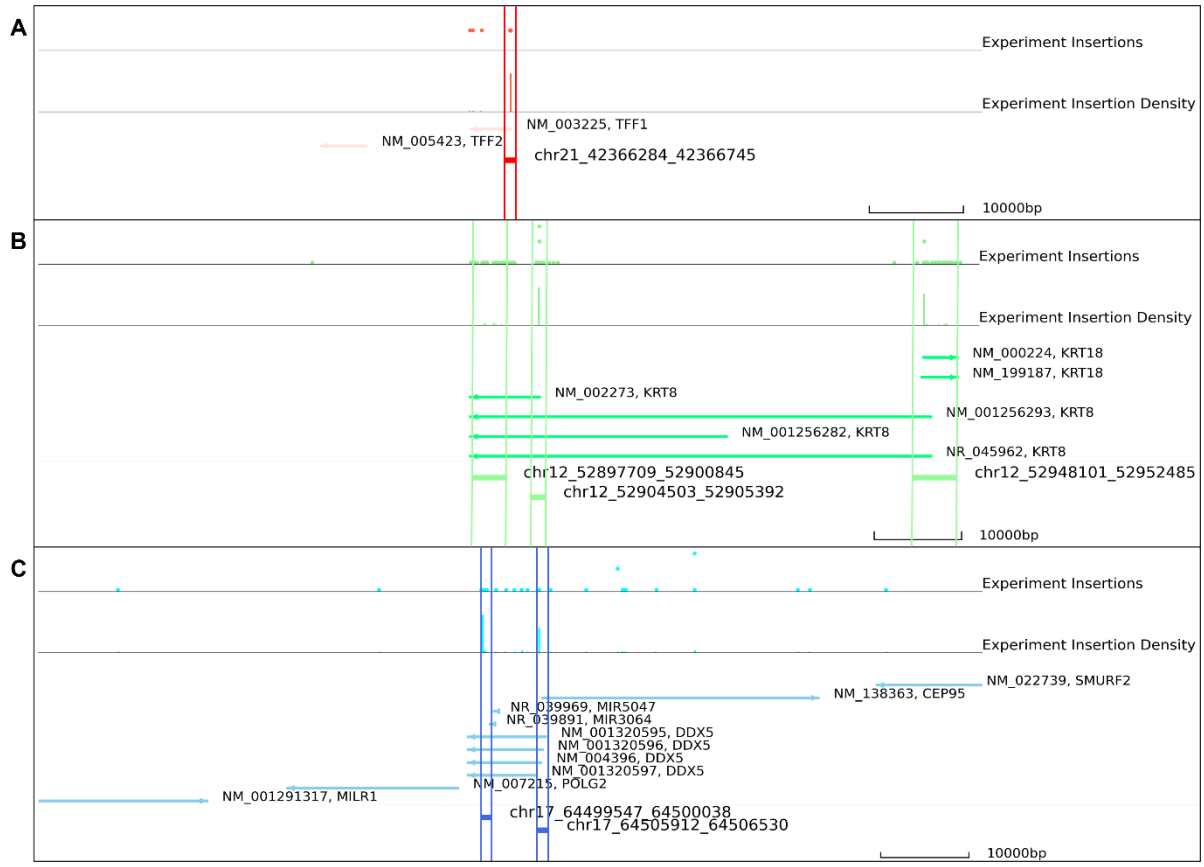


Figure 4.8: Some long-read calling card peaks are located at ER binding sites and show a limited number of insertions in individual cells, in favour of many peaks from many different cells.

Peaks found adjacent to TFF1 **(A)**, KRT8/18 **(B)**, and DDX5 **(C)**. All these peaks overlap with ER ChIP-Seq peaks. Identification of peaks adjacent to bona fide ER binding locations indicates that the fusion protein can redirect wild-type HyPB insertions to genuine ER binding sites. However, these peaks do not contain calling card insertions that stack on top of each other, which indicates that these peaks are generated from many insertions with low 'score' values. The peaks will be called significant due to the number of insertions across different cells being higher than the expected value calculated for that region of the genome by chance.

Across the KRT8/18 locus, three long-read ERα calling card peaks were identified (Figure 4.8 B). The first peak contained 97 insertions, each with a score of one, resulting in a flat experimental track relative to the other peaks. The other two peaks contained over 500 insertions each with varying scores, producing much higher-density and experiment insertion signals than the first peak. Although KRT8/18 aren't classical ERα targets, they are expressed in ERα-positive cells and can be modulated by ERα activity (Meng et al., 2009; Wilson and Dering, 2004; Nguyen et al., 2018; Ha et al., 2011; Iyer et al., 2013). The overlap with ChIP-Seq peaks and many insertions support these long-read peaks as genuine ERα binding events.

Finally, two peaks were identified in the DDX5 gene (Figure 4.8 C). These peaks appeared lower in the experimental track than in the surrounding, non-significant insertion regions, but inspection showed 20-30 insertions across different cells contributing to each peak. Both overlapped with ChIP-Seq peaks in both datasets. DDX5 is a known co-regulator of ERα, and its involvement in ERα transcriptional programmes makes these strong candidates for genuine insertions (Fuller-Pace and Ali, 2008; Wortham et al., 2009; Nicol and Fuller-Pace, 2010).

These examples show that long-read ERα calling card recover peaks at established ERα target genes and co-regulators, with overlap to ChIP-Seq peaks and support from multiple independent cell barcodes. While some peaks appear visually flat on experimental insertion tracks due to the way that track height is calculated, closer inspection revealed that substantial numbers of single insertions across many cells contribute to them. These results demonstrate that long-read calling cards can map ERα binding information with confidence, complementing the overlap and signal analyses presented above.

4.3. Discussion

4.3.1. Advancing ERα Calling Card Technology through long-read sequencing

The work presented in this chapter demonstrates how systematic optimisation and strategic redesign can overcome key limitations in adapting calling card technology for ERα in breast cancer cells. By combining protein engineering, cell culture optimisation, and sequencing innovation, we established a framework that both alleviates the constraints of existing methods and highlights new opportunities for single-cell ERα calling card analysis.

A central outcome was the development of a novel long-read sequencing pipeline that resolved the inefficiency of circularisation and ligation steps required for short-read calling cards. By amplifying calling card transcripts directly from RNA fractions and sequencing with long-read technology, we demonstrated that genuine HyPB-ERα insertions can be recovered when short-read methods fail. This establishes proof-of-principle that long-read sequencing can serve as a viable alternative for calling card experiments, particularly in contexts where circularisation is inefficient or material is limited.

The addition of a Kozak sequence to the C-terminal fusion construct markedly improved experimental outcomes, leading to higher full-length protein expression, greater cell viability and increased RNA yield. This modification reduced the presence of truncated protein products and ensured more robust translation of the fusion protein. As a result, cells transfected with the C-terminal construct were healthier and produced more RNA than previous bulk calling card assays.

These findings highlight two advances: first, that the construct optimisation through Kozak sequence addition improved protein expression and cell viability, and second, that long-read sequencing provides a practical solution to the inefficiencies of short-read circularisation (Moudgil et al., 2020), enabling recovery of genuine insertion events. These insights clarify the constraints of ERα calling card design and sequencing, while also pointing towards future directions such as expanding long-read approaches to achieve single-cell resolution in patient-derived samples.

4.3.2. Enhancing fusion protein expression improves viability and RNA yield

The addition of a Kozak sequence to the C-terminal HyPB-ERα fusion construct markedly improved experimental outcomes, increasing full-length protein expression, reducing truncated products (Figure 4.2), and leading to greater cell viability and RNA yield (Appendix Table 8.9). These results highlight the importance of translation initiation efficiency for maintaining construct stability and ensuring sufficient material for downstream calling card analysis. By contrast, N-terminal transfection consistently produced poor viability and low RNA recovery, indicating that these cells were more sensitive to expression of this construct, and were unable to sustain the viability despite optimisation of co-transfection and selection conditions.

Although the Kozak sequence could not be introduced into the N-terminal fusion without altering the encoded threonine residue, future work could explore codon substitutions that preserve protein function while enabling Kozak motif addition. For example, replacement with a small amino acid like alanine or glycine may be tolerated, but the functional importance of the endogenous threonine must be carefully considered. Such redesigns would test whether improved translation initiation could rescue N-terminal construct performance.

These findings show that construct optimisation through Kozak sequence addition enhances protein expression and cell viability, providing a more reliable platform for calling card

experiments. They also identify a clear avenue for future work: targeted codon redesign to enable Kozak consensus in the N-terminal fusion construct, which could determine whether the observed sensitivity is intrinsic to fusion orientation or design or simply a consequence of inefficient translation.

4.3.3. Cell passage investigation and puromycin dose optimisation

Optimisation of puromycin selection revealed that lower concentrations improved the survival of N-terminal fusion co-transfected cells. Concentrations of 1.5 µg/mL consistently yielded sufficient RNA for sequencing (Figure 4.4), and even 0.5 µg/mL achieved complete cell death within 7 days. Using the lowest effective dose therefore maximised the recovery of viable cells while maintaining selection pressure on non-transfected cells. Despite this improvement, the viability of N-terminal transfections remained insufficient for calling card analysis and additional bulk calling card sequencing was not carried out.

Cell passage number also influenced the viability of transfected cells; high passage cells were more robust than low passage cells, explaining discrepancies between optimisation experiments and the initial bulk ERα calling cards. This difference likely reflects clonal evolution during culture, where more resilient cells are selected for over time with successive passages (Anon; Selich et al., 2016; Hynds, Vladimirou and Janes, 2018; Hughes et al., 2007; Gisselsson et al., 2019). Generally, experiments should not be conducted on cells with too high passage numbers to avoid recording artefacts from acquired traits developed over time as genuine cell-line behaviour (Hughes et al., 2007; Xie et al., 2012; Wenger et al., 2004). For calling card experiments, aimed at modelling in vivo breast cancer, low-passage cells remain preferable despite their greater sensitivity.

Finally, co-transfection efficiency assays provided useful estimates of plasmid delivery, though interpretation is complicated by cell doubling during the 48-hour culture period. Using the BrokenHeart SRT reporter ensured that only the co-transfected cells would fluoresce red for counting, but the expansion of cultures means that the measured efficiency likely underestimates the true proportion of cells receiving both plasmids. This is because, unlike the TdTomato SRT, the BrokenHeart SRT cannot be inherited by daughters. MCF7 cells double approximately every 24 hours, with some evidence claiming the doubling time can vary up to 48 hours (Sutherland, Hall and Taylor, 1983; Larsson et al., 2020; Jakesz et al., 1984). It is therefore possible that 2-4x more cells than were counted could have descended from a co-transfected cell, but they couldn’t be detected as the inserted SRT does not contain the dsRed coding sequence. Alternative reporters like TdTomato would not resolve the issue, as single-transfected cells would also express the red fluorescent protein.

These findings show that the puromycin dose and passage number both affect cell viability and that co-transfection efficiency can be estimated but not precisely quantified under current conditions. While these optimisations improved recovery for the C-terminal construct, they were insufficient to rescue the N-terminal fusion, reinforcing the conclusion that further protein design changes will be required to achieve functional calling cards with this construct.

Taken together, the co-transfection efficiency determination was the best we could achieve with the materials we had, but it may not completely reflect co-transfection efficiency.

4.3.4. Limited insertions per cell constrained single-cell analysis, necessitating considering samples as bulk calling cards for evaluation

Given that the dedicated calling card fraction of the 10X eluate had been exhausted during attempts to adapt the short-read protocol (Moudgil et al., 2020), we proceeded with the transcriptomic fraction, which had already undergone PCR amplification. While this may have reduced the proportion of genuine calling card transcripts, it nonetheless allowed recovery of insertion events sufficient for downstream analysis where circularisation and short-read methods could not. To maximise interpretability, we analysed the dataset in pseudo-bulk form, pooling insertions across cells. This approach increased calling card recovery sufficiently to enable peak calling and enabled direct comparison with bulk calling card datasets using established pipelines.

Analysis of the long-read dataset revealed that the number of insertions recovered per cell was substantially lower than expected, given the average recovery using 10X genomics (De Simone et al., 2025; Zheng et al., 2017). Even under relaxed stringency, the maximum recovery was 14 insertions per cell in 120 cells, compared with typical single-cell RNA-seq datasets that detect hundreds of features per cell (Svensson, Vento-Tormo and Teichmann, 2018; Luecken and Theis, 2019). The low recovery meant that individual cell profiles could not be distinguished from background, limiting confidence in single-cell level interpretation

The pseudobulk analysis confirmed that genuine ERα-associated peaks could be detected even from a sub-optimal sample, demonstrating the feasibility of long-read calling cards for mapping binding events. Although the low-per-cell insertion recovery prevented single-cell resolution in this experiment, the results establish proof of principle that with improved library preparation and sequencing depth, long-read calling cards could generate the first single-cell ERα binding profiles.

4.3.5. Long-read calling cards reproduce bulk overlap patterns: Overlap with ChIP-Seq peaks but stronger concordance with ChIA-PET anchors

Overlap analysis of long-read ERα calling cards revealed the same pattern observed in bulk experiments: modest to good overlap with ChIP-Seq peaks (Holding et al., 2019; ENCODE Project Consortium, 2012) but consistently stronger overlap with ChIA-PET anchors. Across untrimmed datasets, overlap with ChIP-Seq ranged from 9-22%, which is lower than bulk calling cards (Figure 8.12). This suggests that reduced overlap does not necessarily indicate a lack of function but reflects both dataset choice and inherent variability in peak calling. Differences in culture conditions may also contribute: GEO ChIP-Seq was performed in hormone-stripped media with estradiol treatment, whereas our long-read experiments were conducted in full media due to growth limitations.

It is more complicated to assess the effect of cell culture conditions for the Encode dataset (ENCODE Project Consortium, 2012), as the exact conditions for the cells used in ChIP-Seq are not publicly available. The standard cell culture protocol used for MCF7 cells in other Encode projects does indeed use hormone-stripped media and supplements with E2. However, the JSON with metadata for the specific experiment linked to ENCFF063JMY does not state that any treatment was used for any of the samples. It is hard to be sure of the truth, given the lack of clarity in the published material, but if these cells were not treated with E2, this could be a contributing factor to the higher overlap observed with ENCODE peaks compared to GEO (Holding et al., 2019).

In contrast, overlap with ChIA-PET anchors was consistently high with 63% of long-read calling card peaks overlapping with either anchor (Fullwood et al., 2009; Kent et al., 2002). This highlights the ability of long-read calling cards to capture ERα's long-range chromatin interactions similarly to bulk calling cards (Section 3.2). The higher overlap with ChIA-PET compared to ChIP-Seq may reflect the structural properties of the HyPB-fusion protein, which inserts within ~1 kb of TF binding sites (Moudgil et al., 2020; Wang et al., 2012a). ERα binding can often involve looping contacts across chromatin; these sites may be easier to detect in noisy datasets than local binding events that are measured by ChIP-Seq.

It is important to note that the very high overlap observed for untrimmed UMI-filtered peaks could be inflated by the small number of peaks remaining after filtering (10 out of 506). Similarly, the clustering of ChIA-PET anchors within short genomic distances can increase apparent overlap, as a single long-read peak may overlap multiple nearby interactions. Our analysis accounted for this by reporting the percentage of overlapping long-read peaks

regardless of multiple overlapping anchors, but the proximity of anchors remains a caveat when interpreting high overlap percentages.

These findings show that long-read calling cards reproduce the characteristic overlap patterns of bulk experiment: modest to good overlap with ChIP-Seq peaks but strong overlap with ChIA-PET anchors. This consistency across datasets supports the conclusion that long-read sequencing provides a viable alternative to circularisation-based methods, enabling direct recovery of ERα binding information even for challenging samples.

4.3.6. ChIP-Seq signal at the centre of long-read calling card peaks reflects overlap between the datasets

To complement the overlap analysis, we examined ChIP-Seq signal intensity at long-read calling card peaks. GEO datasets showed little enrichment: Average $\log_2$ fold change values remained close to baseline, and heatmaps revealed a weak signal across loci (Figure 4.6). This is consistent with the low percentage of overlapping long-read calling cards with GEO peaks.

In contrast, ENCODE datasets show clear enrichment of ChIP-Seq signal both at the centre of and within nearby regions. Across unfiltered insertions, the average $\log_2$ fold change reached ~3.25 at the centre of peaks, with a broad enrichment profile extending ±10 kb (**Error! Reference source not found.** A). This pattern mirrors bulk calling cards, though the central peak is less sharply defined, reflecting the high levels of noise in long-read datasets.

Applying UMI filtering increased the average $\log_2$ fold change across the entire 20 kb region (**Error! Reference source not found.** C). This effect arises because filtering removes low-support insertions (single-UMI events), leaving only loci supported by multiple UMIs. As a result, the remaining peaks show a stronger ChIP-Seq signal, confirming that filtering improves specificity.

These results show that ENCODE ChIP-Seq signal is enriched at long-read calling card peaks (Appendix Figure 8.15), particularly after UMI filtering. GEO datasets did not show enrichment, consistent with their lower overlap. The ENCODE enrichment supports the function of the HyPB-ERα fusion protein and demonstrates that long-read calling cards can recover genuine ERα binding hotspots detectable by ChIP-Seq.

4.3.7. Limited overlap between bulk and long-read calling cards highlights the need for reproducibility metrics

Comparison of long-read and bulk ERα calling card datasets revealed that overlap between peaks was lower than expected (Appendix Figure 8.16 and Figure 8.17). Most long-read peaks were located at loci detected in bulk experiments, as only one peak overlapped within 1000 bp. This could suggest that long-read sequencing recovers additional sites; it is equally possible that the low overlap reflects the limited number of replicates and small data size. Without statistical testing, we cannot conclude that the loci are truly independent.

In ChIP-Seq analysis, reproducibility is routinely assessed using saturation curves to determine whether peak calling is still in the linear range or has reached saturation with the maximum number of peaks that can be detected in the data (Boyd et al., 2023; Jung et al., 2014; Sundaram et al., 2016). The IDR framework is then applied to the down-sampled replicate datasets to determine how many peaks can be called at each step (Li et al., 2011; Jalili et al., 2015). These approaches provide quantitative metrics for overlap and reproducibility, but calling card analysis currently lacks equivalent tools, as we discussed in Chapter 3. Developing an adapted pipeline, for example, applying IDR to down-sampled calling card datasets, would allow determination of whether peak detection is saturated and provide a statistical basis for comparing overlap between bulk and long-read calling card experiments.

Rather than indicating inconsistency, the limited overlap observed between ERα calling card datasets highlights the need for such reproducibility metrics for calling cards. Both bulk and long-read experiments were constrained by material and sequencing depth, and additional replicates combined with an IDR-like framework would likely increase the number of significant peaks called and provide a more complex picture of ERα binding. Establishing these methods will be essential for future work to benchmark calling card reproducibility against established standards in ChIP-Seq.

Across the KRT8/18 locus, three long-read ERα calling card peaks were identified (Figure 4.8 B). The first peak contained 97 insertions, each with a score of one, resulting in a flat experimental track relative to the other peaks. The other two peaks contained over 500 insertions each with varying scores, producing much higher-density and experiment insertion signals than the first peak. Although KRT8/18 aren't classical ERα targets, they are expressed in ERα-positive cells and can be modulated by ERα activity. The overlap with ChIP-Seq peaks and many insertions support these long-read peaks as genuine ERα binding events.

Finally, two peaks were identified in the DDX5 gene (Figure 4.8 C). These peaks appeared lower in the experimental track than in the surrounding, non-significant insertion regions, but inspection showed 20-30 insertions across different cells contributing to each peak. Both overlapped with ChIP-Seq peaks in both datasets. DDX5 is a known co-regulator of ERα, and its involvement in ERα transcriptional programmes makes these strong candidates for genuine insertions (Fuller-Pace and Ali, 2008; Wortham et al., 2009; Nicol and Fuller-Pace, 2010).

These examples show that long-read ERα calling card recover peaks at established ERα target genes and co-regulators, with overlap to ChIP-Seq peaks and support from multiple independent cell barcodes. While some peaks appear visually flat on experimental insertion tracks due to the way that track height is calculated, closer inspection revealed that substantial numbers of single insertions across many cells contribute to them. These results demonstrate that long-read calling cards can map ERα binding information with confidence, complementing the overlap and signal analyses presented above.

4.3.8. Long-read calling card peaks are located at biologically relevant ERα peaks and highlight confidence in long-read calling cards

Long-read calling cards recovered peaks at canonical ERα targets (GREB1, TFF1) (Sun, Nawaz and Slingerland, 2007; Haines et al., 2020; Hodgkinson et al., 2018; Cheng, Michalski and Kommagani, 2018; Baron et al., 2007; Amiry et al., 2009; Yi et al., 2020), cancer-associated genes (KCNJ3 (Rezania et al., 2016; Kammerer et al., 2016), VMP1 (Amirfallah et al., 2019; Inaki et al., 2011; Persson et al., 2017), KRT8/18 (Meng et al., 2009; Wilson and Dering, 2004; Nguyen et al., 2018; Ha et al., 2011; Iyer et al., 2013)), and coregulators (DDX5 (Fuller-Pace and Ali, 2008; Wortham et al., 2009; Nicol and Fuller-Pace, 2010)). The overlap of these peaks with ChIPseq datasets provides strong evidence that the insertions represent genuine ER binding events. This result demonstrates that long-read calling cards capture both direct ER targets and loci relevant to breast cancer biology, reinforcing their utility for mechanistic studies.

Beyond biological relevance, our confidence in peaks depends on how insertions are distributed. Peaks supported by many cells, even if each insertion has a score of one, indicate reproducibility across independent replicates. Conversely, peaks with high scores from single barcodes reflect multiple insertions in the same cell, which may indicate highly active sites but provide less evidence of reproducibility. Both metrics contribute, but reproducibility across cells is the stronger indicator of genuine binding.

Taken together, these findings show that longread calling cards can recover biologically meaningful ERα peaks and highlight the need for clear reproducibility metrics. Future work should establish combined measures of cell-level replication and UMI support, providing a robust framework for peak confidence analogous to IDR in ChIP-Seq analysis.

4.3.8.1. Future work

Future work should build directly on the methodological advances described in this chapter by refining construct design, optimising experimental conditions, and developing reproducibility frameworks for calling card analysis. At the construct level, codon redesign of the N-terminal fusion to enable Kozak consensus could test whether improved translation initiation rescues performance, complementing the gains already achieved with the C-terminal construct. In parallel, further optimisation of puromycin selection and passage number will be important to maximise viability in single-cell experiments, ensuring that low passage cells can be used without compromising recovery.

At the sequencing and library preparation stage, systematic evaluation of primer designs and amplification strategies will be needed to confirm whether correct amplification improves insertion recovery. Establishing robust single-cell libraries enriched for genuine ER-directed insertions will allow long-read sequencing to achieve sufficient depth for confident peak calling. Filtering strategies should also be benchmarked, including UMI stringency and down-sampling, to balance specificity and sensitivity. These steps will clarify how thresholds affect overlap with ChIP-Seq and ChIA-PET datasets and ensure that signal enrichment reflects genuine binding events.

Beyond individual datasets, reproducibility must be addressed with quantitative metrics. In ChIP-Seq, reproducibility is routinely assessed using saturation curves to determine whether peak calling remains in the linear range and by applying the IDR framework to replicate datasets. Developing an adapted pipeline for calling cards, for example, applying IDR to down-

sampled datasets, would provide a statistical basis for comparing overlap between bulk and long-read experiments and determine whether peak detection is saturated. Combined measures of cell-level replication and UMI support will further strengthen confidence in peak calling, establishing reproducibility standards for calling cards analogous to those used in ChIP-Seq.

Long-term, integrating the HyPB domain into the endogenous ER locus of stable cell lines or patient-derived xenografts would overcome the limitations of co-transfection efficiency. This approach would increase the proportion of cells capable of inserting calling cards, improving RNA yield and heterogeneity captured. With appropriate controls and E2 conditions, such models could more faithfully replicate endogenous ER binding.

Finally, these methodological improvements will enable the next stage of this project: linking ER binding heterogeneity to microenvironmental factors. In particular, applying single-cell long-read calling cards under hypoxia or other stress conditions will allow direct investigation of how the tumour microenvironment shapes ER binding programmes. This will achieve the original aim of dissecting ERα binding heterogeneity in breast cancer at single-cell resolution.

5. Creating an MCF7 spheroid model for hypoxia

5.1. Introduction

The overarching aim of this thesis is to utilise calling cards to investigate ERα redistribution under hypoxia. Following the previous chapter's establishment of ERα long-read calling cards, this chapter focuses on establishing a physiologically relevant 3D model for studying ERα signalling under hypoxia. Future applications of ERα calling cards would benefit from integration with 3D culture systems, as this would enable exploration of the complexity of heterogeneous ERα binding under hypoxia in a way that more faithfully replicates clinical disease. Extensive research shows that 3D cell culture systems respond to drug treatment and other biological assays in ways that more closely resemble clinical samples than 2D culture methods (Koedoot et al., 2021; Kenny et al., 2007; Stock et al., 2016; Cavo et al., 2018; Guillen et al., 2022; Ludwik et al., 2024). Among 3D systems, spheroids provide a highly accessible and reproducible first model, avoiding the ethical and technical challenges of PDX or organoids (Zanoni et al., 2020; Goldrick et al., 2023; Han, Kwon and Kim, 2021; Idrisova, Simon and Gomzikova, 2022). Combining ERα calling cards with 3D culture methods would enable capturing the binding heterogeneity of ERα in therapy-resistant and sensitive breast

cancer cells over a period of time at single-cell resolution across an impaired oxygen concentration gradient at the core of cultures. This chapter, therefore, establishes spheroid culture as the first accessible 3D cell culture model for integrating ERα calling cards.

Hypoxia is commonly identified in patient tumours (Asano et al., 2018; Shamis et al., 2022; Collin et al., 2021; Fox et al., 2006) and has been linked to poor prognosis and resistance to anti-cancer therapies (Collin et al., 2021; Fox et al., 2006; Shamis et al., 2022; Yang et al., 2015; Jia et al., 2015; Li et al., 2020b; Park and Lee, 2017). MCF7 spheroids present a controllable system to model hypoxia development and study ERα functions when oxygen is limited due to spheroid size (Palma et al., 2016; Muñoz-Galindo et al., 2019; Rofstad et al., 1996; Mukomoto et al., 2020; Yakavets et al., 2020; Singh et al., 2016). It is possible to show the development of hypoxia through staining with the 2-nitroimidazole EF5. EF5 is a molecule that is reduced in low oxygen environments to form adducts (Wang et al., 2012b; Pugh-Toole et al., 2022; Mikhail, Eetezadi and Allen, 2013; Lao et al., 2015; Grimes et al., 2014; Pires et al., 2012) similar to the molecule pimonidazole, which has been well established as a marker for hypoxia (Swartz et al., 2022; Varia et al., 1998; Ragnum et al., 2015). Given that ERα signalling under hypoxia is implicated in endocrine therapy resistance (Yang et al., 2015; Ryu, Park and Lee, 2011; Kurebayashi et al., 2001), examining its behaviour under hypoxia at single-cell resolution is directly relevant to patient outcomes. This clinical relevance provides the rationale for using spheroids to capture ERα binding redistribution during hypoxia. Breast cancer cells transfected with calling card reagents could be grown to a size where hypoxia begins to develop naturally due to the diameter of the spheroid, which has been reported previously (Palma et al., 2016; Muñoz-Galindo et al., 2019; Rofstad et al., 1996; Mukomoto et al., 2020; Yakavets et al., 2020; Singh et al., 2016). Transfection with calling card reagents before hypoxia establishment would therefore utilise the longitudinal recording function of calling cards (Yen et al., 2023; Cammack et al., 2020), marking how ERα binding changes in response to stimuli such as hypoxia. This approach provides a novel means of tracking ERα redistribution during the onset of hypoxia in a reproducible 3D model

EF5 has been used to demonstrate the development of hypoxia in MCF7 spheroids previously in a pre-print (Pugh-Toole et al., 2022), but it was not used to characterise the development of hypoxia in MCF7 spheroids as we do in this section. Therefore, while hypoxia has been observed in these spheroid models, there is little consistent information on the initiation and development of hypoxia in spheroids (Beckers et al., 2025). Addressing this gap is therefore essential before integrating MCF7 spheroid models with ERα calling cards. This chapter addresses that gap by systematically characterising hypoxia development in MCF7 spheroids and validating them for ER studies

Building on this rationale, we next examined our complex RNA-seq dataset in which we identified an ERα-Dependent Hypoxic Response (EDHR) that outlines several ERα target genes that are specifically upregulated during hypoxia (Malcolm et al., 2025). In this pre-print, two of the genes identified, *SCNN1B* and *SCNN1G*, stood out due to the recently reported role of the protein they form, EnAC, as a negative regulator of MCF7 cell growth (McQueen et al., 2025; Ware et al., 2021). EnAC is a trimer of three subunits, the α, β and γ subunits. Inhibition of the EnAC with 5 mM amiloride resulted in increased growth in several breast cancer cell lines, including MCF7s. While these publications do not highlight the β and γ subunits, we found that high levels of their expression correlate with worse recurrence-free survival, in contrast to the decrease in overall survival reported for the α subunit (Ware et al., 2021). This distinction between subunit-specific outcomes has not been explored in the context of hypoxia, positioning our findings as an extension of current knowledge.

It remains unclear whether ERα persists in hypoxic regions of spheroids, as some publications report its activity and expression (Kapara et al., 2021; Muñoz et al., 2010), whereas others claim that ERα is not the main driver of the proliferation of MCF7 spheroids (Chen, Liu and Yan, 2022; Whitman et al., 2019). Clarifying this uncertainty is central to validating spheroids as a model for ERα redistribution heterogeneity under hypoxia.

This chapter aims to validate an MCF7 spheroid model for studying ERα signalling under hypoxia using the ERα calling card reporter. Our objectives to achieve this aim are:

1. Validate an MCF7 spheroid model for the development of hypoxia using fluorescent imaging
2. Confirm the expression of ERα within the hypoxic region of MCF7 spheroids
3. Demonstrate the relevance of hypoxic MCF7 spheroids for studying the effects of low oxygen on the ERα transcriptional programme by targeting the hypoxia-induced ERα target EnAC for specific inhibition across a range of concentrations.

Objective 1 establishes our model, Objective 2 validates ERα persistence, and Objective 3 demonstrates functional relevance in hypoxia research before integration with ERα calling cards.

We combine MCF7 spheroid culture with immunostaining, quantitative image analysis and pharmacological inhibition to address these aims. There are many variations of 3D cell culture utilising different types of 3D scaffolds, as well as media combinations (Singh et al., 2016; Pulze et al., 2020; Kelm et al., 2003; Alhawarat et al., 2019; Nagpal et al., 2015; Doublier et

al., 2012; Tevis et al., 2017; Imamura et al., 2015; Beckers et al., 2025; Yakavets et al., 2020; Jardin et al., 2023; Chen, Liu and Yan, 2022). Some groups use phenol red-free media (Yakavets et al., 2020; Chen, Liu and Yan, 2022; Jardin et al., 2023; Morgan et al., 2018), others use phenol red (Singh et al., 2016; Pulze et al., 2020; Kelm et al., 2003; Alhawarat et al., 2019; Nagpal et al., 2015; Doublier et al., 2012; Tevis et al., 2017; Imamura et al., 2015; Beckers et al., 2025) Others add other supplements such as B27 (Abd El-Sadek et al., 2024; Paolillo et al., 2019; Wang et al., 2014a; Alhawarat et al., 2019), anti-biotics (Pugh-Toole et al., 2022; Singh et al., 2016; Pulze et al., 2020; Chen, Liu and Yan, 2022; Nagpal et al., 2015; Doublier et al., 2012), and insulin and other growth factors (Kelm et al., 2003; Alhawarat et al., 2019; Abd El-Sadek et al., 2024). This degree of variation highlights the need for a clear, reproducible model; our approach defines such a method, tailored for integration with ERα calling cards.

We therefore focused our efforts on a spheroid method and set out to define a standard MCF7 spheroid method that could be used for future applications with ERα calling cards. A recent publication took a similar seeding approach to us (Beckers et al., 2025), highlighting the need for this work in the field.

5.2. Results

5.2.1. MCF7 cells seeded at 500 and 1000 cells/well generate spheroids 400 to 600 µm in diameter after 7 days

MCF7 spheroids were grown in a ULA 96-well plate at varying seeding densities to optimise control of spheroid diameter. We defined a target diameter range of 400-600 µm at day 7 in culture, based on reports that hypoxia develops within this range and that MCF7 spheroids form complex cell-to-cell connections by day 7 (Yakavets et al., 2020; Singh et al., 2016; Pulze et al., 2020; Palma et al., 2016).The number of cells seeded in ULA wells also impacts the final spheroid diameter (Chen, Liu and Yan, 2022; Yakavets et al., 2020); we therefore needed to determine which seeding density should be used as the primary determinant of spheroid suitability.

Spheroids were seeded at 500 to 10,000 cells/well, and were imaged from day 2 to 7 to determine the diameter over the course of 7 days (Figure 5.1).

We found that seeding 500 and 1000 cells/well to generate MCF7 spheroids created spheres within the desired range for our downstream experiments; by day 7, these spheroids consistently spanned 400 to 600 µm.

Image analysis showed that spheroids seeded from 5,000 and 10,000 cells/well exceeded the range of reported hypoxia development (Yakavets et al., 2020; Singh et al., 2016; Palma et al., 2016) by day 7. We excluded these densities to limit necrosis, which is a known concern at larger spheroid diameters (Yakavets et al., 2020; Lee et al., 2010; Däster et al., 2017; Gong et al., 2015), and to ensure that spheroids establish mature cell-to-cell connections before exhibiting hypoxia (Pulze et al., 2020). In this model, necrosis is assessed indirectly through characteristic morphological features: such as core cavitation, loss of cellular cohesion and hollowing; rather than through a dedicated necrosis marker, which should be considered when interpreting the extent of cell death

Here, we have shown that seeding ULA wells in 96-well plates with 500 and 1000 cells/well results in the growth of spheroids between 400 and 600 µm in diameter after 7 days in culture with standard media.

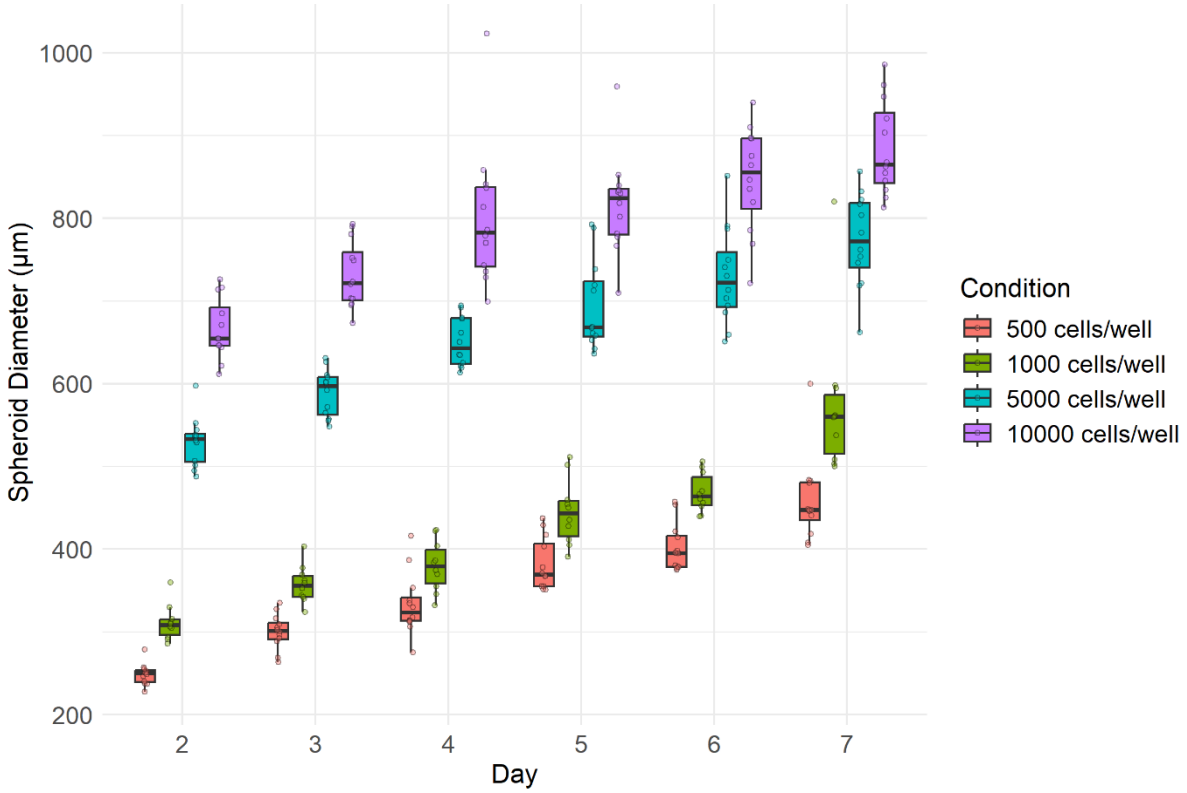


Figure 5.1: **MCF7 cells seeded at 500 and 1000 cells/well grow within the ~400-600 µm range where hypoxia is reported to develop**

MCF7 spheroids (n = 12) were imaged over the course of 7 days to determine a seeding density that produces spheres that grow between 400-600 µm. Both 5000 and 10000 cells/well grew outside of this size range and varied in size between 630 and 1000 µm. Whereas 1000 and 500 cells/well ranged between 400 and 600 µm in diameter at day 7, making these the most suitable seeding densities for further work.

5.2.2. MCF7 spheroids grown in the same media conditions as 2D culture grow to the desired size range in 7 days

Having demonstrated that our model can grow spheres within the range where hypoxia develops, we next tested several media conditions to define the standard system for our spheroid model. Establishing standard conditions is necessary, given the variation in methods across MCF7 spheroid publications (Singh et al., 2016; Pulze et al., 2020; Kelm et al., 2003; Alhawarat et al., 2019; Nagpal et al., 2015; Doublier et al., 2012; Tevis et al., 2017; Imamura et al., 2015; Beckers et al., 2025; Yakavets et al., 2020; Jardin et al., 2023; Chen, Liu and Yan, 2022). We compared the standard 2D cell culture condition (DMEM +Phenol Red + 10% FBS) with Phenol Red Free DMEM, as this would allow future exploration of ±E2 treatments in the model.

In addition to comparing DMEM with or without Phenol Red, we tested other conditions reported in the literature: DMEM -Phenol Red supplemented with B27 and Glutamax (Pulze et al., 2020; Alhawarat et al., 2019; Kang et al., 2019; Mokhtari et al., 2021). These conditions were tested alongside the standard DMEM + Phenol Red + FBS condition.

Brightfield images of spheroids were taken (Figure 5.2) and the diameter of spheroids on day 7 was plotted to compare the effects of media supplements (Figure 5.3).

All spheroids, except the 1000 cell/well + B27 and Glutamax, grew within the 400 to 600 μm range across the 7 days. Mean diameters differed significantly between conditions (ANOVA, $p < 0.001$). Post hoc testing showed that spheroids seeded from 500 cells/well in Phenol Red DMEM were significantly smaller than those in Phenol Red-free media (letters e vs. d), whereas cells seeded at 1000 cells/well were comparable (c). All spheroids grown with additional supplements were significantly larger than other conditions (letters a and b).

While B27 supplementation produced larger spheroids, it offers no further clear benefit compared to full media, so it was not taken forward.

In testing conditions reported in the literature, and evaluated them against our criteria, we found that Phenol Red DMEM produces spheroids in the desired range most consistently, so we selected this as our standard condition. By selecting a system that grows the smallest spheroid within our desired range, we can examine the earliest stages of hypoxia development in these spheroids. Furthermore, this choice of media aligns with the majority of published breast cancer research, enabling direct comparison in future work.

All subsequent experiments were carried out using the same FBS lot to ensure consistency. Based on these findings, downstream analyses used 1000 cells/well seeded in DMEM + Phenol Red + 10% lot-controlled FBS, which reproducibly generated spheroids within the diameter range where hypoxia can develop (Yakavets et al., 2020; Singh et al., 2016).

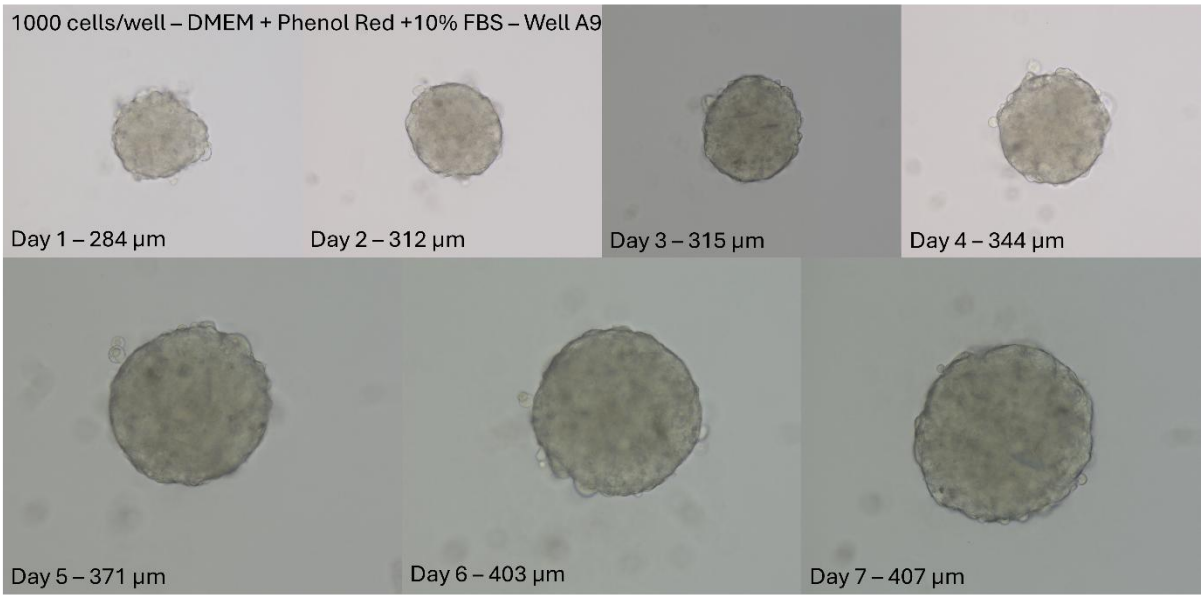


Figure 5.2: **MCF7 Spheroids initially form cell aggregates that develop into defined spheroids after 7 days.**

This figure shows the growth of a representative MCF7 spheroid over the course of 7 days. Measuring spheroid diameter on day 1 produces high levels of variance across spheres as the cells are still aggregating together and have not formed complete cell junction connections. By day 2, cells usually combine to a single sphere aggregate that develops cell connections over days 3 and 4. By day 5, most spheres have continuous surfaces with more developed cell contacts.

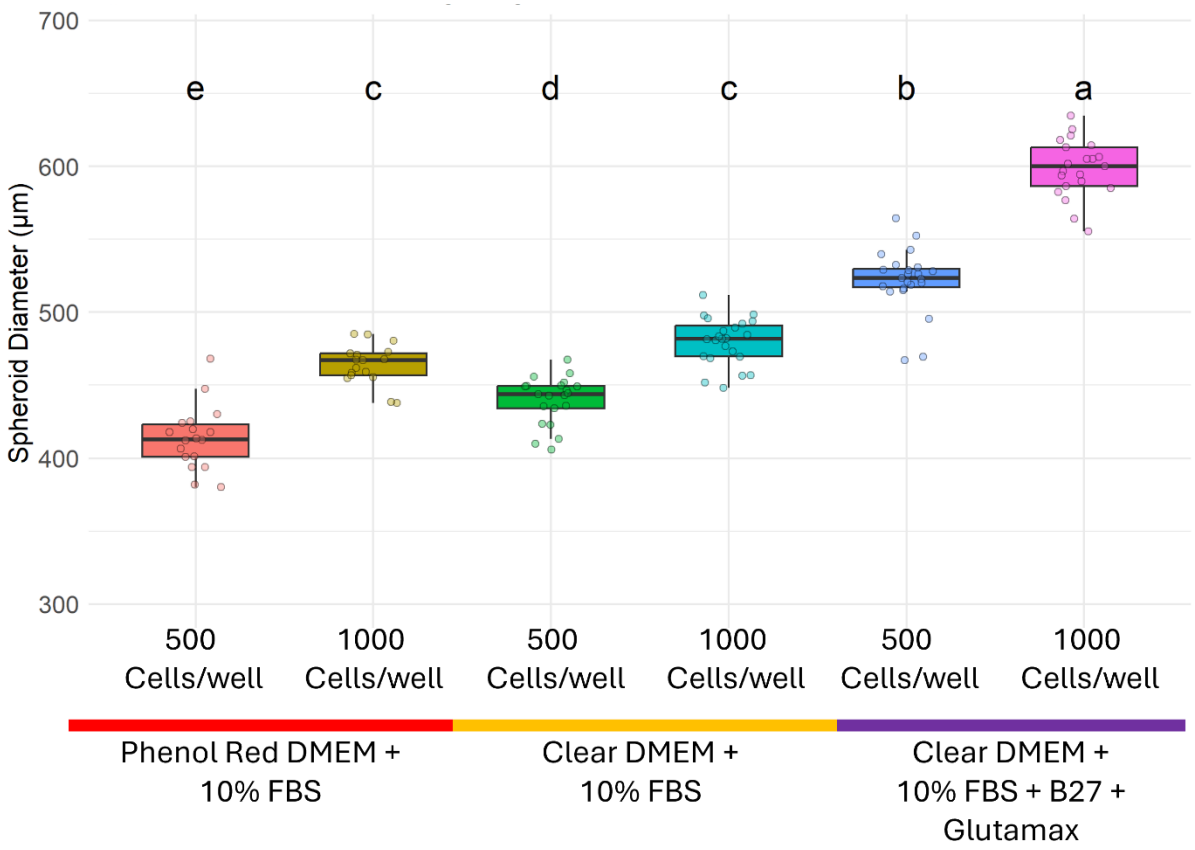


Figure 5.3: **Culturing MCF7 spheroids in DMEM with or without phenol red is sufficient for growing spheroids >500 µm in diameter after 7 days.**

MCF7 spheroids were grown in a variety of cell culture media for 7 days in (i) DMEM + Phenol Red + 10% FBS, (ii) DMEM - Phenol Red + 10% FBS, (iii) DMEM - Phenol Red + 10% FBS + 1x B27 + Glutamax. Spheroids were then imaged on day 7, and their diameter was measured to compare with the other conditions. Bars represent mean ±SEM. Different letters above bars indicate statistically significant differences between conditions (ANOVA, P > 0.001; Tukey post hoc). Both sets of DMEM with only +10% FBS added produced spheres that were between 400 and 500 µm, the desired size range for downstream hypoxic analysis. DMEM without phenol red and supplemented with B27 and Glutamax grew larger than the other conditions, with spheres growing between 500 and 650 µm.

5.2.3. MCF7 spheroids demonstrate hypoxia development correlated with culture time and spheroid size

With the establishment of standard MCF7 spheroid culture conditions for future experiments, we proceeded to characterise the development of hypoxia in our model.

MCF7 spheroids were cultured for 13 days, and beginning on day 7, spheroids were incubated with the hypoxia marker EF5 before fixation with PFA at days 7, 9 11 and 13. This time course allowed us to measure hypoxia development across the size range in which spheroids are reported to develop hypoxia (Yakavets et al., 2020; Singh et al., 2016). EF5 is reduced in low oxygen conditions, forming adducts that are detected through immunostaining with an Anti-EF5 antibody, conjugated to AF488 (Wang et al., 2012b; Pugh-Toole et al., 2022; Mikhail, Eetezadi and Allen, 2013; Lao et al., 2015; Grimes et al., 2014; Pires et al., 2012), thereby monitoring the presence or absence of hypoxia within spheroids. EF5 is therefore a promising candidate to monitor hypoxia development in our spheroid model.

Cryosectioning and immunostaining of fixed MCF7 spheroids treated with EF5 reveal a clear progression of hypoxia development over the course of days 7 to 13 in culture (Figure 5.4). At day 7, spheroids averaged ~450 μm in diameter and showed no EF5 signal consistent with the vehicle controls. By day 9, spheroids reached ~500 μm in diameter and exhibited initial EF5 fluorescence in cells at the centre of spheres, marking the onset of hypoxia. EF5 signal increases at day 11, with a considerable proportion of the spheroid exhibiting fluorescence accompanied by structural damage at the core, including gaps between layers of cells that indicate necrosis as defined in section 5.2.1 (Bell et al., 2001; Huang, Yu and Tang, 2020; Mukomoto et al., 2020; Mangani et al., 2025). By day 13, EF5 signal is widespread in spheroids, forming halos of signal throughout the core, and central cell death had produced hallowed regions within spheroids. Some peripheral cells also showed EF5 signal, suggesting hypoxia extended toward the spheroid surface.

Taken together, these results show that EF5 immunostaining can be used to monitor the development of hypoxia in MCF7 spheroids across the published diameter range for hypoxia (Yakavets et al., 2020; Singh et al., 2016).

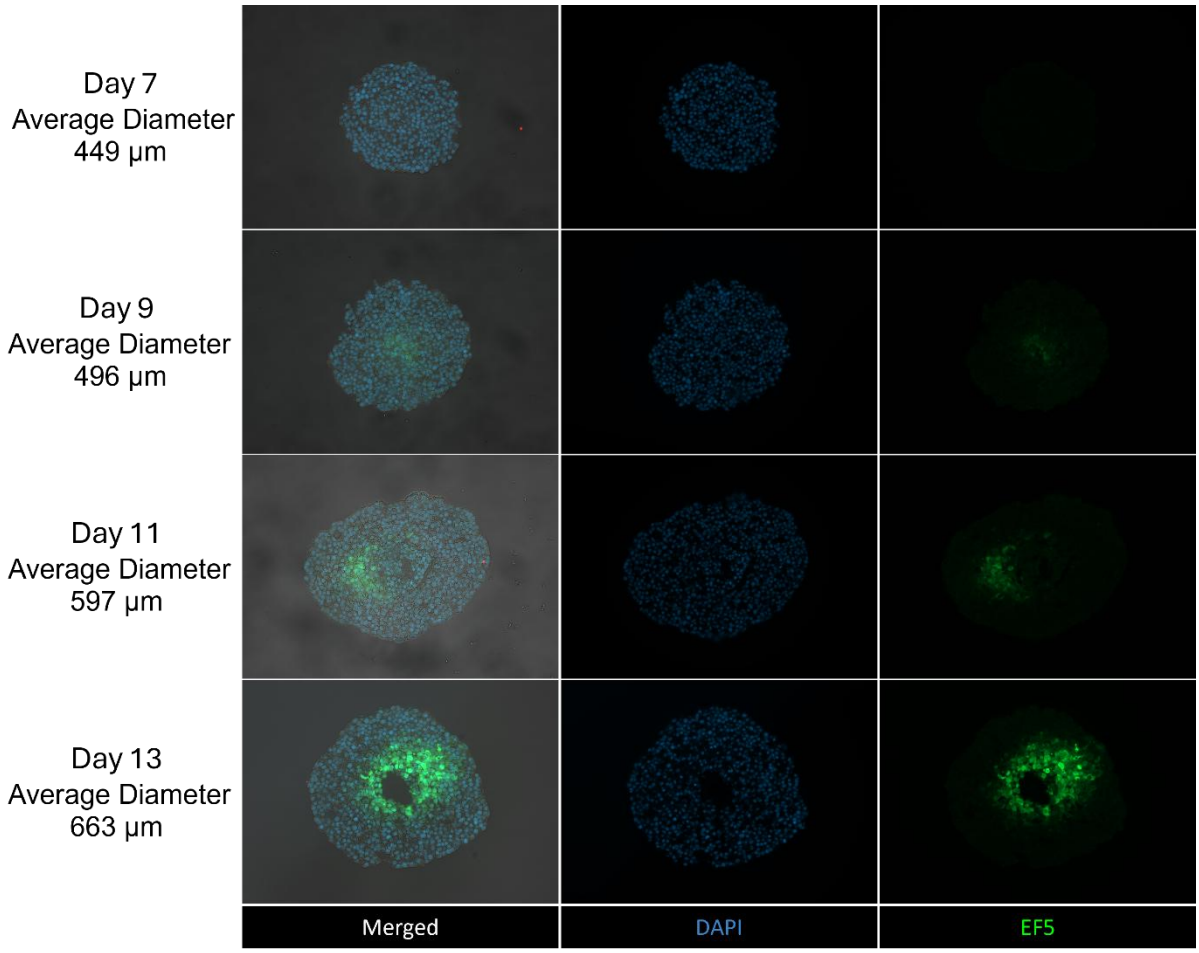


Figure 5.4: **MCF7 spheroids show EF5 fluorescent signal once they reach ~500 µm in diameter, with signal intensity and distribution increasing as spheroids grow larger.**

The average diameter was determined in ImageJ, from images of spheroids before fixation. Spheroids grown to an average diameter of 449 µm by day 7 did not show an EF5 signal at the core. By day 9, spheroids reached ~500 µm and exhibited initial EF5 fluorescence in central cells. At day 11, the EF5 signal was increased and associated with structural damage at the core, including gaps between cell layers. By day 13, the EF5 signal was widespread throughout the core, with central cell death producing hollowed regions.

5.2.4. MCF7 spheroids show radial distribution of EF5 signal intensity across time in culture, with the highest signal in spheroid cores

We quantified hypoxia across spheroids using CellProfiler to measure EF5-derived fluorescence intensity at varying depths. Each spheroid was divided into six concentric radial bins, with bin 1 representing the core and bin 6 the surface layer of cells.

By plotting mean intensity within each bin, we assessed how EF5 signal varied from centre to surface (Figure 5.5). Gradients of these mean intensity plots were compared to vehicle-treated

spheroids pooled across days 7-13 (Figure 5.6). For this analysis, the core refers to bins 1-3 and the surface to bins 4-6

Mean EF5 fluorescence was calculated across radial bins to assess the distribution of signal across the spheroid sections. Differences in the radial slope intensity

From day 9 onwards, the EF5 signal was consistently highest in bins closer to the core of spheroids than bins at the surface (Figure 5.5). Gradients of the mean intensity plots were significantly more negative than the vehicle controls (two-sample t-test, Day 7 n = 17, Day 9 n = 19, Day 11 n = 32, Day 13 n= 30; spheroids per group; Figure 5.6). Day 7 spheroids were the only EF5-treated samples that did not significantly deviate from the vehicle-treated spheroids, showing uniformity across bins. By contrast, spheroids at days 9-13 showed overall negative correlations between mean intensity and bin number.

More than 50% of spheroids at days 11-13 showed peak EF5 intensity between bins 2-4, spanning across the mid-core to inner-surface layers of spheroids. This shift indicates the expansion of the EF5 hypoxic region over time, with peak intensity consistently located within the inner regions of spheroids.

Taken together, these results demonstrate that EF5 immunostaining provides a clear readout of hypoxia development in MCF7 spheroids. Day 9 marks the very beginning of this process, with the EF5 signal first appearing in the core of spheroids at ~500 μm diameter. Notably, this onset coincides with the earliest signs of necrotic damage at the spheroid centre, and from this point onward, the EF5 signal progressively spreads into more outer layers. By days 11 and 13, peak intensity shifts toward bins 2-4, reflecting expansion of the hypoxic region as spheroids grow larger. The combination of EF5 signal progression, its spatial shift outward, and the concurrent appearance of necrotic damage is consistent with published reports of hypoxia development in multicellular spheroids (Yakavets et al., 2020; Singh et al., 2016). These findings validate our model as appropriate for studying the onset and spread of hypoxia in MCF7 spheroids.

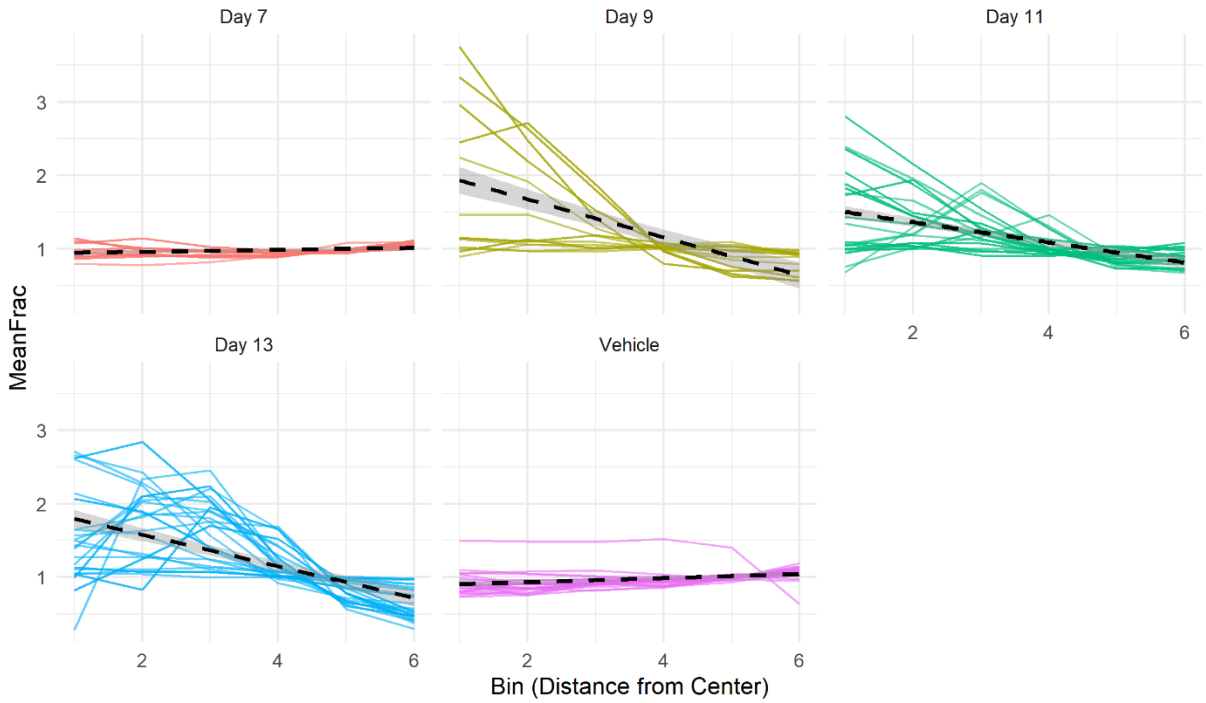


Figure 5.5: **Radial distribution of EF5 signal intensity in MCF7 spheroids shows a negative correlation with distance from the centre of the spheroids from day 9 onwards**

Vehicle-only spheroids pooled across days 7 to 13 showed no EF5 signal with a mean intensity of ~1 across bins. EF5-treated spheroids at days 9-13 showed higher intensity in central bins (1-3), decreasing toward the surface (4-6). The number of spheroids with this negative correlation increased with culture time, with day 13 showing the most pronounced effect. Some spheres on day 9 appear to show a progressive decrease in MeanFrac as bin sizes increase, except one, which shows its highest MeanFrac in bin 2. At days 11 and 13, more spheres with higher Mean intensity in bins 2 and 3 are apparent. Spheroids from each day also appear to show an even distribution of EF5 signal, like vehicle and day 7 spheres, indicating these spheres do not show high levels of signal.

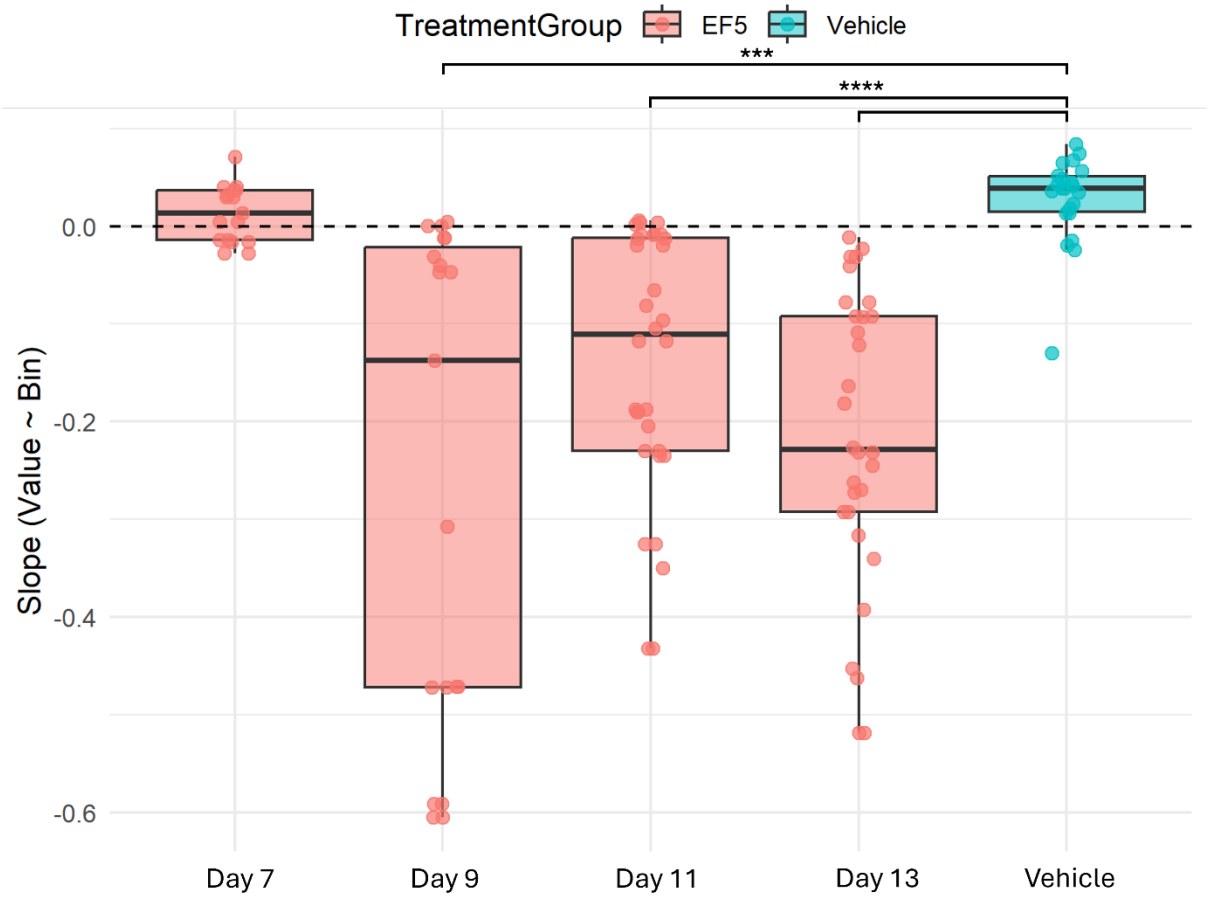


Figure 5.6: **Gradients of EF5 intensity slopes in hypoxic spheroids, day 9 onwards, show a significant negative correlation with bin size compared to vehicle-treated spheres.**

Gradients of EF5 intensity slopes were compared to vehicle controls using two-sample t-tests. Day 7 slopes were not significantly different, whereas day 913 slopes were significantly more negative (***p < 0.001, ****p < 0.0001). This shows that the EF5 signal in hypoxic spheroids has a statistically significant negative correlation with bin number, reflecting increased intensity toward the core.

5.2.5. High fluorescent intensity regions make up to three-quarters of MCF7 spheroids' diameter and show threefold higher intensity compared to controls by day 13

We ran a CellProfiler pipeline to create masked regions based on fluorescent signal intensity to quantify the size and intensity of hypoxic regions relative to the whole spheroid and vehicle-treated controls, complementing the radial distribution analysis above. Using our CellProfiler pipeline, A pool of three representative images for each day of spheroids harvested (Day 7 n = 17, Day 9 n = 19, Day 11 n = 32, Day 13 n = 30), a threshold of 0.9 units was selected to highlight genuine EF5 signal and deselect regions missing any signal. This threshold was then uniformly applied to all images in the dataset.

Our pipeline determined masks for the whole spheroid region and for hypoxic regions. For vehicle-treated spheroids, the ‘vehicle mask’ was equivalent to the whole sphere, serving as an important negative control. CellProfiler functions were then used on all the masks to compare the size and mean intensity within the masked regions. The three masks used were the thresholded ‘High signal’ region, the ‘Whole Sphere’, and the ‘Vehicle’ mask.

Both the area and Feret diameter of high signal regions grow between day 9 and 13, with the most mature spheres making up to a third of the total spheroid diameter (Figure 5.7). The Feret diameter is defined as the maximum distance between any two points on the object's outline when a line of measurement is set at a specific angle; also known as the largest diameter that can be measured between two points of an object. While some High signal regions were called within day 7 spheroids, these are all the result of artefactual staining from aggregates in secondary antibodies that are only a few pixels wide (Appendix Figure 8.20). Only 5 out of the 22 day 7 spheroids stained showed any ‘High signal’. These were retained here for transparency but can be filtered by applying a minimum area threshold, as they represent non-specific staining rather than genuine hypoxia.

By day 13, the area and diameter of high signal regions approached the values measured for the whole spheroid mask, indicating that hypoxia occupied a considerable proportion of the spheroid volume. Day 9 and 11 spheroids showed high signal regions with diameters spanning ~50% of the total spheroid diameter. By day 13, the median Feret diameter of high signal regions reached ~75% of the median diameter of whole and vehicle spheroids, meaning that hypoxic regions spanned three-quarters of the spheroid's diameter and encompassed most of the cells in the culture.

In addition to spatial expansion, the mean fluorescence intensity within high signal regions also increases over time. Day 9 and 11 spheroids showed median intensities of ~1.3-1.4, indicating modest changes despite growth in hypoxic region size. By day 13, however, mean intensities varied from 0.9 – 3.5, with the most intense regions displaying over three times the mean intensity compared to the whole spheroid mask. This demonstrates that hypoxic regions at day 13 not only span most of the spheroid diameter, but also exhibit markedly higher signal intensity, pulling clearly above background and vehicle controls (Figure 5.8). Artefactual signals observed in a minority of day 7 spheroids (Appendix Figure 8.20) were excluded from interpretation.

These results quantitatively show that hypoxic regions expand progressively over time and, by day 13, span up to three-quarters of the spheroid diameter. This expansion coincides with the necrotic damage and outward spread of the EF5 signal from the core, consistent with published reports of hypoxia development in spheroids, thereby validating our model as appropriate for studying the onset and progression of hypoxia.

These analyses quantitatively show that hypoxic regions in MCF7 spheroids expand progressively in both size and intensity. Day 9 marks the onset of hypoxia, with modest increases in signal intensity and region size, while day 13 spheroids exhibit hypoxic regions spanning up to three-quarters of the spheroid diameter and showing mean intensities more than three times higher than earlier time points. These findings confirm that day 13 hypoxic regions rise clearly above background and vehicle controls. Together, these results are consistent with published reports of hypoxia development in spheroids and validate our model as appropriate for studying hypoxia onset and progression.

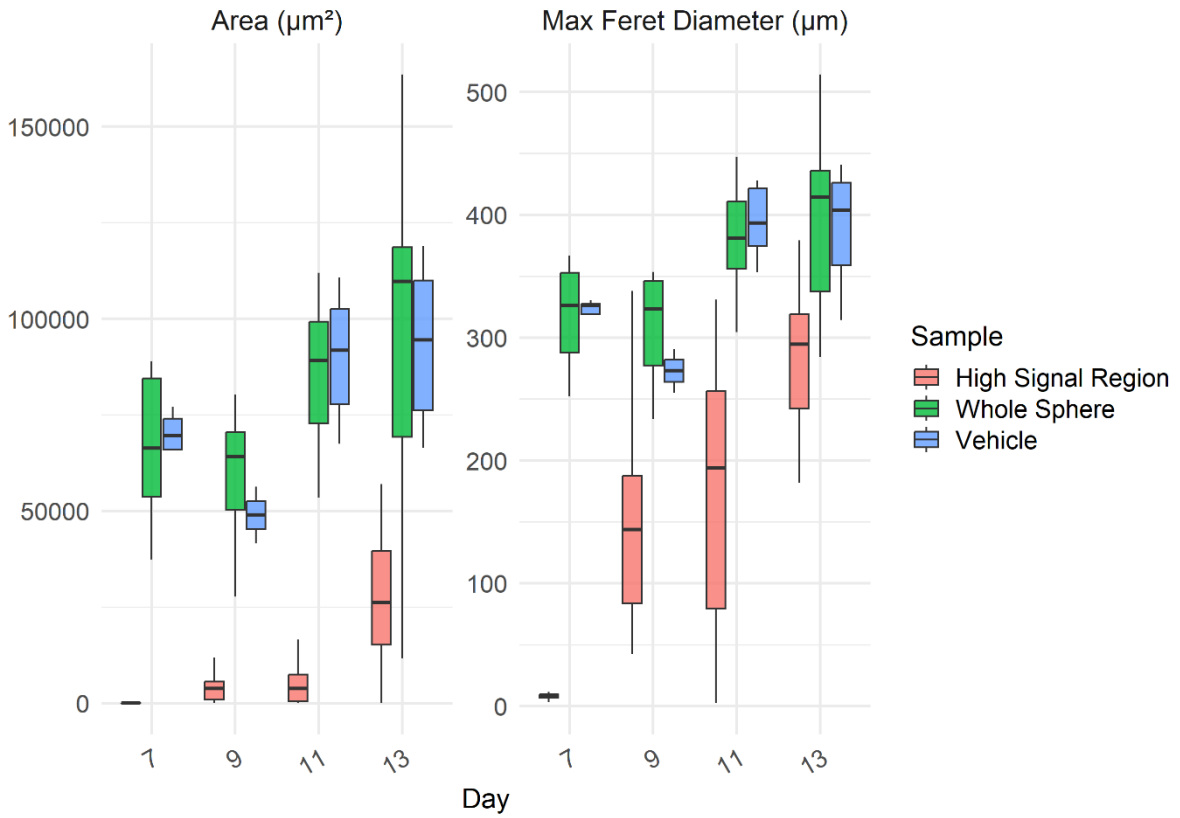


Figure 5.7: The area and diameter of high signal regions of spheres progressively increase between days 7 and 13, with high signal regions spanning large areas of spheroids at day 13.

EF5 images were processed with CellProfiler to isolate high signal regions to compare the size of these regions to the rest of the spheroid. The area of high signal regions is small on day 7, with only a few square micrometres identified. Over time, the area of the high signal region grows to up to ~50000 µm^2 by day 13, forming up to a third of the area of the biggest whole spheroid. The diameter of high signal regions increases at a much faster rate than the area, as the median diameter on day 9 and 11 is almost half the median diameter of whole spheres. By Day 13, the median diameter of high signal regions can make up to three quarters of the diameter of whole sphere regions, making up a considerable proportion of the entire spheroid. Diameter values are reported as Feret diameters, defined as the longest axis across each region.

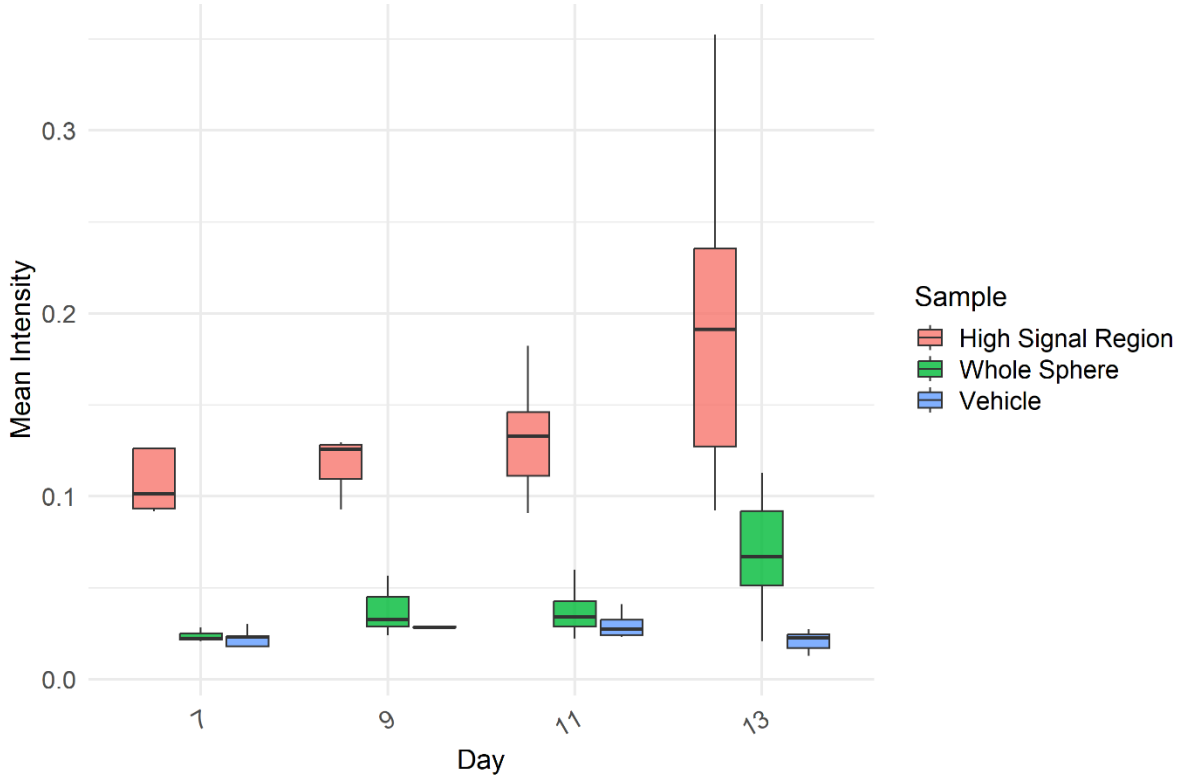


Figure 5.8: **Signal intensity within high signal regions increases from day 7 to 13, with some spheres showing three times the mean intensity on day 13 compared to day 7**

The mean intensity of high signal regions was displayed alongside the whole sphere and the vehicle-treated spheres' mean intensity. The high signal intensity is never lower than 0.9 as this was the threshold value to select high signal regions. Day 13 is the first time that the mean intensity for the whole sphere region crosses over 0.9, with the upper 25% of whole sphere regions crossing the high signal cut-off. The high signal region on day 13 shows up to triple the mean intensity of fluorescence compared to the whole sphere regions, with mean intensity reaching up to 0.35

5.2.6. ERα protein expression is enriched at the spheroid surface and heterogeneous within hypoxic regions

Sequential MCF7 spheroid cryosections were stained with an Anti-ERα antibody to assess ERα protein expression distribution relative to EF5 staining. Representative images day 13 spheroids are shown Appendix Figure 8.19).

ERα protein expression was concentrated in the nuclei of cells at the surface of the MCF7 spheroids. In addition, isolated ERα-positive nuclei were observed sporadically within deeper layers of the spheroid, including the inner surface and outer core regions (Figure 5.9, arrows). These correspond to bins 2-5 within CellProfiler analysis pipelines. When comparing the depth

of these sporadic ERα-positive cells to sequential EF5 staining, some were located at depths consistent with the outer layers of hypoxic regions.

To test whether ERα expression varied systematically with depth, we quantified ERα fluorescence intensity across radial bins using a similar, adapted CellProfiler pipeline applied to EF5 (Appendix Figure 8.21). Although ERα intensity tended to be visibly higher at the surface and lower in the core, this pattern was not statistically significant, as the general pattern was observed in secondary-antibody-only controls, indicating that more work with improved staining is needed to determine the relationship between ERα expression and hypoxia. These results suggest that heterogeneity of ERα protein expression within spheroids, with a trend towards anti-correlation with hypoxia that requires further investigation. However, this trend would be consistent with the literature, which suggests that ERα is downregulated during hypoxia (Ryu, Park and Lee, 2011; Stoner et al., 2002; Cho et al., 2005; Whitman et al., 2019).

These findings show that ERα protein expression is enriched at the spheroid surface but can also be detected sporadically within hypoxic regions. While quantitative analysis did not reveal significant differences across radial bins, the observed heterogeneity highlights the potential for future studies using calling card reagents to probe how hypoxia influences ERα binding and gene regulation in MCF7 spheroids.

Future studies could address this spatial heterogeneity using spatial transcriptomics approaches. Technologies such as Visium or other spatial RNA-sequencing platforms would enable gene expression patterns to be mapped across spheroid sections while preserving spatial organisation. This would allow ERα target gene expression and hypoxia-responsive transcriptional signatures to be quantified across the radial gradient of the spheroid, providing a clearer assessment of how hypoxia influences ERα activity. Combining spatial transcriptomics with ERα calling card approaches could further link spatial variation in transcription factor binding with local transcriptional outputs.

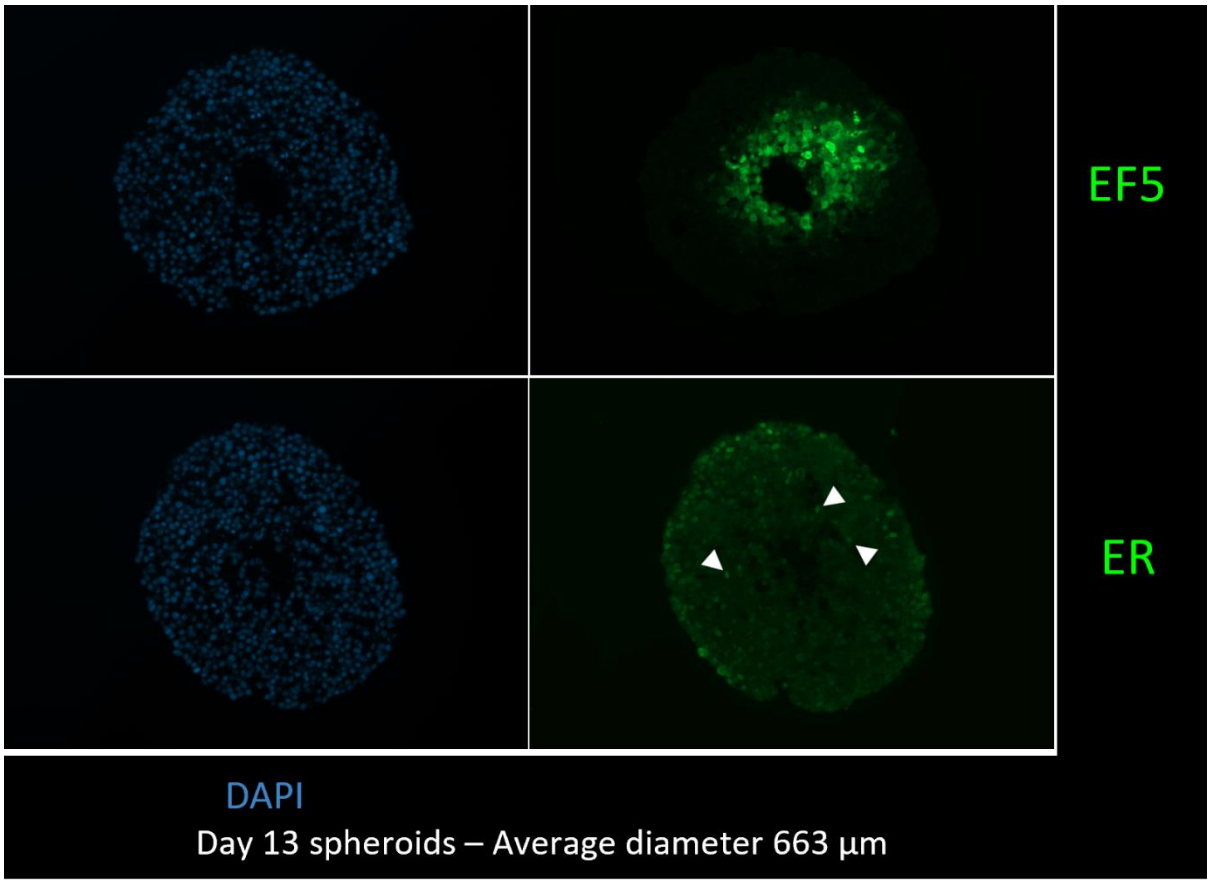


Figure 5.9: ERα fluorescent signal is mainly localised to the nuclei of cells at the surface of spheroids, with sporadic levels of signal within the hypoxic region of MCF7 spheroids

Sequential cryosections of MCF7 spheroids stained with either Anti-EF5 or Anti-ERα antibodies reveal the relationship between hypoxia and ERα expression. ERα signal appears mainly localised in the nuclei of cells at the surface of spheroids, in this image, cells at the top of the sphere show the highest levels of signal. While there are several cells deeper into the sphere that show a fluorescent signal, most of the cells in the core of spheroids do not show high levels of ERα fluorescent staining. However, there are some cells with a fluorescent signal at the same depth where hypoxic signal was detected by EF5 (indicated with white arrow), suggesting that some ERα-expressing cells in MCF7 spheroids could be affected by hypoxia..

5.2.7. Treating spheres with 100 μM of the EnAC inhibitor amiloride impairs spheroid growth and prevents hypoxic core development.

To test our MCF7 spheroid model's capability to explore how hypoxia affects breast cancer, we designed a growth assay that exposes spheroids to Amiloride, an inhibitor of EnAC, a protein whose subunits are upregulated by ER during hypoxia (Malcolm et al., 2025).

A recent publication from our lab utilises a complex RNA-seq panel to explore the relationship between ERα and hypoxia, in which we reveal an ERα-Dependent Hypoxic Response (EDHR) (Malcolm et al., 2025). Our EDHR signature shows that during hypoxia, the regulatory β-and-γ-subunits of the EnAC are upregulated by ERα. The expression of these subunits under hypoxia is interesting in the context of the literature, as other publications have focused on the α-subunit of the EnAC and found it to negatively affect cell proliferation (Ware et al., 2021; McQueen et al., 2025). In contrast, β-and-γ-subunit expression has been associated with poorer incidence-free survival, highlighting the need to understand the cause of this inconsistency in the clinical outcomes.

We incubated two plates of spheroids with a range of amiloride concentrations across the range of EnAC's IC50, which is reported to be ~90 nM (Lu et al., 2008; Noreng et al., 2020; Teiwes and Toto, 2007). One plate was incubated on day 2 to represent treatment during tumour formation, hereby referred to as 'early' treatment, and the other was incubated on day 9 to represent treatment of established, hypoxic tumours, also called 'late' treatment.

Late treatment spheroids treated with 100 μM Amiloride immediately respond to Amiloride, with spheroid growth being reduced for two days, before increasing at a slower rate than other conditions tested (Figure 5.10). Spheres treated with <100 μM grow in diameter at a similar rate across all the days.

Late 100 μM-treated spheres are the only group that has a statistically significant difference in diameter from the vehicle control. Although some 1 μM-treated spheres overlapped in size with 100 μM-treated spheres, overall the groups remained significantly different.

Early treatment spheroids showed the same impaired growth at 100 μM, failing to form mature spheroids (Figure 5.11; Appendix Figure 8.22). Like late treatment spheroids, only the 100 μM Amiloride-treated spheres show significant differences in their diameter over the course of 13 days.

In contrast, early treatment spheroids exposed to 10 µM Amiloride showed the smallest diameter in conditions below 100 µM. On days 12 and 13, there appears to be a pattern in which spheroids treated with the lower amiloride concentrations (10 nM – 10 µM) had slightly larger average diameter than vehicle controls, with average diameter decreasing as amiloride concentration increased. Although not statistically significant, this trend is consistent with reports that EnAC inhibition increases MCF7 proliferation (Ware et al., 2021).

These results show that targeting EnAC in MCF7 spheroids with amiloride can impair spheroid formation and proliferation at >100 µM. However, because 100 µM amiloride is substantially higher than the reported IC50, effects at this concentration may reflect broader perturbations of ion transport and cell physiology rather than EnAC-specific inhibition alone. This limitation should be considered when interpreting growth impairment at high doses.

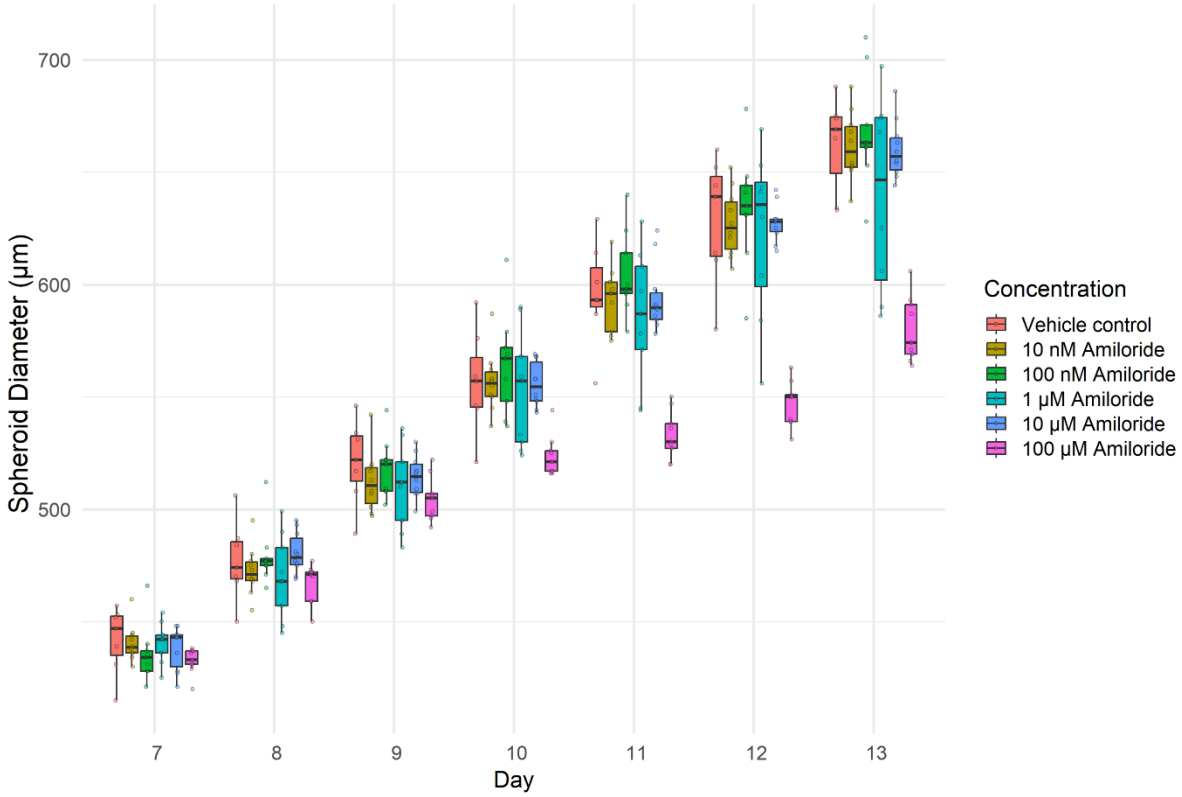


Figure 5.10: Treatment of MCF7 spheroids with 100 µM amiloride on day 9 inhibits growth and reduces spheroid diameter compared to other conditions.

Spheroids were treated on day 9 with a range of amiloride concentrations to assess effects on established cultures. Spheroids treated with <100 µM did not differ significantly from vehicle controls. Spheroids treated with 100 µM were immediately affected, showing reduced growth between days 9-10. From day 10 onwards, 100 µM-treated spheroids were 40-80 µm smaller than other conditions and significantly different in diameter (Tukey test, $p < 0.01$). From day 11 onwards, differences remained highly significant ($p < 0.0001$).

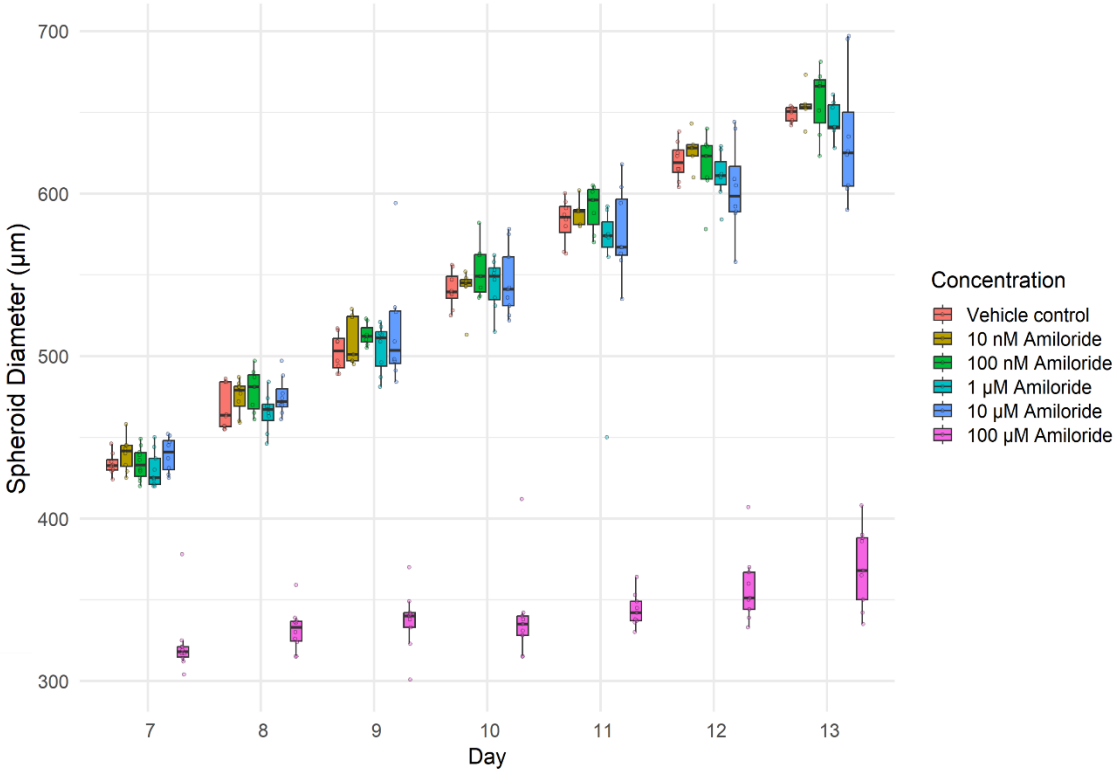


Figure 5.11: Early treatment of MCF7 spheroids with amiloride impairs spheroid formation at 100 µM.

Early treatment of MCF7 spheroids with amiloride impairs spheroid formation at concentrations of 100 µM. Spheroids were treated on day 2 of culture with a range of amiloride concentrations to assess effects on spheroid formation, mimicking treatment during tumour initiation. Spheroids treated with 100 µM amiloride were immediately affected, with growth arrested by day 3 and diameters increasing at a slower rate than in other conditions. Impaired growth persisted to day 13, with increased variance in diameter due to irregular spheroid morphology. Up to day 10, spheroids treated with concentrations <100 µM showed similar growth to vehicle controls. From day 11 onwards, the median diameters of nanomolar treatments remained slightly above vehicle controls, whereas 1 µM and 10 µM treatments fell below the vehicle median. These differences were not statistically significant. In contrast, spheroids treated with 100 µM amiloride were consistently smaller than all other conditions, with significant differences in diameter from day 7 onwards (Tukey test, $p < 0.0001$).

5.2.8. MCF7 spheroids grown in 100 µM amiloride exhibit irregular surface structure and internal branching, preventing proper spheroid formation and hypoxic core development

Spheroids treated with 100 µM amiloride on day 2 of culture were incubated with EF5 before collection and fixation for immunostaining to examine the internal structure of misshapen early treatment spheroids.

MCF7 spheroids treated with 100 μM have an irregular surface and structure at day 7; by day 13, the spheroids begin to disintegrate by shedding cells, although some budding structures retained partial organisation (Figure 5.12). Whereas DMSO-treated spheroids remained spherical with smooth surfaces that expanded consistently over time, amiloride-treated spheroids developed bumps and dents in their surface. However, such structural disruption could reflect the nonspecific effects of high-dose amiloride on cell adhesion and viability, which limits the extent to which these phenotypes can be attributed solely to EnAC inhibition. As they grew past day 7, cells shed into the well and accumulated at the bottom, while others formed rounded buds on the surface. These spheroids progressively lost their structure and further deviated from the typical compact morphology.

Cryosectioning and immunostaining revealed an irregular internal organisation, with spheroids failing to reach the size or organisation required to form a hypoxic core after 13 days (Figure 5.13). Most of the cryosections retained a large, connected mass of cells; however, a minority also contained tube-like clusters of cells that were separated from the main body. These tube-like clusters may partly reflect sectioning artefacts, but their consistent presence in treated spheroids and absence in controls suggests they are the result of abnormal growth patterns, with cells extending outward rather than forming a compact spheroid. Notably, none of the amiloride-treated cells were positive for EF5, consistent with the observation that these spheroids do not reach the depth or volume required for hypoxia. Although we do not infer a mechanism here, prior work associates increased expression of the regulatory β-and-γ-subunits of the ENaC with worse incidence-free survival (Malcolm et al., 2025), which we consider further in the Discussion in the context of amiloride's growth-restrictive effects in MCF7 cells.

Overall, we show that 100 μM amiloride-treated spheres cannot form spheroids with regular, compact morphology and hypoxic core development observed in control spheroids after 13 days in culture.

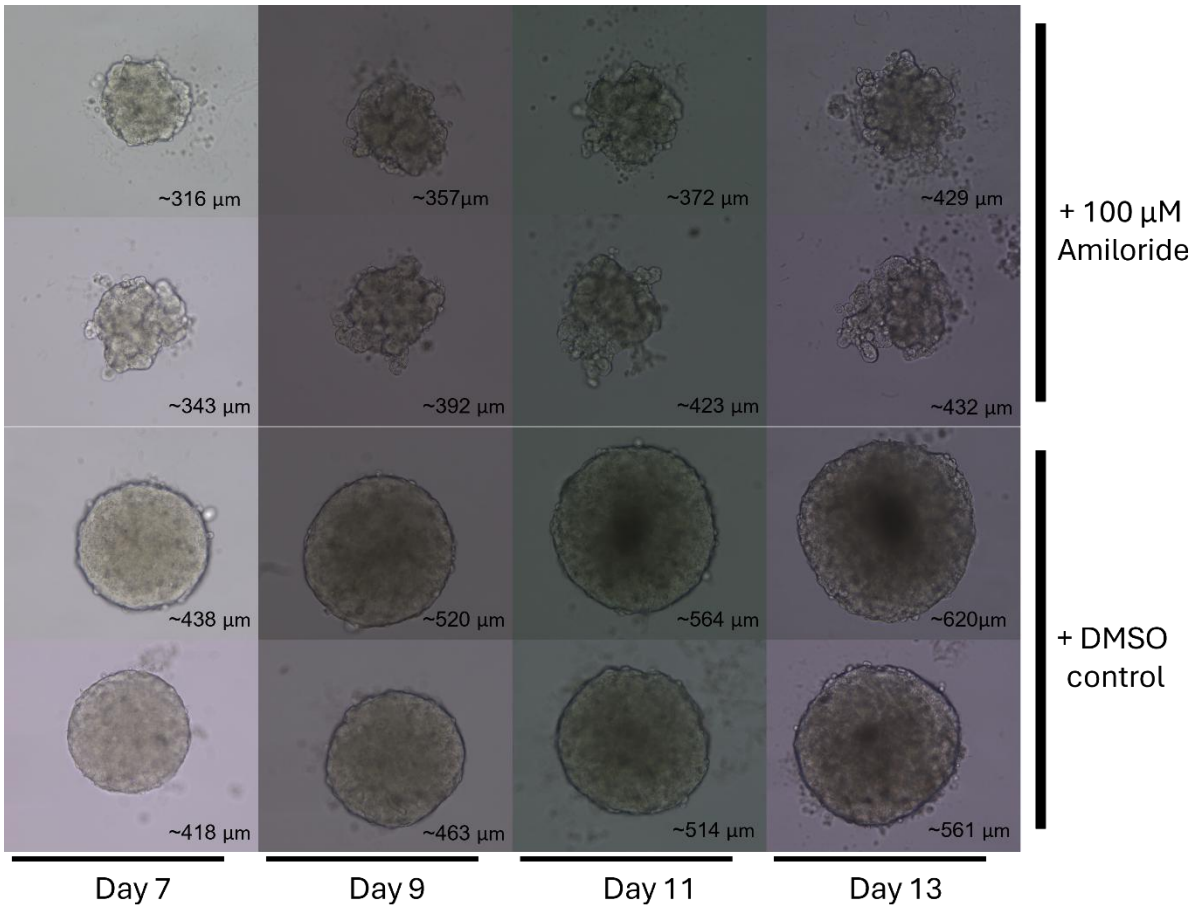


Figure 5.12: Early treatment with 100 µM amiloride disrupts spheroid surface morphology and promotes cell shedding.

Spheroids treated with 100 µM amiloride grew more slowly than DMSO controls, with irregular surfaces evident by day 7. By day 13, treated spheroids showed bumps, dents, and budding structures, with cells shedding into the well. Growth continued in diameter but as irregular buds and extensions rather than a compact spheroid mass.

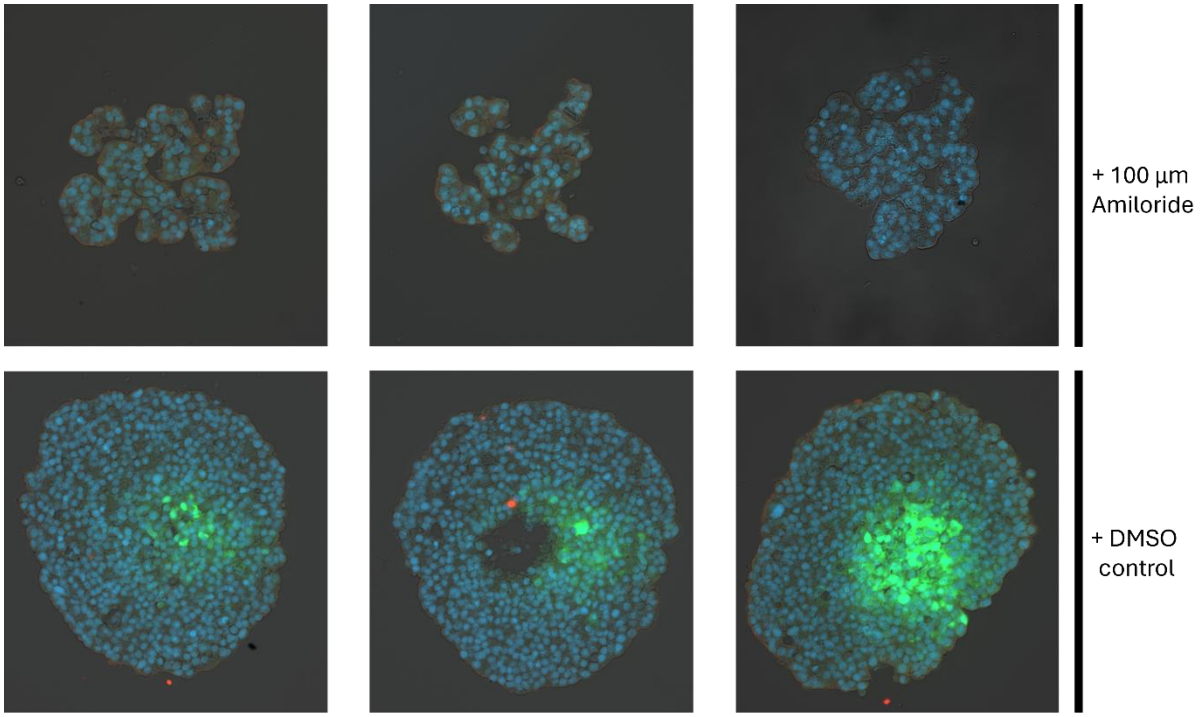


Figure 5.13: Cryosectioning reveals irregular internal organisation and absence of hypoxic cores in 100 µM amiloride-treated spheroids.

Early treatment spheroids were cryosectioned and stained with anti-EF5 and anti-CAIX conjugated antibodies to assess the development of hypoxia in spheres treated with 100 µM amiloride. This sectioning revealed the irregular internal structure of spheres treated with 100 µM amiloride. Spheres create tube-like structures rather than forming a single mass of cells. While there are more globular regions as are observed in the left and right images, cells are still separated into irregularly shaped, disconnected structures. Furthermore, no spheres treated with 100 µM amiloride could exhibit EF5 signal, unlike the DMSO control, indicating treated spheres are incapable of forming hypoxic cores in 13 days.

5.3. Discussion

5.3.1. Summary

In this section, we established a reliable MCF7 spheroid model that can generate hypoxia in the core of spheres due to the diameter impeding oxygen diffusion to be used in conjunction with ERα calling cards in future experiments.

Previous studies reported that MCF7 spheroids develop hypoxia once they reach ~500-600 μm in diameter, with necrotic core formation at later stages (Yakavets et al., 2020; Singh et al., 2016; Palma et al., 2016). Here, we extend and characterise these observations by providing an EF5 time course, quantitative analysis of hypoxic region size and intensity, and evidence that the ERα is expressed within hypoxic regions.

By confirming the ERα expression in hypoxic cells, we have validated our model for future use with long-read, single-cell ERα calling cards to probe whether ERα binding heterogeneity is altered under hypoxia.

We further show that amiloride treatment impairs spheroid growth at 100 μM, highlighting ion channel activity as a contributor to spheroid integrity.

Together, these findings establish a physiologically relevant model for studying hypoxia-dependent ERα activity.

5.3.2. We establish a reliable MCF7 spheroid hypoxia model as a platform for the future study of ERα in low-oxygen environments.

Our EF5 time-course confirmed reproducible hypoxia onset at 500 μm in diameter, consistent with previously published models (Yakavets et al., 2020; Singh et al., 2016; Palma et al., 2016). Establishing controlled spheroid growth before hypoxia onset is a key step in creating a physiologically relevant model to study ERα in low oxygen conditions. MCF7 spheroids have been used to probe the effects of hypoxia before; prior studies show that a necrotic core was observed in spheroids at 500 μm in diameter (Yakavets et al., 2020; Palma et al., 2016) and hypoxia was detected by Ru-dpp fluorescence at 600 μm in diameter (Singh et al., 2016). Quantitative analysis of our spheroids showed hypoxic regions expanding to three-quarters of the total spheroid diameter by day 13, coinciding with necrotic core formation. Both the area and Feret diameter of EF5-positive regions increased progressively, with day 13 spheroids showing hypoxia across most of the core.

Extensive optimisation confirmed a seeding density of 1000 cells/well or less in standard media produced spheroids within 400 to 600 μm in diameter, avoiding premature hypoxia and enabling reproducible baseline formation (Figure 5.3). Consistency in our model ensures that any observations in future experiments reflect experimental variables rather than inconsistency driven by heterogeneous oxygen gradients.

Optimisations on timepoints in this assay's construction were informed by a detailed investigation into the structural development of MCF7 spheroids, which revealed that complex cell-to-cell connections are only apparent after 5 to 7 days in culture (Pulze et al., 2020). These benchmarks guided our strategy to ensure hypoxia does not develop until spheroids have established a stable structure to minimise downstream confounding variables in future assays.

Our EF5 immunostaining time course (Figure 5.4) validated our model's hypoxia onset ~500 µm in diameter, consistent with previously published models (Yakavets et al., 2020; Singh et al., 2016; Palma et al., 2016). Quantitative analysis of CellProfiler pipelines reveals a negative correlation between EF5 mean fluorescence intensity ("MeanFrac") per radial bin and bin size (Figure 5.5 and Figure 5.6), indicating that hypoxia develops in the core and expands as the spheroid grows over time. Consequently, from day 11, spheroids >600 µm in diameter show extensive anti-EF5 signal co-localised with cell structure deterioration and necrotic core formation, confirming that limited oxygen drives structural deterioration over time. Formation of the necrotic core at these stages is also consistent with previous research (Singh et al., 2016), establishing our model as a reliable system to explore hypoxia-dependent processes in breast cancer within a 3D context.

We observe occasional asymmetry in the anti-EF5 signal intensity in spheroid cores that could represent a reduced oxygen gradient at the plate interface, but this does not alter the interpretation of our model. Quantitative analysis of EF5 staining demonstrated that hypoxia could extend to three-quarters of the total spheroid diameter (Figure 5.7). While this may be inflated due to artificial restrictions, due to the plate interface impeding oxygen diffusion, this could be improved in future experiments by suspending spheroids in a matrix or using microfluidic devices to ensure even oxygen diffusion (Mihara et al., 2019; Grist et al., 2019; Pyne et al., 2024; Kamoya et al., 2016; Tevlek et al., 2023). Other technical factors can contribute to the development of hypoxia that are not biologically relevant, such as the depth of the cell culture media in wells (Al-Ani et al., 2018; Hass et al., 2023). Media depth is a particular concern given the inconsistency of evaporation of wells within cell culture plates (Mansoury et al., 2021; Grosch et al., 2016; Das et al., 2016; Maxwell et al., 2023), resulting in inconsistent oxygen availability for spheroids across the plate. Over the course of 13 days, we did observe varying well volumes due to inconsistent evaporation, which likely contributed to the edge effects observed in the outer wells (see methods Section 2.6).

Across each day, variability in the EF5 signal was observed between individual spheroids, reflecting differences in sectioning depth and imaging. While we aimed to always section the core of spheres, it was not always technically possible. Sequential slices were taken to gather

as much contextual information as possible, and in some cases, peripheral regions were included that showed reduced EF5 signal compared to more central slices. It is therefore possible that some of the spheres at day 9, 11 and 13 that did not show a change in mean intensity with bin depth were the result of less central core sampling.

Taken together, these data demonstrate that we have provided a robust hypoxic spheroid model for investigating ERα binding and activity in low oxygen conditions. To further validate our model as a platform for future ERα calling cards to map binding heterogeneity in hypoxic conditions, we went on to examine ERα expression throughout our spheroids.

5.3.3. ERα fluorescent signal co-localises with EF5 in sequential cryosections, enabling future application with ERα calling cards

ERα expression in cells exhibiting hypoxia is an essential requirement for the application of calling card methods to our MCF7 spheroid hypoxia model. The literature surrounding ERα distribution in 3D cultures is inconsistent; some studies suggest ERα expression is restricted to the surface of spheroids (Muñoz et al., 2010; Kapara et al., 2021; Whitman et al., 2019), whereas others report ERα is not required in 3D MCF7 culture (Chen, Liu and Yan, 2022), as ERα+ cells can activate alternative survival pathways under hypoxia.

We demonstrate that the anti-ERα fluorescent signal is localised predominantly at the surface of MCF7 spheroids, but is additionally detectable in inner layers, including regions exhibiting hypoxia (Figure 5.9). This expression pattern suggests that ERα can persist in low oxygen conditions in our model. Although ERα expression is reduced under hypoxia (Ryu, Park and Lee, 2011; Stoner et al., 2002; Cho et al., 2005; Whitman et al., 2019), we show that ERα expression remains detectable in hypoxic cells, consistent with the EDHR identified in our previous publication (Malcolm et al., 2025). This parallel supports the relevance of our model in chromatin-binding studies, as it indicates that ERα may engage distinct regulatory elements under hypoxia compared to normoxia. This model, therefore, represents a platform for testing whether hypoxia modifies ERα binding patterns; a key question for understanding hormone receptor signalling in low-oxygen environments.

Although the anti-ERα signal was lower than EF5, this is consistent with the reduced ERα expression in 3D culture (Muñoz et al., 2010) and technical factors inherent to the AF488 channel, such as autofluorescence, which is common in biological tissues (Shah et al., 2017; Morita et al., 2019; Gooz and Maldonado, 2023; Ostrander et al., 2010). The choice of secondary antibody and channel, therefore, introduced background that reduced contrast, but the clear nuclear localisation of ERα signal and its absence in secondary-only controls confirm

that the staining represents true ERα protein expression. These technical considerations explain the lower signal-to-noise ratio, undermining the conclusion that ERα is expressed within hypoxic regions of MCF7 spheroids.

Taken together, these analyses show that the ERα protein expression is enriched at the spheroid surface but can also be detected within hypoxic regions located approximately 100-200 μm from the spheroid surface, overlapping with EF5-positive zones. Qualitatively, this pattern overlaps with EF5-positive zones, supporting the conclusion that ERα is downregulated but persists in hypoxia. However, quantitative radial intensity analysis did not show a statistically significant enrichment of ERα signal above background autofluorescence (Appendix Figure 8.21), reflecting the modest contrast achieved with the AF488 secondary antibody. This highlights the need for improved resolution between the true nuclear signal and background noise in future experiments. Despite these technical constraints, the consistent nuclear localisation that was absent in secondary-only controls confirms that the observed staining represents genuine ERα protein expression, validating our model for future calling card studies to probe ERα binding heterogeneity under hypoxia.

5.3.4. The application of our hypoxic MCF7 model to study hypoxia-driven ERα transcriptional targets reveals that high-dose amiloride treatment impairs spheroid formation and growth.

Amiloride treatment reveals how the activity of an ion channel specifically upregulated by ERα under hypoxic influences spheroid growth and development in our model. We focus on a set of genes specifically upregulated by ERα during hypoxia, and not normoxia, that were clinically relevant: SCNN1B and SCNN1G (Malcolm et al., 2025). This approach enabled the assessment of how arresting the function of an ion channel expressed under hypoxia alters spheroid architecture. All analyses excluded spheroids from outer wells to avoid plate-related edge artefacts; see Methods for details.

High-dose amiloride (100 μM) significantly affected spheroid growth (Figure 5.11 and Figure 5.12) and prevented spheroids from reaching the size and organisation required for hypoxic core formation (Figure 5.13), indicating that ion channel activity contributes to spheroid structural integrity. The reported IC50 for EnAC is between 80 to 100 nM (Lu et al., 2008; Noreng et al., 2020; Teiwes and Toto, 2007), reflecting inhibition of electrophysiological activity in the channel, whereas 100 μM has been proposed in prior studies as a pharmacologically relevant concentration in cellular assays (Mukherjee et al., 2020; Sermet-Gaudelus et al., 2007; Giorgetti et al., 2021; Baquero and Gilbertson, 2011; Yu et al., 2012). The high-dose effect is likely driven by combined inhibition of other ion channel targets such as NHE1,

consistent with broader amiloride activity at higher concentrations (>1 μM) (Teiwes and Toto, 2007).

Although day 13 early treatment of spheroids with EnAC's IC50 concentration (100 nM) appeared to show increased spheroid growth that negatively correlated with amiloride concentration (Figure 5.12), this pattern was only observed when edge wells were included. Because edge wells exhibited position-dependent artefacts (evaporation and altered oxygen diffusion) and were excluded from all final analyses, we did not perform statistical comparisons based on those wells. The apparent increase, therefore, should be interpreted cautiously as a plate position artefact rather than a biological response. A dedicated, position-controlled experiment could be designed to test whether low-dose ENaC inhibition modulates spheroid formation under rigorously uniform oxygen diffusion and media depth (e.g., central-well layouts, matrix suspension or microfluidics), but this was beyond the scope of the present study.

Taken together, our findings show that high-dose amiloride (100 μM) prevents MCF7 spheroids from reaching the size and organisation required to form hypoxic cores, highlighting the importance of ion channel activity for spheroid integrity. This concentration has been used in prior studies as a pharmacologically relevant dose, and our results extend beyond that work by demonstrating its impact in a 3D hypoxic model. Importantly, ENaC biology is closely linked to HIF signalling in epithelial tissues (Dizin et al., 2021; Baloglu et al., 2020; Zhou et al., 2023; Nag and Resnick, 2017), and both β-and-γ-subunit expression have been associated with worse incidence-free survival in breast cancer (Malcolm et al., 2025). By characterising the development of hypoxia in spheroids, our model provides a framework to explore how ENaC activity contributes to ERα-driven growth under oxygen stress. This integration is particularly relevant given the conflicting reports that α-ENaC is proposed as a therapeutic agent (McQueen et al., 2025; Ware et al., 2021; Evans and Sloan-Stakleff, 2000; Leon et al., 2013; Berg et al., 2022), but the regulatory subunits have also been reported to indicate worse recurrence-free survival (Malcolm et al., 2025). Our data suggest that the impact of amiloride is context dependent: while high doses disrupt spheroid growth and prevent hypoxia, low-dose treatment near the ENaC IC50 may promote proliferation, in contrast to published findings (McQueen et al., 2025; Ware et al., 2021). In this way, our model presents as a useful model to explore these disparities in an accessible and reproducible system to reconcile these contradictions to test whether hypoxia modifies the impact of ion channel inhibition on ERα-driven growth.

5.3.5. Future work

The work in this chapter establishes a reproducible hypoxic spheroid model, demonstrates ER localisation within hypoxic regions, and explores the impact of amiloride treatment on spheroid growth. We demonstrate this model's future applicability to mechanistic studies of hormone-driven adaptation to oxygen stress; each of these chapters' findings provides a foundation for further development.

5.3.5.1. Expanding model systems and microenvironmental readouts

Future studies should extend beyond MCF7 spheroids to validate findings in additional ERα+ breast cancer lines such as T47D, as well as in ERα-low models to determine whether hypoxia-driven ERα activity is unique to luminal phenotypes. Incorporating patient-derived 3D cultures will strengthen translational relevance once mechanistic clarity is achieved in cell line models.

In parallel, the role of the microenvironment warrants closer examination. Measuring extracellular pH, lactate accumulation, and ion flux will clarify whether amiloride's effects are mediated indirectly through metabolic or ionic shifts that feed back on HIF signalling. Such assays would provide a more integrated view of how ion channel activity intersects with hypoxia biology and ERα function.

5.3.5.2. Refinement of the spheroid hypoxia model

Our EF5 time course confirmed reproducible hypoxia onset at ~500 μm diameter, consistent with published models, and future work should refine this system by improving oxygen diffusion and minimising artefacts caused by plate position and evaporation. Approaches such as suspending spheroids in matrices or using microfluidic devices could ensure more uniform oxygen gradients. Further validation with additional hypoxia markers (e.g. pimonidazole, HIF-1α) would strengthen confidence in hypoxia onset and progression. Establishing reproducible thresholds for hypoxic core formation across replicates will provide a robust baseline for downstream mechanistic studies.

5.3.5.3. ERα correlation under hypoxia

We showed that ERα is detectable in hypoxic regions, overlapping with EF5-positive zones, although quantitative enrichment above background autofluorescence was modest. Future experiments should improve signal resolution through alternative fluorophores, enabling statistical confirmation of ERα enrichment. Application of ERα calling cards in spheroids will

allow mapping of ERα binding heterogeneity across oxygen gradients, providing insight into how hypoxia alters ERα transcriptional targets.

5.3.5.4. Studying ERα activity during hypoxia with calling cards

Combining ERα calling cards with other 3D culture methods offers a powerful way to extend these findings. Calling cards provide a longitudinal record of transcription factor binding, and when applied to spheroid culture or patient-derived xenografts (PDX), they could capture the binding heterogeneity of ERα in therapy-resistant and sensitive breast cancer cells over a period of time. Because hypoxia develops naturally once spheroids reach ~500-600 µm in diameter, (Palma et al., 2016; Muñoz-Galindo et al., 2019; Rofstad et al., 1996; Mukomoto et al., 2020; Yakavets et al., 2020; Singh et al., 2016), transfection or transduction with calling card reagents before hypoxia onset would allow ER binding events to be marked as oxygen gradients emerge. This approach would enable single-cell resolution mapping of how ERα binding changes in response to hypoxia and other stimuli, directly addressing the question of whether ERα engages distinct regulatory elements under impaired oxygen conditions.

5.3.5.5. Amiloride treatment in spheroids

High-dose amiloride (100 µM) treatment impaired spheroid growth and prevented hypoxic core formation, highlighting the importance of ion channel activity for spheroid integrity. However, the biological significance of this effect is complicated by off-target inhibition at high concentrations. Future work should therefore include a full dose-response analysis to distinguish specific ENaC inhibition from broader ion channel effects. Genetic approaches such as ENaC subunit knockdown or rescue experiments will help determine whether the observed phenotype is attributable to ENaC inhibition or other targets such as NHE1. Parallel culture under normoxic and hypoxic conditions will clarify whether amiloride's effects are modified by oxygen availability, addressing conflicting reports of increased viability in MCF7 cells. Beyond spheroid size, future studies should quantify structural integrity, apoptosis, and proliferation markers to provide a more comprehensive assessment of amiloride's impact.

Together, these avenues of future work will consolidate the robustness of the spheroid model, clarify the mechanistic relationship between ER and HIF under hypoxia, and resolve the context-dependent effects of amiloride, while expanding the relevance of this system to diverse breast cancer models and microenvironmental conditions.

6. General discussion

6.1. Recap of thesis aims

The research described in this thesis aimed to engineer and validate a single-cell platform to study ERα regulation of the genome within the breast cancer microenvironment by developing an ERα calling card reporter. This platform was designed to be applied to a characterised 3D spheroid model, enabling the study of how microenvironmental cues shape ERα activity and the contribution of binding to endocrine therapy resistance.

The motivation for this work arose from the literature surrounding ERα endocrine therapy resistance, which consistently highlights the heterogeneity of ERα activity and binding across the genome as a driver of tumorigenic pathways (Generali et al., 2006; Yan et al., 2018; Ryu, Park and Lee, 2011; Heddleston et al., 2010; Padró et al., 2017; Vaupel, 2004; Muz et al., 2015; Semenza, 2012; Wang et al., 1995). Breast cancer cells do not exist in isolation, but within a dynamic tumour microenvironment that interacts with and modulates cancer progression (Hanahan and Weinberg, 2011; Quail and Joyce, 2013; Mittal, Brown and Holen, 2018; Li, Tsang and Tse, 2021; Denkert et al., 2018). Microenvironmental cues, including nutrient availability, stromal interactions and oxygen tension, exert profound influences on gene regulation. Among these, hypoxia is a particularly potent stimulus, inducing both spatial and temporal heterogeneity (Yan et al., 2018; Heddleston et al., 2010; Yun and Lin, 2014; Chen et al., 2018). This results in diverse cellular states developing across patient tumours, with populations of cells responding to hypoxia by adopting more pro-tumorigenic programmes than others (Azizi et al., 2018; Yang et al., 2023). Responding to these cues involves epigenomic and transcriptomic changes that promote tumour progression (Batie and Rocha, 2019; Thienpont et al., 2016; Di Giovannantonio et al., 2025; Batie et al., 2019).

On this basis, we identified hypoxia as the first microenvironmental cue to target, given its well-established role in driving heterogeneity and its clinical relevance in breast cancer. A platform capable of measuring the effects of hypoxia on a single-cell profile of ERα binding would therefore be a valuable tool to characterise the initiation and progression of endocrine therapy resistance over a period of time.

To achieve this, we reviewed available methods. We identified scCC as a promising approach, as it provides the ability to record TF binding events at single-cell resolution while simultaneously capturing the transcriptomic state through RNA-Seq. This dual function allows the binding profile of ERα to be directly connected to the microenvironmental state of the same

cell, with hypoxia-responsive genes serving as proxies for oxygen status (Malcolm et al., 2025). We were therefore motivated to develop a platform for studying the genomic regulation coordinated by ERα, and to establish a reproducible breast cancer spheroid culture system in which hypoxia develops naturally as spheroids reach critical diameters (Yakavets et al., 2020; Singh et al., 2016).

Specifically, the relationship between ERα and hypoxia is of particular interest for the application of our calling card platform. While HIF-1α upregulation in response to hypoxia has been associated with reduced ERα protein levels, ERα remains transcriptionally active under these conditions (Malcolm et al., 2025; Padró et al., 2017; Ryu, Park and Lee, 2011; Jehanno et al., 2022; Yang et al., 2015). Rather than a simple loss of function, ERα and hypoxia signalling pathways appear to coordinate gene regulation, sometimes co-occupying regulatory elements or influencing each other's gene targets (Yang et al., 2015; Ryu, Park and Lee, 2011; Generali et al., 2006; Padró et al., 2017; Min et al., 2022; Cho et al., 2005). For example, coordinated hypoxic responses have been described to promote the regulation of amino acid transporters, like SNAT2, instead of ERα, resulting in increased survival following endocrine therapy, contributing to resistance (Morotti et al., 2019). These expression patterns can then be reinforced by epigenetic remodelling that further supports lasting changes to gene regulatory programmes (Yang et al., 2015; Morotti et al., 2019). More recently, our lab has described an ERα-dependent hypoxic response (EDHR), associated with transcriptional changes linked to poor patient outcomes and has enriched adjacent to EMT regions of patient tumours (Malcolm et al., 2025). Other studies have shown that prolonged exposure to hypoxia mimetics ($CoCl_2$) can result in HIF compensating for reduced ERα activity, further underscoring the dynamic interplay between these pathways (Jehanno et al., 2022; Canhasi, Tina and Eremo, 2023).

These findings highlight the importance of understanding ERα regulation within the oxygen-restricted microenvironment of breast cancer. Hypoxia-driven ERα heterogeneity in breast cancer requires experimental solutions that faithfully recapitulate the behaviour of patient tumours while remaining reproducible and amenable to genetic manipulation. While many 3D cell culture methods exist, spheroid culture offers the lowest technical complexity and is widely accessible, making it an ideal system for mechanistic studies (Däster et al., 2017; Ham et al., 2016; Raza et al., 2017; Yang et al., 2025; Sachs et al., 2018).

In summary, this thesis addresses a critical mechanistic gap by developing a single-cell ERα calling card platform and an accessible hypoxic MCF7 spheroid model. Together, these tools provide a reproducible framework to study how microenvironmental cues, specifically hypoxia,

shape ERα binding heterogeneity and contribute to the progression of endocrine therapy resistance.

6.2. Summary of main findings

Platform Validation

- Both N- and C-terminal HyPB-ERα fusion proteins can cleave SRTs and insert them into the genome of co-transfected cells (Figure 3.4).
- Only the C-terminal HyPB-ERα fusion protein term generates sufficient signal to call genuine ERα-directed calling card peaks, highlighting the importance of fusion orientation (Figure 3.5).
- Across five bulk ERα calling card RNA replicate libraries, up to 935 significant peaks were called, and each replicate is required for identifying significant peaks.
- Reproducibility depends primarily on the number of biological replicates rather than sequencing depth, with multiple replicates required for robust peak calling in line with previous calling card studies (Figure 3.8).

Benchmarking against existing methods

- The genomic feature distribution was comparable to ERα ChIP-Seq, indicating no major bias in insertion sites (Figure 3.12, Figure 3.13, and Figure 3.14).
- Up to 70% of Bulk ERα calling card peaks are within 1000 bp of ChIP-Seq peaks (Figure 3.15)
- Up to 81% of Bulk ERα calling card peaks overlap with either end of a ChIA-PET interaction within 1000 bp (Figure 3.23).
- Hypergeometric testing confirms that overlap analysis with ChIP-Seq and ChIA-PET anchors is statistically significant (Appendix Table 8.4 and Table 8.6).
- Heatmaps spanning 20 kb across bulk ERα calling card peaks showed clear central ERα ChIP-Seq signal enrichment (Figure 3.16, Figure 3.17, and Figure 3.18).
- Several CC peaks were found near known ERα target genes, including: GATA3, FOXA1 and GREB1 (Figure 3.25)
- Motif analysis recovered known ERα-associated motifs, including FOXA1, GATA3 and RXR (Figure 3.26).

Single-cell platform development

- A key finding was that integrating long-read sequencing into the pipeline provided a solution to a challenging step in single-cell analysis and avoided material loss.
- Across samples, 77% of UMI-filtered ERα long-read calling card peaks and 63% of non-filtered peaks overlap with ChIA-PET anchors (**Figure 8.12**).

- Overlap between long-read and bulk ERα calling cards was limited, likely reflecting stochastic insertion and small dataset size, highlighting the need for reproducibility metrics like IDR in calling card analysis.

Characterising a hypoxic spheroid model

- The MCF7 spheroid model consistently developed defined hypoxic and necrotic core formation at diameters of >600 μm, as confirmed by EF5 staining (Figure 5.4).
- Hypoxia, determined by EF5 staining, was anti-correlated with ERα levels and positively correlated with spheroid depth (Figure 5.5 and Figure 5.6).
- ERα expression is confirmed in the nucleus of cells at the surface of spheroids, consistent with the literature, by immunostaining; further investigation is required to determine significance (Figure 5.9).
- Treatment with 100 μM of amiloride impairs spheroid growth upon addition and prevents the depth and organisation required for hypoxic core formation (Figure 5.10, Figure 5.11, and Figure 5.13).

6.3. ERα calling cards are the first tool that can be used for the simultaneous study of single-cell ERα binding and its cell fate consequences over time

In this thesis, we have extensively shown the concordance between ChIA-PET (Fullwood et al., 2009), the gold standard of TF binding - ChIP-Seq (Holding et al., 2019; ENCODE Project Consortium, 2012), and our ERα calling card peaks, confirming our platform's ability to map ERα binding in breast cancer cells. Between 66.7% and 81% of calling card peaks overlapped with ChIA-PET anchors, and up to 70% overlapped with ChIP-Seq peaks (Holding et al., 2019; ENCODE Project Consortium, 2012). Although overlap was lower for long-read calling card peaks, the ENCODE ChIP-Seq signal was enriched at the centre of both bulk and long-read peaks.

Furthermore, ERα's interaction with canonical target genes, GATA3, FOXA1 and GREB1 was identified in both long-read and bulk ERα calling cards (Sun, Nawaz and Slingerland, 2007; Haines et al., 2020; Baron et al., 2007; Hurtado et al., 2011). Motif analysis confirmed that enrichment of the ERE motif among these peaks, consistent with known ERα's known binding preferences. In combination with ChIP-Seq signal enrichment and similar genomic feature distribution patterns, these data confirm that our HyPB-ERα fusion protein effectively maps ERα binding sites.

These findings are consistent with established strategies for validating calling card reporters (Wang et al., 2012a; Moudgil et al., 2020; Wang, Johnston and Mitra, 2007) and together confirm that redirected insertions represent genuine ERα binding events rather than undirected HyPB activity. Given that only one calling card peak overlaps between long-read and bulk calling card peaks, most long-read peaks represent distinct, non-overlapping loci. This reflects the stochastic insertion and limited sampling, underscoring the need for reproducibility metrics rather than indicating novel binding sites.

Our single-cell ERα system is capable of simultaneously capturing TF binding and transcriptomic data in a single experiment, unlike any other method described to date (Moudgil et al., 2020). Although scCUT&Tag has been successfully applied to TFs (Bartosovic, Kabbe and Castelo-Branco, 2021), its performance depends on high-quality, TF-specific antibodies and can exhibit variable sensitivity and sparsity at true binding sites, particularly in heterogeneous samples (Wu et al., 2021; Bartosovic and Castelo-Branco, 2023; Bartosovic, Kabbe and Castelo-Branco, 2021). In contrast, the ERα calling cards are antibody-independent and cumulatively record TF binding events over a period of time, which helps to capture rare transient or intermittent ERα occupancy that may be missed in snapshot assays. Critically, our implementation is multi-omic in nature as it also couples TF binding with the transcriptome from the same single-cells, linking ERα cistrome activity to cell state. In practical terms, scCut&TAG provides a complementary, high-resolution map of accessible, TF-bound chromatin at a single-cell timepoint, while calling card offers a historical record of ERα engagement with regulatory elements coupled with simultaneous RNA profiling. These methods are complementary: scCut&TAG can validate locus-specific ERα engagement and refine peak boundaries, and ERα calling cards can reveal how binding accumulates across time and associates with transcriptional programmes in 3D spheroid models. By combining them, we can distinguish persistent from transient ERα binding, quantify how microenvironmental cues (hypoxia) shift binding-expression coupling in single-cells, and construct a more complex mechanistic atlas of ERα regulation.

The current limitation of the ERα calling card system lies in the insertion number rather than in the genuine biological signal, as the number of insertions is restricted to the number of SRT copies delivered to the cell. Sensitivity is therefore constrained by the coverage of binding sites, not by the ability of the fusion to redirect insertions. Based on prior calling card publications, increasing replicates from five to ten or twelve can substantially improve coverage, approaching the sensitivity of ChIP-Seq (Wang et al., 2012a; Moudgil et al., 2020). This suggests that optimisation of replicate number, rather than sequencing depth alone, will

be the most effective strategy for enhancing statistical confidence in peak calling and for distinguishing genuine binding events from background insertions.

While scCut&Tag provides high-resolution cistromic maps (Bartosovic, Kabbe and Castelo-Branco, 2021) and calling cards go further to connect, binding to transcriptomic state, spatial sequencing adds positional context within the tumour microenvironment (Hu et al., 2023; Ståhl et al., 2016; Williams et al., 2022; Maynard et al., 2021). Recent advances described in a pre-printed publication show that single-cell multiscale foot printing has revealed the modular organisation of DNA regulatory elements, demonstrating that regulatory architecture can be dissected at single-cell resolution and linked to chromatin accessibility patterns across tissues (Hu et al., 2023). Incorporating these insights into our calling card sequencing platform would allow ERα binding events to be mapped not only to their genomic targets but also to their physical location within hypoxic spheroids or patient tumours. In this framework, scCut&Tag contributes locus-specific binding validation, spatial sequencing provides positional context, and ERα calling cards uniquely connect binding to transcriptomic state and preserve a historic record of occupancy. Integrating all three would enable the dissection of both temporal and spatial heterogeneity in ERα regulation, producing a more complete mechanistic atlas of how microenvironmental cues shape endocrine therapy resistance.

Taken together, our single-cell ERα calling card platform provides a reproducible and complementary tool to study hypoxia-induced ERα binding heterogeneity. When combined with our hypoxia spheroid model and orthogonal methods such as scCut&Tag and spatial sequencing, this system offers a powerful framework for future mechanistic studies of ERα regulation in breast cancer.

6.4. Single-cell ERα calling cards applied to a hypoxic spheroid model can be used to explore hypoxia-driven ERα heterogeneity over time

This thesis established two complementary foundations: a validated ERα single-cell calling card reporter and a reproducible MCF7 spheroid model that develops defined hypoxic and necrotic cores. While these tools were not experimentally combined here, their integration represents the next logical step. Applying single-cell ERα calling cards within hypoxic spheroids would allow longitudinal mapping of ERα binding events (Cammack et al., 2020) as oxygen gradients emerge and change, linking cumulative binding records to the transcriptomic state of the cells. This approach moves beyond data collection to mechanism definition: distinguishing persistent versus transient ERα occupancy under hypoxia, identifying binding-

expression couplings that drive adaptation, and pinpointing regulatory programs that could be targeted to disrupt endocrine therapy resistance. Additionally, the application of calling cards to 3D culture systems enhances the method because the spheroids more faithfully recapitulate in vivo conditions (Mazzucchelli et al., 2024; Guillen et al., 2022; Cheng, Leung and Singleton, 2022). Previous calling card reporters have only been applied to either 2D cell culture or AAV infection of whole mice (Wang et al., 2012a; Moudgil et al., 2020; Cammack et al., 2020; Lalli et al., 2022). A combination of our single-cell ERα calling cards would therefore provide a middle ground between these techniques, utilising a culture method that replicates in vivo conditions but is more accessible than animal experiments.

Spheroid culture provides an accessible and reproducible 3D system compared to organoids or PDX, requiring only non-adherent vessels or simple plate treatments (Vinci et al., 2012; Ekert et al., 2014). These vessels can either be purchased from suppliers directly, or standard plates can be treated with commonly available reagents in breast cancer research labs, like agarose and collagen, to prevent cell adherence (Friedrich et al., 2009; Lovitt, Shelper and Avery, 2014; Hongisto et al., 2013). This makes spheroids a practical platform for many laboratories, including those with limited resources (Bray et al., 2024; Xu et al., 2023b; Devi, 2025; Fu et al., 2025). By contrast, single-cell calling cards are not yet widely accessible, relying on specialised sequencing and analysis pipelines. The work here delivered the prerequisites for integration:

- Engineering a functional ERα reporter capable of single-cell binding and transcriptomic capture.
- Benchmarking evidence of ERα-directed insertions.
- A hypoxic spheroid model with reproducible core formation and spatial ERα expression variation.

These achievements establish feasibility and rationale. The combined application belongs in future methods, where experimental detail and validation will be expanded.

The validation of our ERα reporter provides the foundation to extend calling cards to other nuclear receptors that are relevant to breast cancer, such as PGR and Androgen Receptor (AR), which share regulatory architectures and are documented to crosstalk with ERα (Mohammed et al., 2015; Finlay-Schultz et al., 2017; Swinstead et al., 2016; Ponnusamy et al., 2019; Paakinaho and Palvimo, 2021). The generation of additional reporters with alternative transposases would enable multiplexed reporters to simultaneously map multiple

TFs alongside the transcriptome of single-cells in one experiment, surpassing ChIP-Seq's single-factor, bulk restraints. It is also possible to expand to other TFs, like the key pioneer factors for ERα binding FOXA1 and GATA3.

Utilisation of these reporters could enable the generation of the single-cell breast cancer steroid receptor atlas, a multiplexed dataset capturing NR and pioneer factor binding in the same single-cells, paired with transcriptomic profiles. This atlas would go further than any previous work by resolving the cross-factor co-occupancy, binding-expression coupling and how insertion accumulation changes over a period of time in a single assay.

FOXA1 is an example of the complexity that can be explored, as ERα and FOXA1 peaks show ~50% overlap, but FOXA1 loss results in the elimination of almost all ERα binding (Carroll et al., 2005; Hurtado et al., 2011). Adding FOXA1 and GATA3 reporters would quantify how cofactor co-occupancy with ERα varies across hypoxia gradients, testing whether certain sites persist under oxygen stress while others are transient.

A simpler immediate step is to deconvolute complexity within bulk peak overlaps at single-cell resolution through ERα calling cards. Pairing single-cell calling cards with complementary single-cell cistromic methods could reveal whether co-occupancy inferred from bulk represents true single-cell engagement across microenvironmental niches.

Together, our work on the ERα calling cards platform and MCF7 spheroid model provides the groundwork for future integration, but their independent value is already clear. The reporter demonstrates that ERα binding can be captured alongside transcriptomic state at single-cell resolution, and the spheroid model can be used to show how hypoxia and spatial gradients shape ERα expression. What follows is the opportunity to connect these strands: to use calling cards within spheroids not only to catalogue binding events, but to uncover mechanisms as persistent ERα occupancy under hypoxia or altered pioneer factor co-occupancy, that may drive therapy resistance. This trajectory points toward a multiplexed steroid receptor atlas, extending beyond ERα to PGR, AR, FOXA1, and GATA3, and offering a single-cell framework for dissecting how microenvironmental cues rewire nuclear receptor programs in breast cancer.

6.5. Broader applications of ERα calling cards and the standardised MCF7 spheroid model

Our platform has a broad range of applications that extend beyond hypoxia and ERα endocrine therapy research. The ability to record TF binding at single-cell resolution and link it to the transcriptomic state makes the system relevant to any tissue where ERα signalling shapes cell fate. Neurological (Bains, Cousins and Roberts, 2007; Cordey and Pike, 2005; Maioli et al., 2021; Spence et al., 2011), Cardiovascular (Pare et al., 2002; O'Lone et al., 2007), Ovarian and Endometrial (Hu, Zhong and He, 2005; Shabani et al., 2007) disorder research is just some examples that stand to benefit from our ERα calling cards model.

In neuronal tissue, ERα signalling modulates synaptic plasticity and mitochondrial function, where sex-specific differences in neurodegeneration correlate with ERα target gene expression (Bains, Cousins and Roberts, 2007; Cordey and Pike, 2005; Maioli et al., 2021). Astrocytes interact with neuronal and microglial cells to coordinate ERα signalling in a way that has clinical implications for patients, providing a protective effect when stimulating ERα (Bains, Cousins and Roberts, 2007; Cordey and Pike, 2005; Maioli et al., 2021; Spence et al., 2011). Using calling cards in this context would allow mapping of cell-type-specific enhancers, identifying how they are rewired in neurodegeneration, and testing how treatments such as E2 or tamoxifen alter binding and downstream fate decisions.

Cardiovascular cells also display protective ERα signalling, with distinct activation of ERα- and ERβ- target gene pathways in smooth muscle and endothelial cells (Visniauskas et al., 2023; Pare et al., 2002; O'Lone et al., 2007; Aryan et al., 2020). Our model enables tracking of how ERα enhancer activation changes during hypoxia and/or ischemia, and it combines with single-cell sequencing to link activation with gene expression changes across cell types. This would provide mechanistic insight into how ERα contributes to vascular adaptation and injury response.

Finally, in endometrial and ovarian tissues, ERα works in concert with the pioneer factor FOXA1, like its role in the breast. This context-dependent upregulation drives ERα-directed transcription with prognostic relevance depending on subtype (Tangen et al., 2014; Droog et al., 2017; Ge et al., 2023; Chen, Li and Ou-Yang, 2022). Applying our platform here would allow the dissection of ERα binding and cell fate consequences as cells transition to hyperplastic and malignant states, clarifying how pioneer factors co-occupancy shapes disease progression.

The activity of ERα across diverse tissues highlights the broader utility of our system. By providing single-cell maps of ERα binding linked to transcriptomic outcomes, the platform addresses knowledge gaps left in bulk assays; it reveals which cell types drive pathogenic ERα programmes, how binding rewires under stress or treatment, and which enhancer-gene couplings represent actionable mechanisms for patient care. This moves beyond simply "knowing more" to enabling targeted hypothesis testing about intervention points in ERα-driven disease.

6.6. Limitations of this work

Like any experimental method, both calling card reporters and spheroid cell culture systems have general limitations that apply across any context they are used, in addition to specific constraints observed in the application

6.6.1. General limitations of the platforms used

HyPB inserts only at TTAA tetranucleotides, which restricts sensitivity in genomic regions that lack nearby TTAA insertion sites (Chen et al., 2020; Mitra, Fain-Thornton and Craig, 2008; Ding et al., 2005; Moudgil et al., 2020). The flexibility of the fusion protein linker enables deposition within 1000 bp, partially overcoming this restriction by increasing the chance of nearby TTAAs, at the cost of reduced precision. As a result, calling card insertions are often located adjacent to, rather than within, the binding motif of the TF, since the HyPB fusion partner cannot access that motif if the TF is bound to it (Moudgil et al., 2020).

The exogenous expression of ERα from plasmids risks perturbing natural gene expression and chromatin accessibility. The original calling card authors note that integrating HyPB into the endogenous locus avoids superphysical levels of expression; however, this is much more time-consuming and costly (Moudgil et al., 2020). While high expression levels can be burdensome, they may also improve insertion coverage through increased representation in ERα heterodimers, representing a trade-off rather than a strict limitation.

As we have demonstrated in Chapter 3, the number of replicates is key to determining genuine TF directed calling card insertions (Moudgil et al., 2020). Sensitivity is therefore a concern regarding the detection of rare insertion events, and careful experimental design is required.
The calling card method is highly sensitive to the amount of donor plasmid DNA delivered during transfection, and in our experiments, this resulted in very low numbers of insertions per cell. With only a handful of insertions per cell, single-cell peak calling becomes unreliable, as genuine binding events cannot be distinguished from background noise. This limitation is

consistent with previous single-cell genomics studies, which show that sparse coverage reduces sensitivity to rare events and necessitates aggregation strategies (Buenrostro et al., 2015; Shi et al., 2022; Chen et al., 2019). Aggregating cells by type or cluster ("pseudobulking") increases statistical power and enables recovery of high-confidence peaks, as demonstrated in single-cell ATAC-Seq and RNA-Seq analyses (Buenrostro et al., 2015; Squair et al., 2021; Ji et al., 2020; Granja et al., 2021). However, aggregation comes at the cost of obscuring rare cell states, so its value lies in balancing sensitivity with resolution depending on the biological question.

Despite these limitations, previous calling card reporters (Wang et al., 2012a; Moudgil et al., 2020; Lalli et al., 2022; Cammack et al., 2020), and our benchmarking against existing ERα binding datasets provides confidence that, with sufficient replicates, our system can reliably map ERα binding.

Spheroids are the most basic form of modern 3D cell culture techniques and rely on cells that do not differentiate and form in vivo structures, inherently limiting their ability to recapitulate the in vivo breast cancer environment compared to other systems (Bhattacharya, Calar and de la Puente, 2020; Fisher and Rao, 2020). Multicellular tumour spheroids do exist to incorporate immune cell interactions into breast cancer development, but most models only use a single-cell type, neglecting a key cancer interaction and progression pathway (Lee et al., 2023b; Costa et al., 2016). As a result, spheroid models can lack typical breast heterogeneity and microenvironment control without the use of specialised microfluidic devices (Grist et al., 2019; Pyne et al., 2024).

However, even with these limitations, spheroids have been found to show more realistic pharmacokinetics compared to 2D cultures and can be used to reliably create physiological gradients with high reproducibility and affordability (Imamura et al., 2015; Nagelkerke et al., 2013; Friedrich et al., 2009; Hirschhaeuser et al., 2010).

6.6.2. Limitations specific to this thesis

MCF7 cells transfected with Lipofectamine 3000 and selected with puromycin showed high levels of sensitivity and cell death, requiring large deviations from published methods for calling card reporting (Moudgil et al., 2020). Data from another PhD student in our lab shows that transfected MCF7 cells that receive selection with blasticidin show increased viability compared to puromycin-selected cells. Therefore, developing a blasticidin SRT and selection would represent an opportunity to improve SRT yield by overcoming the sensitivity observed during puromycin selection.

Our N-terminal fusion constructs failed to yield sufficient mRNA, likely due to impaired protein function and cellular stress, as discussed in previous chapters. In repeated attempts at bulk calling cards, N-terminal transfected cells showed higher cell death, consistent with the slightly reduced functionality in validation experiments. While the N-terminal construct could insert SRTs, the C-terminal fusion was more effective and was therefore used in downstream analysis. Further investigation is needed to confirm the mechanism of stress. Currently, it's unclear whether N-terminal cells failed because the puromycin selection was too harsh, or if the N-terminal fusion was inherently less or non-functional.

The number of significant calling card peaks called across both datasets combined is lower than the expected number of peaks called in an average ChIP-Seq experiment (Bailey et al., 2013; Kharchenko, Tolstorukov and Park, 2008), largely due to the loss of replicates during sample processing. Increasing replicate number improves statistical power and peak recovery, as we and others have shown (Moudgil et al., 2020). However, as observed in ChIP-Seq, peak discovery likely follows a saturation curve: additional replicates yield diminishing returns once most genuine binding sites have been captured (Landt et al., 2012; Hansen et al., 2015; Jung et al., 2014; Boyd et al., 2023). This means replication is necessary to reach sensitivity, but replication alone cannot indefinitely increase peak counts or overcome insertion bias

Previous calling card studies have shown that HyPB has affinity for BRD4 binding sites and have used the undirected transposase to map BRD4 super enhancers (Moudgil et al., 2020). Given that BRD4 and ERα have been known to co-occupy genomic loci (Nagarajan et al., 2014; Lee et al., 2017b), this affinity of HyPB could increase the background signal in samples, specifically at BRD4-ERα co-binding sites. In our data, this background complicates peak calling by inflating undirected insertions at ERα binding sites. We observe an enrichment of BRD4 co-binding sites compared to ChIP-Seq, but not complete domination of peaks, indicating that other genuine binding sites can still be called. Evidence from previous publications indicates that the HyPB-BRD interaction is a direct protein-protein interaction (Gogol-Doring et al., 2016; Moudgil et al., 2020; Yoshida et al., 2017), raising the theoretical possibility of mutating the interaction domain or adopting alternative transposases to reduce the background. However, given BRD4's central role in ERα-driven breast cancer biology (Nagarajan et al., 2014; Feng et al., 2014; Zheng et al., 2022; Hanker, Sudhan and Arteaga, 2020), disrupting this interaction could have unintended consequences for the very mechanisms under study. Careful evaluation of alternative transposases, like *SleepingBeauty* and Tn5, or engineered variants, may therefore be necessary to balance reduced background

with preservation of the relevant biology. However, as we have demonstrated, our platform is sufficient for mapping ERα binding events despite the background insertions and affinity for BRD4.

Long-read calling cards captured ~100 distinct peaks that did not overlap with bulk datasets. Several explanations are possible. These peaks may represent loci located >1000 bp from ChIP-Seq peaks, reflecting genuine ERα binding sites that were undersampled in ChIP-Seq and therefore appear as false negatives. Alternatively, they could arise from technical artefacts, since our longread libraries were generated from amplified RNA templates, which may introduce bias in peak recovery. A statistical calculation of expected overlap by chance would help clarify whether these distinct peaks fall within the range of random sampling.

We also observed that longread peaks overlapped ~40% fewer ChIP-Seq sites than bulk calling cards. This disparity is striking and warrants further investigation. One possibility is that the smaller pool of longread peaks simply reduces overlap by chance. Another is that longread calling cards capture different aspects of ERα binding, more akin to interaction anchors measured by ChIA-PET, which in our data showed higher overlap than ChIP-Seq. This suggests that longread calling cards may preferentially recover binding events associated with chromatin interactions rather than all ChIP-Seq peaks. Increasing the replicate number would improve statistical power, but replication alone cannot resolve whether the observed differences reflect biology, technical bias, or stochastic sampling. Future work should therefore combine longread calling cards with orthogonal assays and formal overlap modelling to determine the origin and significance of these distinct peaks.

Deriving biological significance from amiloride treatment is complicated by off-target inhibition. At 100 μM, amiloride also inhibits NHE1, making it unclear how much ENaC contributes to the observed phenotype. Further work is required to disentangle these effects.

Single-cell libraries were generated, but sample processing errors resulted in extremely low complexity: most cells contained only a single read or peak. This prevented meaningful single-cell analysis, since genuine binding events could not be distinguished from background noise at the level of individual cells. Although the dataset was insufficient for robust bulk analysis, future improvements in library preparation, replicate number, and barcode strategies could overcome these limitations. Incorporating proper controls and optimised barcoding would allow single-cell longread calling cards to deliver reliable maps of ERα binding linked to transcriptomic state, finalising the single-cell pipeline outlined in this thesis.

6.7. Future directions

This thesis established both a single-cell ERα calling card platform and a reproducible 3D spheroid culture system, creating a foundation that can be immediately extended in future experiments. Together, these tools provide a platform ready to explore hypoxia-driven ERα heterogeneity and endocrine therapy resistance in more complex models.

The findings presented here suggest that a future iteration of single-cell long-read calling cards has the power to explore the deeper complexity of ERα binding heterogeneity that we have not yet captured. Given that both long-read and bulk calling cards are unique sets of insertions with high levels of concordance within the expected variance of gold-standard binding assays, ERα calling cards represent a powerful tool whose full potential is still to be demonstrated. After overcoming initial protocol challenges, we are confident that running ER calling cards with more replicates and optimised processing will generate highly valuable data to study endocrine therapy resistance.

6.7.1. Development of multicellular tumour spheroids to more faithfully recapitulate the in vivo cell-to-cell interactions

The current MCF7 hypoxic spheroid model will more faithfully recapitulate the tumour microenvironment than 2D cell culture (Fisher and Rao, 2020; Lee et al., 2023b; Imamura et al., 2015), but there are still key cell interactions missing, such as the interactions between immune cells, fibroblasts and stromal cells (Edmondson et al., 2014; Biju, Priya and Francis, 2023).

Adapting our model to co-culture multiple cell types into a single spheroid would be better to represent the tumour microenvironment (Froehlich et al., 2016; Han, Kwon and Kim, 2021; Vakhshiteh et al., 2023; Yakavets et al., 2020; Franchi-Mendes, Lopes and Brito, 2021; Dhandapani et al., 2023). Fibroblasts and macrophages are known to support tumour growth and immune evasion (Hu et al., 2022; Pan et al., 2020; Pittet, Michielin and Migliorini, 2022), and spheroid models have been used in the past to investigate how these cell types can influence tumorigenesis and cancer progression (Franchi-Mendes et al., 2021; Franchi-Mendes, Lopes and Brito, 2021). Incorporating these cell types into our MCF7 tumour model, then applying single-cell calling cards, would allow us to test how immune and stromal interactions reshape ERα binding programs under hypoxia. The expected outcome would be identification of cell-type specific ERα binding changes that correlate with pro-tumorigenic signals, such as angiogenesis, immunosuppression and therapy resistance (Zhang et al.,

2021c; Kim et al., 2022; Colegio et al., 2014; He and Zhang, 2021; Franchi-Mendes, Lopes and Brito, 2021).

6.7.2. Use fusion proteins with other transposase reporters to multiplex TF binding analysis

A unique challenge in ERα calling cards is the co-binding of BRD4 at ERα binding sites, increasing background insertions at genuine peaks. This could be mitigated by utilising an alternative transposase, such as *SleepingBeauty* or Tn5. Although these alternative transposases were tested by the Mitra group and PiggyBac was found to be the best (Moudgil et al., 2020; Wang et al., 2012a), BRD4 co-binding may not have been an issue in their system. Other enzymes may avoid any BRD4-induced bias and improve signal-to-noise.

Future work could extend beyond ERα to create reporters for pioneer factors FOXA1 and GATA3, or nuclear receptors such as PGR and AR. Multiplexing these reporters would enable a single-cell atlas of ER and its partners, revealing how protein-protein interactions adapt under hypoxia. Unlike scCut&Tag, which requires antibodies, scCC reporters can be genomically encoded, offering the possibility of stable, inducible reporters in vivo. Recent work on extrachromosomal DNA amplification (Pradella et al., 2025), suggests that it's possible to engineer self-replicating SRTs that could provide a constant supply of insertions, potentially overcoming current limitations in insertion number.

6.7.3. Adapting calling cards to record 3D interactions in the same way as ChIA-PET

Another exciting application is to adapt calling cards to capture transcriptomic, interatomic, and ERα binding data in one experiment. Peak calling could be weighted using Bayesian models that incorporate public ChIP-Seq datasets as priors, allowing us to distinguish direct ERα binding from peaks deposited through long-range chromatin contacts. Protein-protein interatomic adaptations could be explored by designing reporters for ERα cofactors and testing whether co-insertions occur at interacting loci. Rather than being prohibitively challenging, this approach could begin with the integration of existing public ChIP-Seq and ChIA-PET data, providing a framework to resolve which peaks represent genuine binding versus 3D proximity. Through slight experimental changes, it may be possible to examine 3D interactions directly. At harvest, once cells are processed for single-cell or bulk sequencing, excess cells can be fixed and processed in a method like ChIA-PET or HiChIP. This would give clarity to which calling card peaks are located at ERα binding sites and which have been deposited at long-range DNA brought into proximity to ERα.

6.7.4. Single-cell ERα calling cards coupled with 3D cell culture and spatial sequencing can revolutionise ERα binding heterogeneity research

The ultimate future application of this work will require an AAV-based delivery model to introduce ERα calling card reporters into patient-derived organoids/xenografts. Using AAV delivery ensures stable integration and long-term recording (Cammack et al., 2020), which is essential for experiments that track binding heterogeneity as seen in spheroids that grow and develop hypoxic cores.

By sampling patient-derived spheroids that are resistant or sensitive to treatment, ERα calling cards could capture binding heterogeneity and link it to endocrine therapy resistance. These spheroids can be transduced with calling card reagents and grown to a size where hypoxia naturally develops. Applying 10x single-cell RNA sequencing alongside calling cards would generate detailed profiles of each cell, revealing how ER binding changes across the spheroid.

Importantly, this analysis would not only identify heterogeneity but also test whether ERα binding shifts along defined cell state trajectories. Diffusion map embedding followed by diffusion pseudotime (DPT) analysis could be used to test the hypothesis that ER binding progressively changes as cells move through hypoxia-induced trajectories, rather than appearing as random heterogeneity.

Spatial sequencing of sequential tissue sections would then confirm whether the DPT trajectories correspond to hypoxic regions of the spheroid. This integration would allow us to directly associate ERα binding changes with both spatial location and developmental trajectory. Finally, comparing hypoxic ERα signatures from spheroids with existing tumour datasets would test whether the observed binding programs correlate with patient tumours, providing translational relevance.

7. Conclusion

In conclusion, we have developed a novel, versatile platform for studying ERα binding and transcriptomic heterogeneity at single-cell resolution using long-read sequencing, which is now ready to be applied to our characterised hypoxic MCF7 spheroid model.

We have created a powerful and accessible set of tools that utilises our novel pipeline to answer key questions surrounding how ERα heterogeneity contributes to disease progression. Future application of our platform in the hypoxic MCF7 spheroid model will allow us to dissect

how hypoxia-driven ERα heterogeneous binding and transcriptional evolution occur during hypoxia. This application, in turn, will clarify their role in disease progression and endocrine therapy resistance.

Our ERα calling cards are a unique tool that fills a critical gap in single-cell mechanistic ERα research (Akhoundova and Rubin, 2022; Fang et al., 2024). To date, there are no single-cell datasets that directly profile ERα binding, meaning that our understanding of ERα regulation has been limited to bulk assays or indirect inference from transcriptomic data (Li et al., 2023; Kaya-Okur et al., 2019; Hainer et al., 2019). This absence has prevented researchers from resolving how ERα binding varies between individual cells, how rare resistant states emerge, or how binding programs persist under microenvironmental stress. By providing the first framework to record ERα binding at single-cell resolution, our platform enables questions that could not previously be asked: which binding events are transient versus persistent, how binding rewiring correlates with rare resistant cell states, and how hormone receptor crosstalk influences ERα binding. The applications of our platform are wide-ranging within breast cancer research, including endocrine therapy resistance initiation, and beyond, with potential use in other tissues such as the endometrium and ovaries (Tangen et al., 2014; Droog et al., 2017; Ge et al., 2023; Chen, Li and Ou-Yang, 2022). We have also characterised the development of hypoxia in MCF7 spheroid models and demonstrated its application to hypoxic breast cancer research through the inhibition of the hypoxic ERα target gene EnAC with Amiloride (Malcolm et al., 2025; Ware et al., 2021; McQueen et al., 2025). This highlights the broader utility of combining mechanistic binding assays with functional perturbations, which is essential for linking ERα binding heterogeneity to downstream phenotypes.

Importantly, our simplified 3D culture model enhances research accessibility to low-middle SDI regions where breast cancer burdens are predicted to rise (Wanigasekara et al., 2023; Gunti et al., 2021; Matei et al., 2024). This accessibility applies specifically to the spheroid model, which requires only modest resources compared to organoids or xenografts. Ensuring that such models remain reproducible and affordable should be a key consideration for global equity in cancer research (Arnold et al., 2022; GBD 2021 Forecasting Collaborators, 2024; GBD Collaborators, 2025; Kim et al., 2025).

Looking ahead, a combination of our methods with spatial sequencing, ChIA-PET and multiplexed calling card reporters could yield the first single-cell ERα binding atlas. This atlas would be capable of mapping the spatial and temporal heterogeneity of ERα binding caused by hypoxia, enabling a mechanistic deconstruction of how microenvironmental factors contribute to the heterogeneity of ERα and its binding partners. Our methods are a resource

that can generate new hypotheses and guide the rational design of interventions to combat endocrine therapy resistance in ERα+ breast cancer.

Our work, therefore, lays the foundation for future studies that can resolve the earliest molecular events in ERα therapy resistance and explore how they might be targeted. By establishing both a novel single-cell calling card platform and an accessible spheroid model, this thesis opens the door to a new generation of mechanistic studies that can inform precision oncology. In doing so, we hope that this work can be taken forward and utilised to develop novel therapies that enhance the efficacy of treatments and translate into real differences in patients' clinical outcomes.

8. Appendices

8.1. Chapter 3 results appendices

8.1.1. Fusion protein cloning primer list

Table 8.1: A list of cloning primers used to create N- and C-terminal HyPB-ERα fusion proteins.

Primer name	**Primer Sequence (5' – 3')**
C term NEBuilder Fwd	5' – CATCGACATGTGCCAGAGCTGTTTCAAGTTGGGCGGCGGCGCCCCCGCCGTGGGCGGCGGACCCAAGGCCG CGGATAAAATGACCATGACCCTCCACAC - 3'
C term NEBuilder Rev	5' -GAAAGCTGGGTCTAGATATCTCGATCAGACTGTGGCAGGGA – 3'
New ESR1 N term NEB Fwd	5' –GAACCAATTCAGGCGCGCCGCCACCATGACCATGACCCTCCACAC - 3'
New ESR1 N term NEB Rev	5' - GTCGTCCAGGCTGCTGCCCATTTTATCCGCGGCCTTGGGTCCGCCGCCCACGGCGGGGGCGCCGCCGCCCA ACTTGACTGTGGCAGGGAAACC - 3'

8.1.2. Fusion protein cloning plan diagram

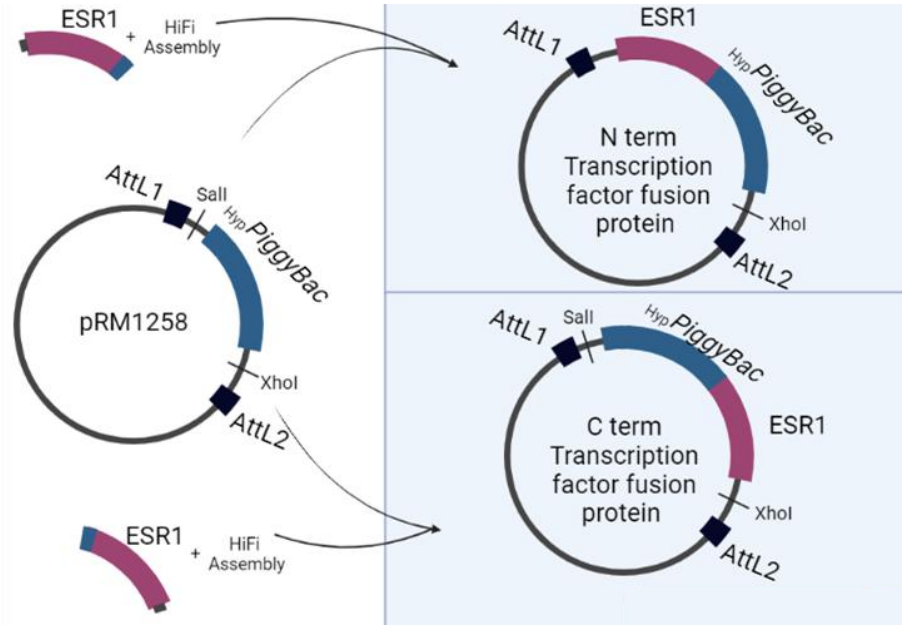


Figure 8.1: A diagram of N-and-C terminal cloning of ERα using NEBuilder HiFi Assembly.
pRM1258 WT-HyPB was digested in two separate reactions with SalI and XhoI respectively. After purification, these are mixed with inserts from N-term HiFi assembly PCR and HiFi assembly master mix. N-terminal cloning destroys the SalI recognition site whereas C-terminal cloning will not destroy XhoI after digestion.

8.1.3. Full Western blots

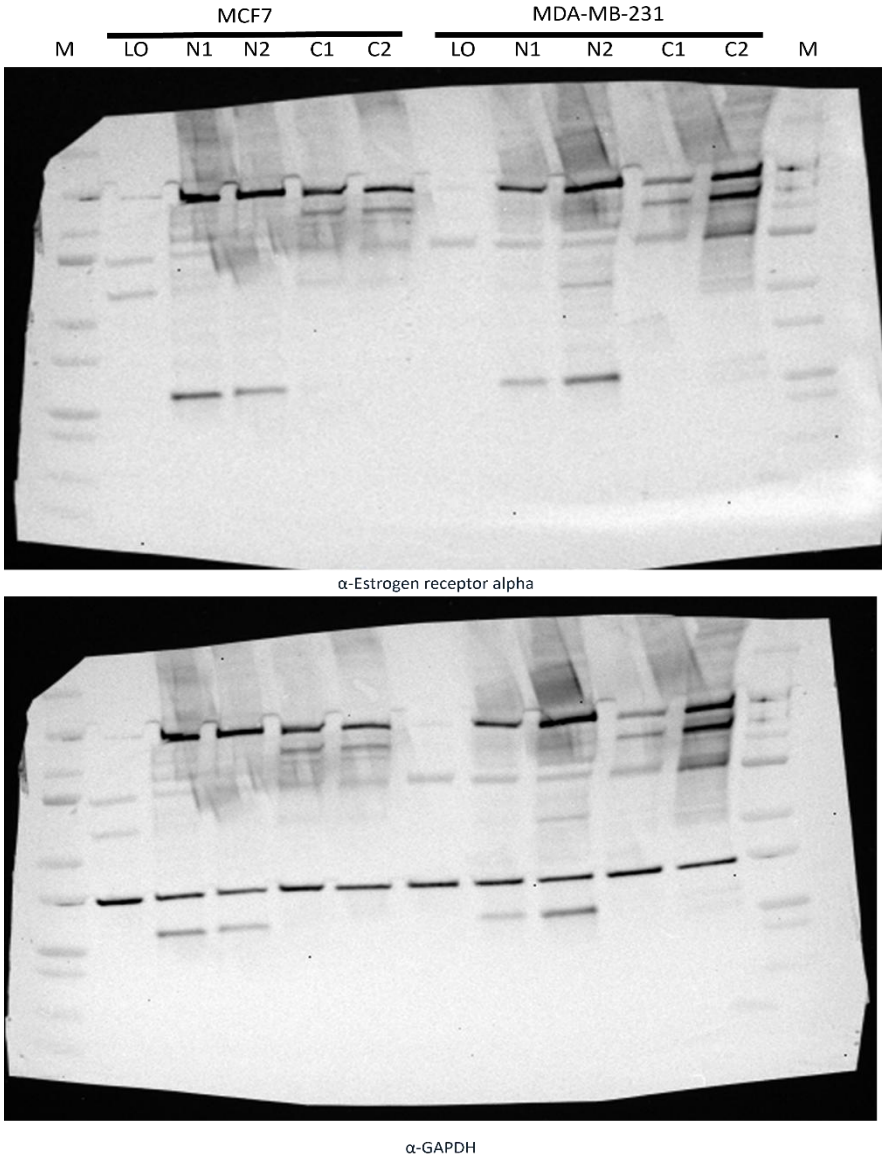


Figure 8.2: Fusion protein expression plasmids can express the full-length protein as shown by the presence of a band of the expected weight on an anti-ERα (N-terminal) Western blot.

Following gateway cloning of C- and N-terminal fusion protein genes into expression plasmids, the plasmids were transfected into MCF7 and MDA-MB-231 cells and cell lysates were collected after 72 hours for Western blotting with the Abcam α-Erα antibody (ab3575). **M** = Marker, **LO** = Lipofectamine only, **Nx** = N-terminal fusion protein transfection, replicate x, **Cx** = C-terminal fusion protein transfection, replicate x. The molecular weight of the full fusion protein is 136 kDa, as indicated by the first arrow; each transfection shows a band at this molecular weight, whereas the lipofectamine does not. The second arrow indicates the endogenous ER band at 66 kDa, while expression of this band can be seen in the empty vector and N-fusion condition, it is not apparent in the C-fusion conditions. The membrane was re-blotted with an α-*GAPDH* antibody as a loading control for this blot (band found at 37 kDa), demonstrating equal protein loading in each lane

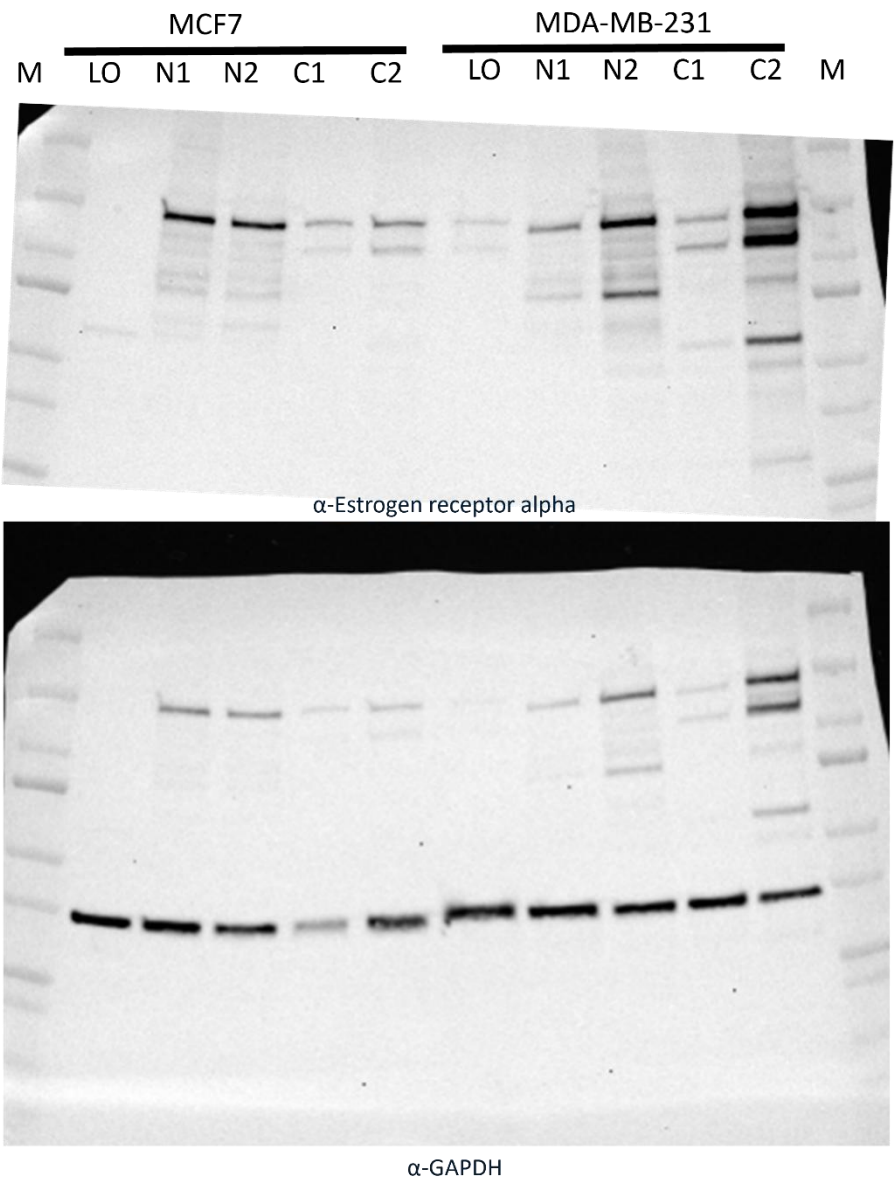


Figure 8.3: Fusion protein expression plasmids can express the full-length protein as shown by the presence of a band of the expected weight on an anti-ERα (C-terminal) Western blot.

Following gateway cloning of C- and N-terminal fusion protein genes into expression plasmids, the plasmids were transfected into MCF7 and MDA-MB-231 cells and cell lysates were collected after 72 hours for Western blotting with the Santa Cruz α-Erα antibody (sc-8002). **M** = Marker, **LO** = Lipofectamine only, **Nx** = N-terminal fusion protein transfection, replicate x, **Cx** = C-terminal fusion protein transfection, replicate x. The molecular weight of the full fusion protein is 136 kDa, as indicated by the first arrow; each transfection shows a band at this molecular weight, whereas the lipofectamine does not. The second arrow indicates the endogenous ER band at 66 kDa, while expression of this band can be seen in the empty vector and N-fusion condition, it is not apparent in the C-fusion conditions in MCF7 cells. Interestingly, there appears to be induction of endogenous ER expression in both C-terminal fusion protein transfections in MDA-MB-231 cells. The membrane was re-blotted with an α-GAPDH antibody as a loading control for this blot (band found at 37 kDa), demonstrating equal protein loading in each lane except for C1 and C2 MCF7 lanes, in which there appears to be some misloading as shown by the weaker GAPDH band.

8.1.4. Puromycin kill curve

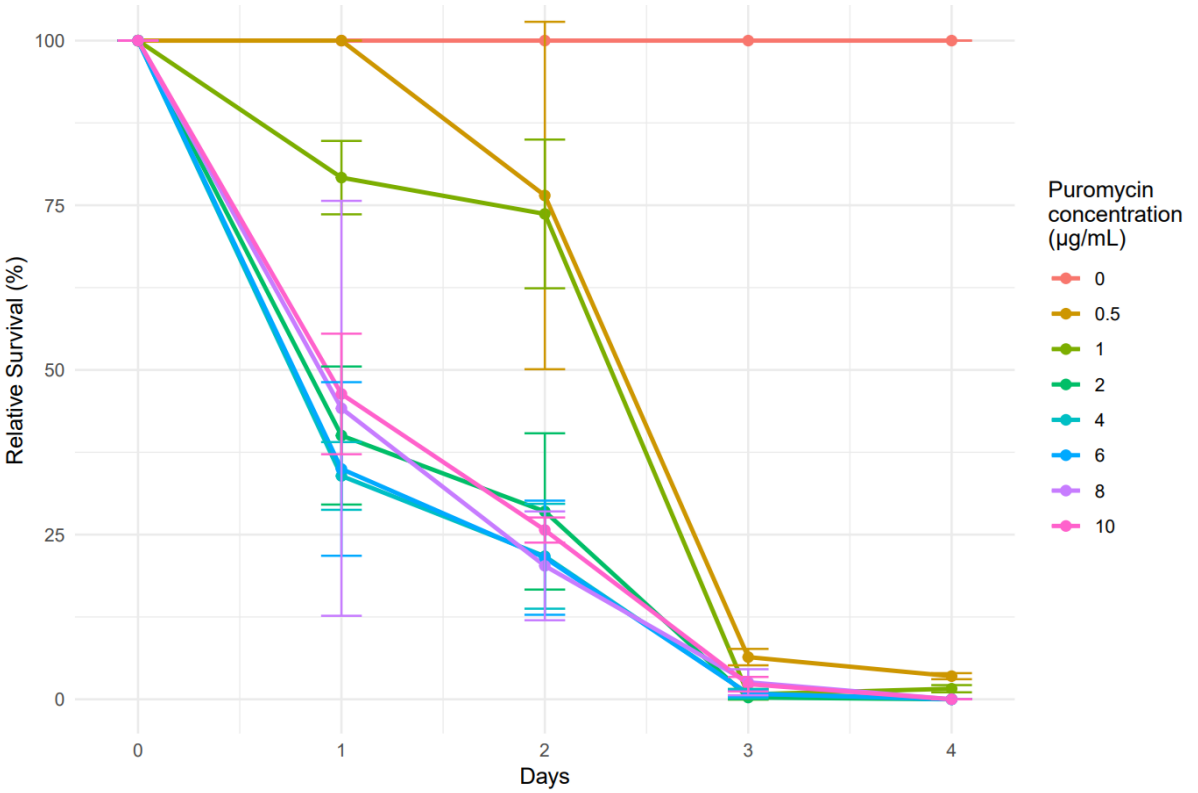


Figure 8.4: All concentrations of puromycin from 2 µg/mL to 10 µg/mL killed all MCF7 cells after one week; therefore, 2 µg/mL was selected to use in calling card selections. Cells were seeded in a 24-well plate and were imaged every day for a week after treatment with puromycin at varying concentrations. Cells at four specific regions of each well were counted to determine relative survival. Most conditions resulted in no viable cells after 7 days; 0.5 and 1 µg/mL resulted in the slowest rate of death over this time; however, there did appear to be a few cells remaining after 7 days (less than 1%). Given that 2 µg/mL was the lowest concentration that achieved 100% killing by day 7, it was selected to carry forward into future experiments. Results have been restricted to day 4 as the biggest changes in cell death occur in this timeframe. Error bars show standard deviation.

8.1.5. High-sensitivity tape station traces

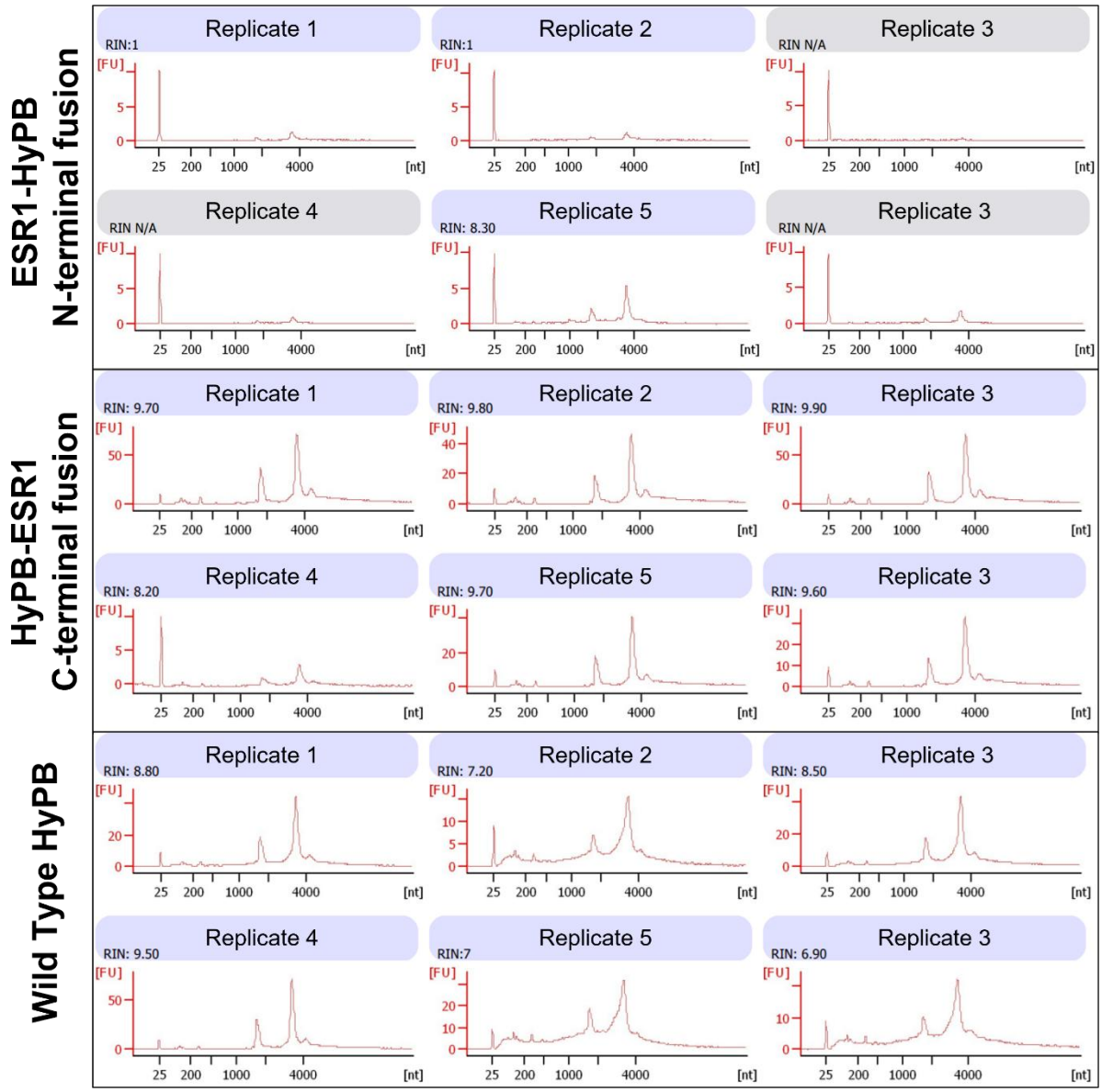


Figure 8.5: High-sensitivity TapeStation total RNA traces show good recovery of RNA for C-terminal fusion and WT HyPB bulk calling cards, whereas N-terminal fusion calling cards fail to produce sufficient quality RNA.
Cells transfected with N-terminal fusion protein plasmids showed reduced viability throughout puromycin selection, which is reflected in the low RNA yield shown on the TapeStation trace. The amount of RNA across replicates was not sufficient for at least 50 ng/µL required to carry forward. C-terminal fusion protein and wild-type traces on the other hand, are sufficient to carry forward, except for the C-terminal fusion replicate 4, which also produced insufficient RNA.

8.1.6. Bulk Calling Cards RNA quality

Table 8.2: **RNA recovery for N-terminal fusion calling cards and C-terminal replicate 4 is insufficient to carry forward.**

Cells transfected with N-terminal fusion protein plasmids showed reduced viability throughout puromycin selection, which is reflected in the low RNA yield shown on the TapeStation trace. The amount of RNA across replicates was not sufficient to carry forward. C-terminal fusion protein and wild-type traces, on the other hand, are sufficient to carry forward, except for the C-terminal fusion replicate 4, which also produced insufficient RNA.

Sample	**RNA Concentration (ng/ul)**	**RNA integrity number (RIN)**	**Live Cell Count (x10^4 cells)**
HyPB-ERA C Replicate 1	190	9.7	79.3
HyPB-ERA C Replicate 2	108	9.8	N/A
HyPB-ERA C Replicate 3	172	9.9	N/A
HyPB-ERA C Replicate 4	8	8.2	N/A
HyPB-ERA C Replicate 5	92	9.7	N/A
HyPB-ERA C Replicate 6	78	9.6	N/A
WT *HyPB* Replicate 1	136	8.8	36.4
WT *HyPB* Replicate 2	101	7.2	N/A
WT *HyPB* Replicate 3	163	8.5	N/A
WT *HyPB* Replicate 4	159	9.5	N/A
WT *HyPB* Replicate 5	248	7	N/A
WT *HyPB* Replicate 6	154	6.9	N/A
ERA-HyPB N Replicate 1	5	1	6.5

ERA-HyPB N Replicate 2	7	1	N/A
ERA-HyPB N Replicate 3	3	N/A	N/A
ERA-HyPB N Replicate 4	5	N/A	N/A
ERA-HyPB N Replicate 5	15	8.3	N/A
ERA-HyPB N Replicate 6	5	N/A	N/A

8.1.7. Calling Card cDNA concentrations

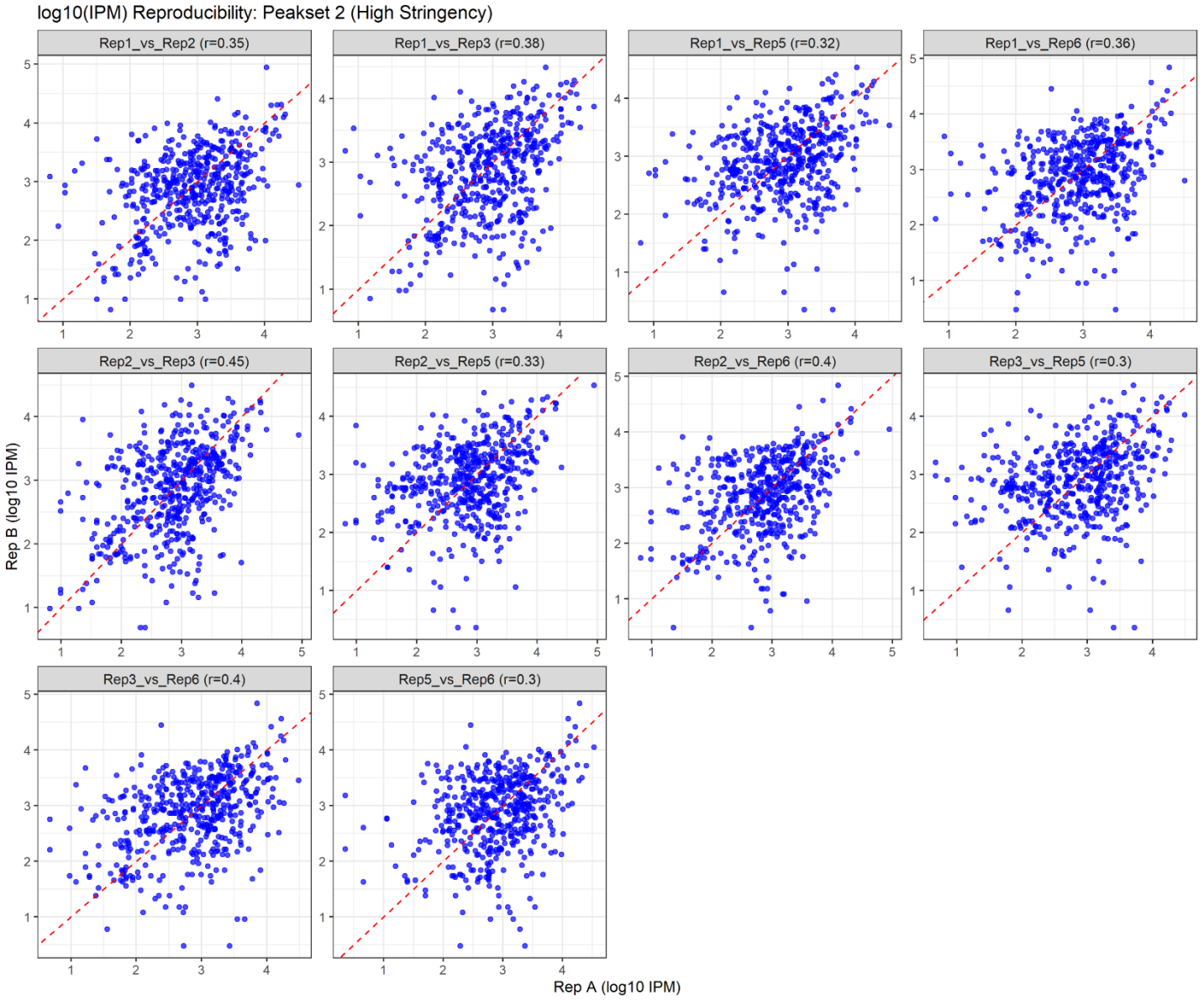
Table 8.3: Final cDNA libraries fall within the expected concentration of 2-4 ng/µl within the 500-700 bp fragment range.

All the replicates processed contain sufficient fragments within the expected calling card range. Concentration was determined using 1:10 dilutions of final libraries on a high-sensitivity DNA TapeStation assay.

Sample	**cDNA Concentration within Calling Card region (ng/ul)**
HyPB-ERA C Replicate 1	3.78
HyPB-ERA C Replicate 2	3.66
HyPB-ERA C Replicate 3	6.15
HyPB-ERA C Replicate 5	2.81
HyPB-ERA C Replicate 6	3.33
WT *HyPB* Replicate 1	2.65
WT *HyPB* Replicate 2	2.17
WT *HyPB* Replicate 3	2.90
WT *HyPB* Replicate 4	3.27
WT *HyPB* Replicate 5	3.05
WT *HyPB* Replicate 6	2.24

8.1.8. Replicate ROC analysis

A


log10(IPM) Reproducibility: Peakset 2 (High Stringency)
Rep1_vs_Rep2 (r=0.35)
Rep1_vs_Rep3 (r=0.38)
Rep1_vs_Rep5 (r=0.32)
Rep1_vs_Rep6 (r=0.36)
Rep2_vs_Rep3 (r=0.45)
Rep2_vs_Rep5 (r=0.33)
Rep2_vs_Rep6 (r=0.4)
Rep3_vs_Rep5 (r=0.3)
Rep3_vs_Rep6 (r=0.4)
Rep5_vs_Rep6 (r=0.3)
Rep B (log10 IPM)
Rep A (log10 IPM)

B

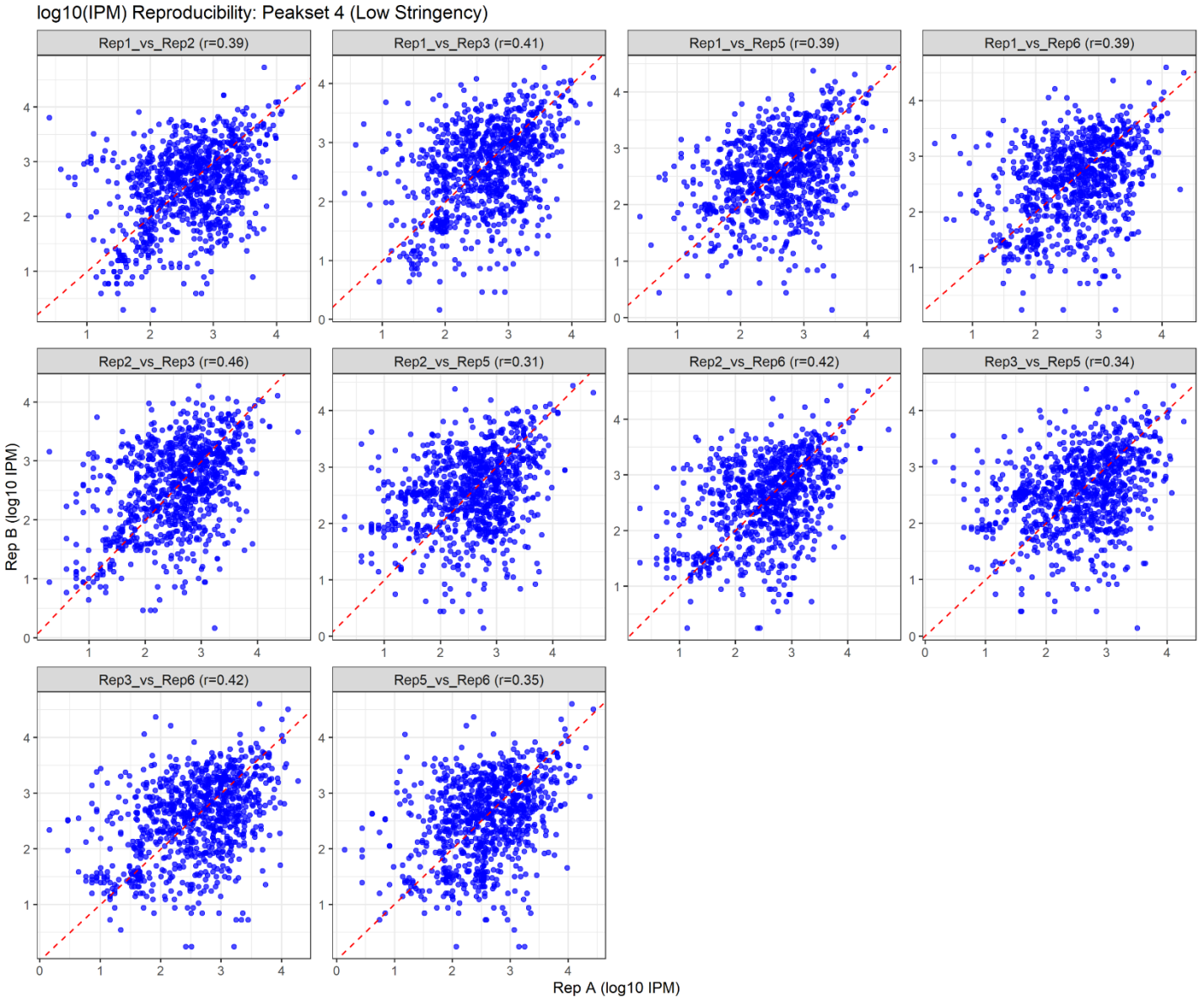

log10(IPM) Reproducibility: Peakset 4 (Low Stringency)
Rep1_vs_Rep2 (r=0.39)
Rep1_vs_Rep3 (r=0.41)
Rep1_vs_Rep5 (r=0.39)
Rep1_vs_Rep6 (r=0.39)
Rep2_vs_Rep3 (r=0.46)
Rep2_vs_Rep5 (r=0.31)
Rep2_vs_Rep6 (r=0.42)
Rep3_vs_Rep5 (r=0.34)
Rep3_vs_Rep6 (r=0.42)
Rep5_vs_Rep6 (r=0.35)
Rep B (log10 IPM)
Rep A (log10 IPM)

Figure 8.6: **Comparative analysis of HyPB-ERα calling card replicates at MACCs peaks.**

(A) Displays the distribution of insertions across the high stringency peak set (n = 506), **(B)** shows the corresponding results for the low stringency peak set (n = 935). Insertions from each replicate were filtered to retain only those overlapping the designated peak set and subsequently normalised to Insertions Per Million (IPM) before plotting. Across both stringency thresholds, replicate comparisons reveal moderate concordance, with correlation coefficients ranging between 0.30 and 0.46. Notably, the strongest correlations (r = 0.45-0.46) are observed between replicates 2 and 3, underscoring reproducibility while also highlighting variability in peak overlap depending on stringency. The calculated significance for all plots in **(A)** in reading order from top left to bottom right is as follows: 7.74×10^{-39}, 4.31×10^{-48}, 3.57×10^{-57}, 9.96×10^{-56}, 7.1×10^{-24}, 2.27×10^{-63}, 1.8×10^{-34}, 6.42×10^{-34}, 3.09×10^{-27}, 2.75×10^{-49}. The calculated significance for all plots in **(B)** in reading order from top left to bottom right is as follows: 5.24×10^{-18}, 5.31×10^{-28}, 1.04×10^{-22}, 4.59×10^{-26}, 3.97×10^{-11}, 6.69×10^{-49}, 6.23×10^{-16}, 2.87×10^{-20}, 2.05×10^{-14}, 7.66×10^{-26}.

8.1.9. Markdown of Calling Card Analysis Pipeline

The code used to analyse these Markdown files will be made available online (https://github.com/JackStenning/Thesis).

8.1.10. Karyotype plots of ER calling cards and ChIP-Seq peaks

Figure 8.7: HyPB-ERα calling card peaks are found across all chromosomes. Using a karyotype plot made with the karyoploteR package, we show that peaks called in the high **(A)** and low **(B)** stringency peaks can be found across all chromosomes in the genome. Some regions may not show calling card peaks as they were listed on the ChIP-Seq blacklist. Some chromosomes appear to be enriched for calling card peaks, such as chromosomes 1 and 20, compared to chromosomes 18 and 22. Peaks can be identified on the Y chromosome in both peak sets; this is likely an artefact caused by repetitive DNA sequence found in the Y chromosome or a transposon fusion.

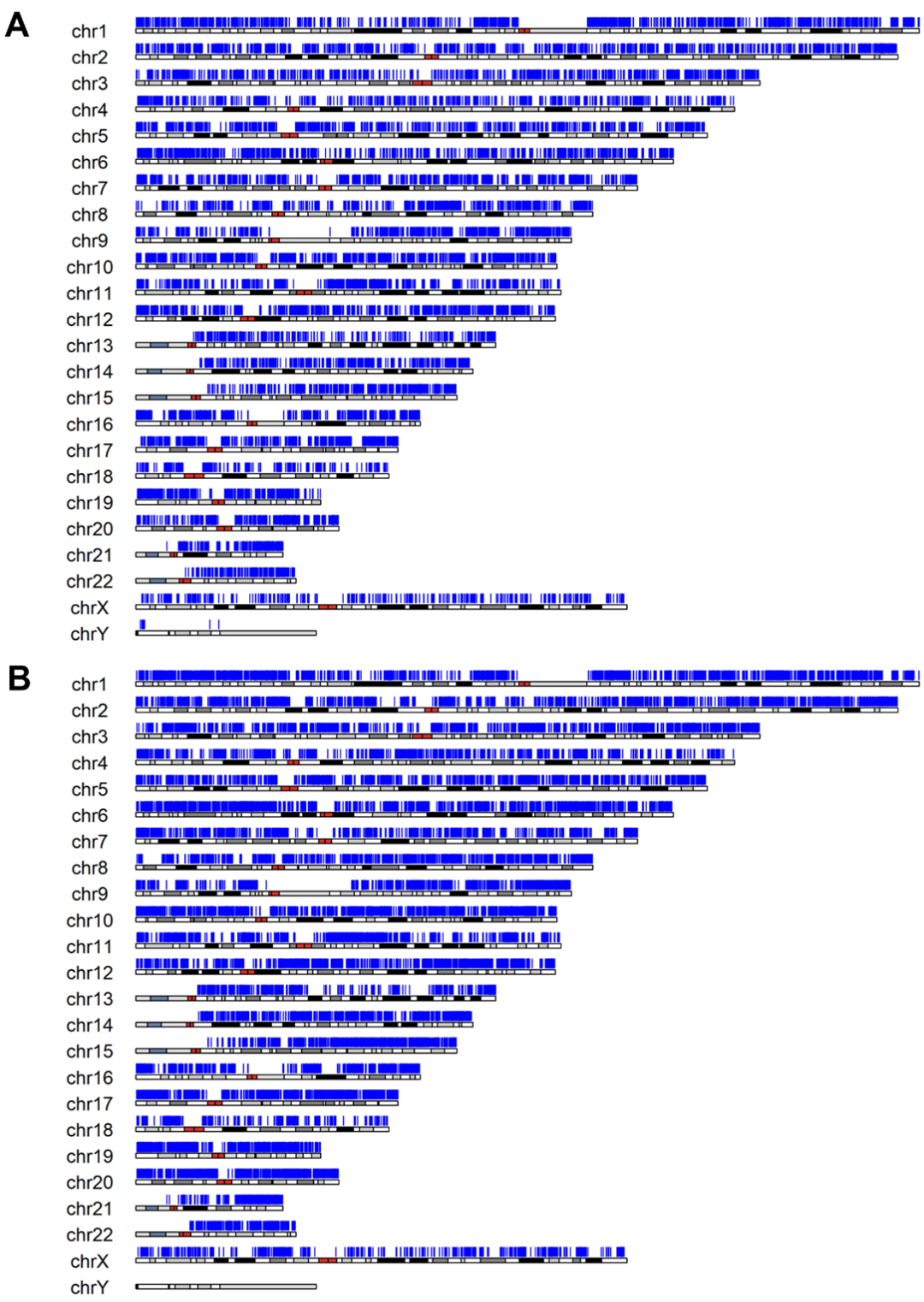


Figure 8.8: ERα ChIP-Seq peaks are found in similar locations in the genome as calling card peaks.

Using a karyotype plot made with the karyoploteR package, we show that peaks called in the GEO **(A)** and Encode **(B)** datasets can be found across all chromosomes in the genome. Blacklisted regions are clear on this plot, as shown by the large regions without any peaks identified, flanked by heavily populated regions. Peaks can be identified on the Y chromosome in the GEO dataset; this is likely an artefact caused by repetitive DNA found in the Y chromosome or a transposon fusion.

8.1.11. Overlap analysis hypergeometric test results

Table 8.4: Hypergeometric test statistics for overlap between HyPB-ERα calling card peaks and ERα ChIP-Seq datasets.

This table reports the enrichment values calculated for peak overlaps shown in Figure 3.15. For each dataset (GSE109820, ENCFF365BIT [Encode 1], and ENCFF063JMY [Encode 2]), the number of overlapping peaks, expected overlaps under a random model, and associated hypergeometric p-values are listed. Calculations were performed using the subset of peaks identified by the PyCallingCards MACCS function at a q-value threshold of 0.05. The results demonstrate highly significant enrichment ($p < 1 \times 10^{-300}$) of ERα calling card peaks within independent ChIP-Seq datasets, supporting the reproducibility of binding site identification across methods.

Sample	N_TTAA	K_TTAA_in_BRD4	n_TTAA_in_sample	k_overlap	Percent_CC_TTAA_in_ChIP	Fold_Enrichment	Hypergeom_p
High_GEO	19214773	32076	5106	1632	31.962	191.4672	0
High_Encode	19214773	48588	5106	1525	29.867	118.1123	0
Low_GEO	19214773	32076	8914	2459	27.586	165.2498	0
Low_Encode	19214773	48588	8914	2310	25.914	102.4815	0

Table 8.5: Hypergeometric enrichment statistics for overlap of HyPB-ERα calling card peaks with BRD4 ChIP-Seq datasets.

This table summarises the statistical values underlying the overlap analysis shown in Figure 3.21. For each BRD4 dataset (Liu et al., 2022 [GSE171908] and Zheng et al., 2022 [GSE186645]), the observed number of overlapping peaks, the expected overlap under a random distribution, and the resulting hypergeometric p-values are reported. Calculations were performed using peaks identified by the PyCallingCards MACCS function. The results demonstrate that ERα calling card peaks are significantly enriched at BRD4 binding sites, with p-values exceeding 1 $x10^{-300}$, confirming that the overlap is far greater than expected by chance and consistently higher than that observed for ERα ChIP-Seq peaks.

Sample	N_TTAA	K_TTAA_in_BRD4	n_TTAA_in_sample	k_overlap	Percent_CC_TTAA_in_BRD4	Fold_Enrichment	Hypergeom_p
High_Liu	19214773	198517	5113	2056	40.2112	38.92108	0e+00
High_Zhang	19214773	198517	5113	2056	40.2112	38.92108	0e+00
Low_Liu	19214773	198517	8931	3282	36.7484	35.56936	0e+00
Low_Zhang	19214773	198517	8931	3282	36.7484	35.56936	0e+00

Table 8.6: Hypergeometric test statistics for overlap between HyPB-ERα calling card peaks and ERα ChIA-PET anchors.

This table reports the enrichment values calculated for peak overlaps shown in Figure 3.23. For each ChIA-PET dataset, the number of overlapping peaks, expected overlaps under a random distribution, and associated hypergeometric p-values are listed. Calculations were performed using peaks identified by the PyCallingCards MACCS function at a q-value threshold of 0.05. The results demonstrate highly significant enrichment of ERα calling card peaks at ChIA-PET anchors ($p < 1 \times 10^{-190}$ for high stringency and $p < 1 \times 10^{-300}$ for low stringency and ChIP-Seq), supporting the robustness of overlap between calling card reporters and independent chromatin interaction datasets.

Sample	N_TTAA	K_TTAA in_ChIA	n_TTAA in sample	k_overlap	Fold Enrichment	log10p_hyper	Fisher Pvalue
Low_CC	19214773	11657215	8914	7078	1.30881535	313.339367	0
High_CC	19214773	11657215	5106	4081	1.31742602	190.886108	1.3E-191
ENCODE	19214773	11657215	48588	36549	1.23990076	1001.59792	0
GEO	19214773	11657215	32076	24677	1.26809739	834.051883	0

8.1.12. Promoter selection data

Note that this work was NOT carried out by our group. It was carried out by Amy Cullen from a previous lab.

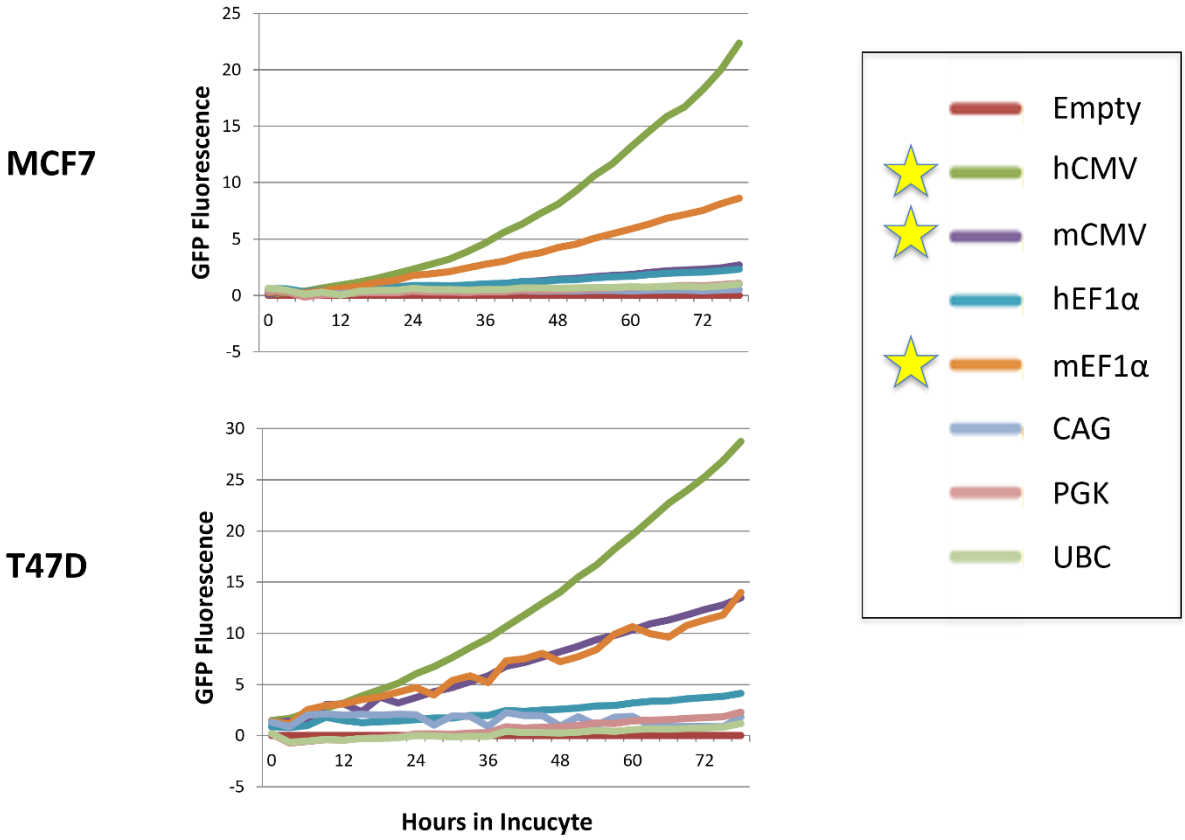


Figure 8.9: **The hCMV promoter drives the highest levels of GFP expression in MCF7 and T47D cells over time.**

Following transduction with a GFP containing plasmid under the control of several promoters delivered by lentiviral particles into MCF7 and T47D cells, the hCMV promoter was found to drive the largest increase in expression as shown through GFP fluorescence. In MCF7s, this was followed by the mEF1α, in T47Ds, the second and third best promoters are tied between mCMV and mEF1α. This work was carried out by Amy Cullen from a previous lab.

8.1.13. Heterogeneous TdTomato expression

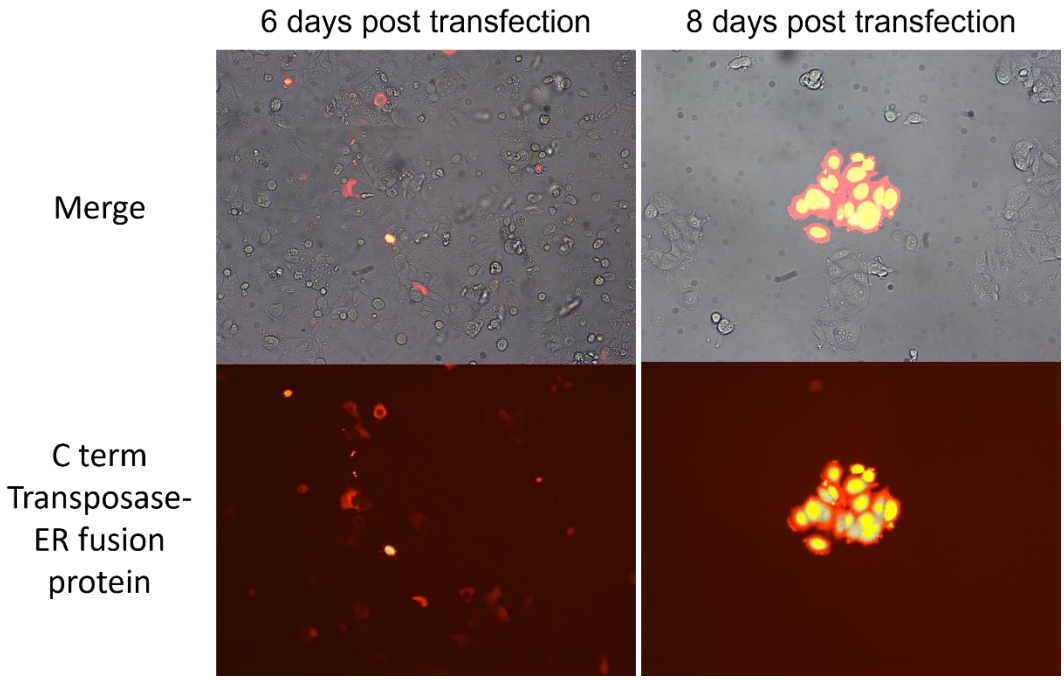


Figure 8.10: Images of cells co-transfected with C-terminal ERα fusion proteins and TdTomato SRT show heterogenous insertion of PiggyBac Transposons.
Cells co-transfected with C-terminal ERα fusion expression plasmids and the TdTomato SRT were imaged every 48 hours for 11 days **(B).** These images show that some of the cells within these plates were far brighter than the others. There appears to be a high level of heterogeneity in the amount of TdTomato expression, indicating a difference in the levels of insertion by HyPB-ER fusion proteins.

8.2. Chapter 4 results appendices

8.2.1. Primer list

Table 8.7: A table detailing the primers used in site-directed mutagenesis and single-cell calling cards
For initial single-cell calling cards, Bio_Long_PB_LTR_3xPT and Bio_IlluminaSeq1_scCC_10X_3xPT were used to amplify calling card transcripts. For the SRTs that were used, these contained a 4 bp misalignment that impaired amplification. A pilot amplification was run using Universal_Long_PB_LTR, and IlluminaSeq1_scCC_10X_3xPT were used instead.

Primer name	**Primer Sequence (5' – 3')**
C_term_Kozak_SDM_Fwd	5' – CCAATTCAGTCGACTGCCACCATGGGCAGCAGCCTG - 3'
C_term_Kozak_SDM_Rev	5' – CAGGCTGCTGCCCATGGTGGCAGTCGACTGAATTGG - 3'
Universal_Long_PB_LTR	5'-[Phos]GCGTCAATTTTACGCAGACTATCTTT - 3'
IlluminaSeq1_scCC_10X_3xPT	5'-[Phos]ACACTCTTTCCCTACACGACGCTCTTCCGATCT - 3'
Bio_IlluminaSeq1_scCC_10X_3xPT	5'-[Phos]ACACTCTTTCCC[BtndT]ACACGACGCTCTTCCGA*T*C*T - 3'
Bio_Long_PB_LTR_3xPT	5'-[Phos]GCGTCAATTTTACGCAGAC[BtndT]ATCTTTC*T*A*G - 3'
scCC_CustomIndex1	5'-GAGACTGGCAAGTACACGTCGCACTCACCATGA - 3'
scCC_PB_CustomRead2	5'-CGTGTAGGGAAAGAGTGTGCGTCAATTTTACGCAGACTATCTTTCTAG - 3'
scCC_P5_adapter	5'-AATGATACGGCGACCACCGAGATCTTCACTCATTCCACACGACTCCTTGCCAGTCTC*T - 3'
scCC_P5_primer	5'-AATGATACGGCGACCACCGAGATC - 3'
scCC_P7_primer	5'- CAAGCAGAAGACGGCATACGAGAT - 3'
scCC_P7_adapter_N701	5'-[Phos]GAGACTGGCAAGTACACGTCGCACTCACCATGATAAGGCGAATCTCGTATGCCGTCTTCTGCTTG - 3'
scCC_P7_adapter_N702	5'-[Phos]GAGACTGGCAAGTACACGTCGCACTCACCATGACGTACTAGATCTCGTATGCCGTCTTCTGCTTG - 3'
scCC_P7_adapter_N703	5'-[Phos]GAGACTGGCAAGTACACGTCGCACTCACCATGAAGGCAGAAATCTCGTATGCCGTCTTCTGCTTG - 3'
scCC_P7_adapter_N704	5'-[Phos]GAGACTGGCAAGTACACGTCGCACTCACCATGATCCTGAGCATCTCGTATGCCGTCTTCTGCTTG - 3'
scCC_P7_adapter_N705	5'-[Phos]GAGACTGGCAAGTACACGTCGCACTCACCATGAGGACTCCTATCTCGTATGCCGTCTTCTGCTTG
scCC_P7_adapter_N706	5'-[Phos]GAGACTGGCAAGTACACGTCGCACTCACCATGATAGGCATGATCTCGTATGCCGTCTTCTGCTTG
scCC_P7_adapter_N707	5'-[Phos]GAGACTGGCAAGTACACGTCGCACTCACCATGACTCTCTACATCTCGTATGCCGTCTTCTGCTTG
scCC_P7_adapter_N708	5'-[Phos]GAGACTGGCAAGTACACGTCGCACTCACCATGACAGAGAGGATCTCGTATGCCGTCTTCTGCTTG

8.2.2. Single-cell calling card primer cocktail

Table 8.8: Calling card SRT amplification primer cocktail

Reagent	**Volume (µL)**
100 µM Bio_Illumina_Seq1_scCC_10X_3xPT primer	5
100 µM Bio_Long_PB_LTR_3xPT primer	5
Low TE Buffer (ThermoFisher, Ref: J75793.AE),	10

8.2.3. RNA concentration and RIN for follow-up bulk ERα calling cards

Table 8.9: RNA concentrations of C-terminal fusions were increased following optimisations, but N-terminal fusion transfections failed to generate sufficient mRNA. C-terminal fusion protein shows an increase in RNA yield compared to the initial bulk calling cards experiment in Chapter 2. The Undirected HyPB control however appears to have a lower average concentration than the previous calling card experiment. Cells transfected with N-terminal fusion protein plasmids continued to show reduced viability throughout the 0.5 µg/ml puromycin selection. The amount of RNA across N-terminal replicates was not sufficient to carry forward.

RNA	**Concentration (ng/ul)**	**RNA integrity number (RIN)**	**Live cell count (x10^4 cells)**
WT-HyPB + Puro Rep 6	49	9.9	36.4
WT-HyPB + Puro Rep 5	110	9.9	N/A
WT-HyPB + Puro Rep 4	96	9.9	N/A
WT-HyPB + Puro Rep 3	120	9.8	N/A
WT-HyPB + Puro Rep 2	76	9.7	N/A
WT-HyPB + Puro Rep 1	124	9.8	N/A
C fusion + Puro Rep 6	198	9.8	79.3
C fusion + Puro Rep 5	124	9.3	N/A
C fusion + Puro Rep 4	140	9.7	N/A
C fusion + Puro Rep 3	259	9.1	N/A
C fusion + Puro Rep 2	203	9.7	N/A
C fusion + Puro Rep 1	130	9.2	N/A
N fusion + Puro Rep 5	6	N/A	6.5
N fusion + Puro Rep 4	7	N/A	N/A
N fusion + Puro Rep 3	7	N/A	N/A
N fusion + Puro Rep 2	7	N/A	N/A
N fusion + Puro Rep 1	5	N/A	N/A

8.2.4. Comparing trimmed and untrimmed long-read calling cards

8.2.4.1. The long-read calling card pipeline approaches the target of captured cells and is comparable to the bulk ERα calling card in the number of peaks called

The difference between the using trimmed and untrimmed reads was clear in the key metrics produced by the Epi2Me single-cell report. Trimmed data yielded only 120 cells from an input of ~10,000, each with a median of one Unique Molecular Identifier (UMI) and one locus detected per cell, far below expectations. In contrast, untrimmed data recovered 4,326 cells, much closer to the 6,000 cells expected from a 10,000-cell target (Zheng et al., 2017).

8.2.4.1. Long-read ERα calling card peaks overlap with ChIA-PET anchors more than ChIP-Seq, reflecting overlap patterns of bulk calling cards

Overlap analysis revealed distinct patterns, with long-read overlap with ChIP-Seq peaks being lower than that with ChIA-PET anchors (Figure 8.11**Error! Reference source not found.**), reflecting the pattern observed in bulk calling cards.

When compared to ChIP-Seq peaks (Figure 8.11**Error! Reference source not found.** A), trimmed datasets overlapped 8-30% depending on the source and filtering, while untrimmed datasets overlapped 9-33%. These values are modest, but they demonstrate that long-read calling cards do recover loci coinciding with established ERα binding sites. Hypergeometric testing confirms that trimmed long-read calling card peaks are not significantly overlapping with ChIP-Seq peaks ($p = 0.85$). In contrast, the signal of ChIA-PET anchors was consistently significantly enriched across trimmed datasets ($p < 1 \times 10^{-7}$).

In contrast, overlap with ChIA-PET anchors was consistently higher for trimmed and untrimmed data (Figure 8.11**Error! Reference source not found.** B) . Trimmed insertions overlapped with 63-67% of anchors, while untrimmed datasets overlapped with 63-66% for unique insertions and 77% for UMI filtered peaks (Figure 8.12 B). This level approaches the 83% observed with bulk ERα, and 72% of previously published SP1 calling card peak overlaps with ChIP-Seq (Wang et al., 2012a). Hypergeometric testing confirms that all untrimmed long-read calling card conditions showed statistically significant enrichment with ChIA-PET anchors ($p < 0.005$), except the UMI-filtered qbed insertions ($p = 0.25$)(Appendix Table 8.10). This degree of concordance highlights the consistency of long-read ERα calling cards with chromatin interaction data.

These results show that while trimming is the standard process and can provide high specificity, it was too restrictive for our dataset and prevented peak calling. Skipping trimming

increased the recovery of cells and insertions, enabling peak calling and overlap analysis. Most importantly, the overlap pattern reproduced what was seen in the bulk experiment: overlap with ChIP-Seq peaks but stronger overlap with ChIA-PET anchors. This consistency supports the conclusion that long-read sequencing provides a viable alternative to circularisation-based methods, allowing direct recovery of calling card insertions and producing overlap profiles that reflect genuine ERα binding activity.

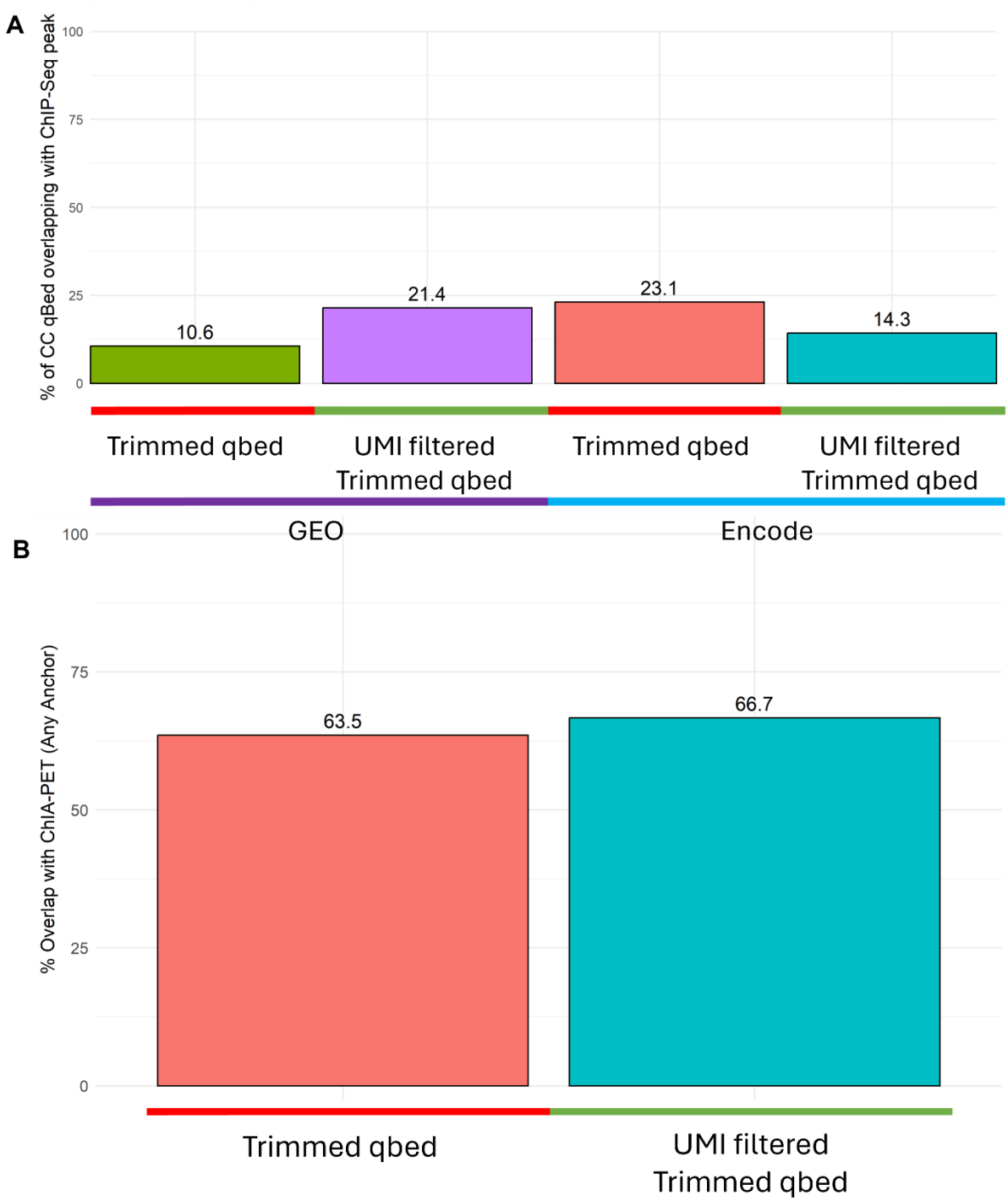


Figure 8.11: Trimmed long-read HyPB-ERα directed calling card insertions show superior overlap (67%) with ChIA-PET anchors compared with ERα ChIP-Seq peaks, which show 3 fold fewer overlapping calling card peaks

(A) Insertions show limited overlap with ChIP-Seq peaks, with GEO overlap increasing from 10.6% to 21.4% after UMI filtering, while ENCODE overlap decreases from 23.1% to 14.3%. No significant enrichment was detected between long-read calling cards and ChIP-Seq ($p = 0.85$) (B) UMI-filtered trimmed insertions demonstrate strong overlap with ChIA-PET anchors (66.7%), compared with around three-fold fewer overlaps observed for ChIP-Seq peaks.

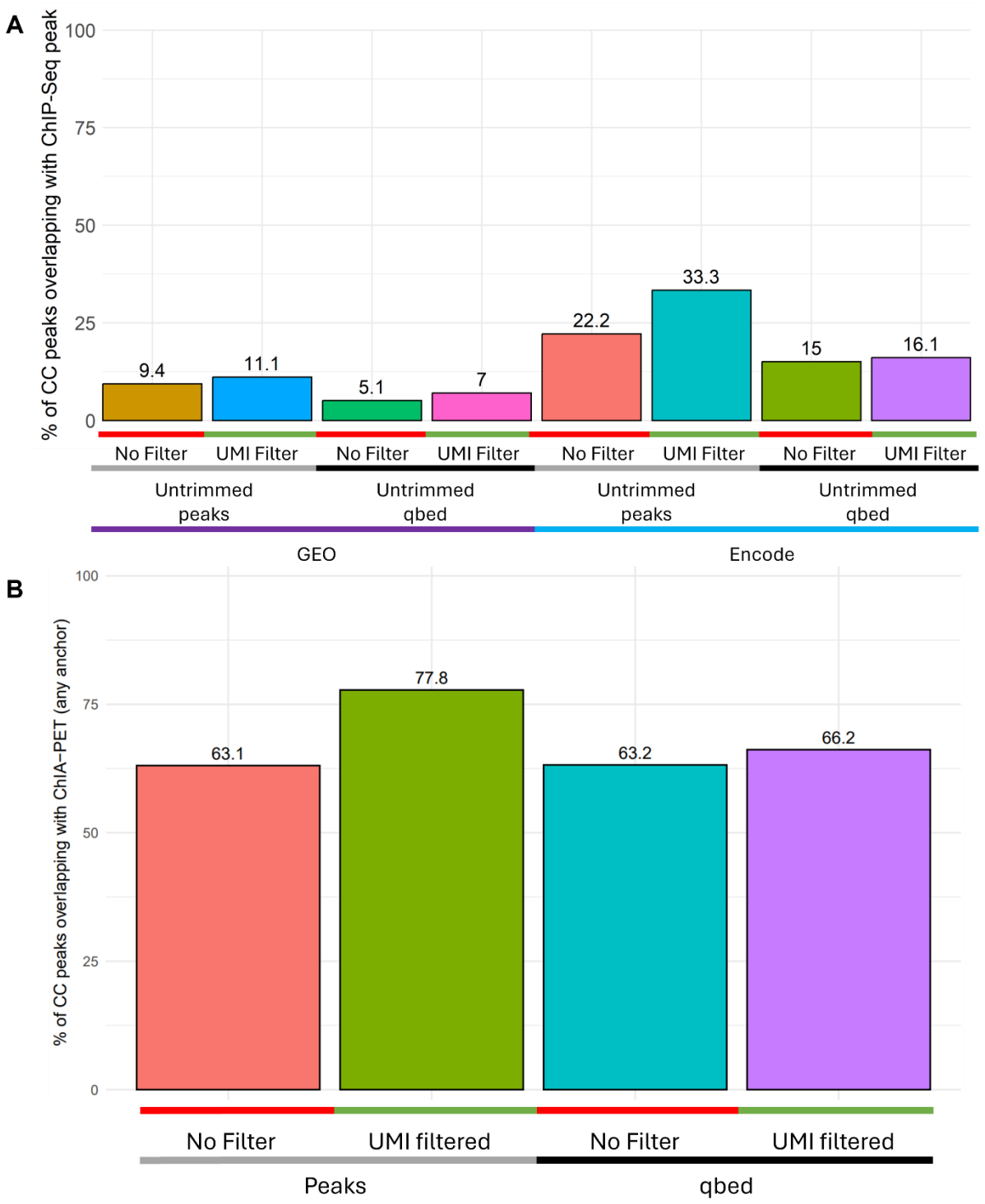


Figure 8.12: 77.8% of untrimmed UMI-filtered long-read HyPB-ERα directed calling card peaks overlap with ChIA-PET anchors compared to 63.1% of non-filtered peaks, while overlap with ChIP-Seq peaks remains limited.
(A) Overlap with ChIP-Seq datasets remains limited, though UMI filtering increases concordance with ENCODE peaks (33% vs 22%) and modestly improves overlap with GEO. **(B)** UMI-filtered peaks show stronger overlap with ChIA-PET anchors (77.8%) compared to non-filtered peaks (63%). Untrimmed calling card overlap with both ChIP-Seq peaks and ChIA-PET anchors was statistically significant ($p < 0.005$) except for UMI-filtered qbed insertions overlap with either ChIA-PET anchor ($p = 0.25$).

8.2.5. Overlap analysis hypergeometric test results

Table 8.10: Hypergeometric test statistics for overlap between long-read HyPB-ERα calling card peaks compared with ERα ChIA-PET anchors and ChIP-Seq datasets.

This table reports the enrichment values underlying the overlap analyses shown in Figures 4.5 and 4.6. For trimmed longread insertions, overlap with ChIP-Seq peaks was limited and not statistically significant (p = 0.85), whereas overlap with ChIA-PET anchors was highly significant, with ~67% of peaks overlapping. For untrimmed longread insertions, only the ChIA-PET overlap was assessed; Unfiltered peaks showed 77.8% overlap compared to 63.1% for non-filtered peaks. Hypergeometric testing confirmed significant enrichment of longread calling card peaks at ChIA-PET anchors (p < 0.005 for most comparisons), while overlap with ChIP-Seq datasets remained modest. These results highlight that longread calling cards preferentially capture chromatin interaction anchors identified by ChIA-PET, with enrichment consistently stronger than for ChIP-Seq peaks.

Sample	N_TTAA	K_in_ChIA	n_in_sample	k_overlap	Percent_overlap	Fold_enrichment	NegLog10_hyper
Untrimmed_GEO	19214773	11657215	1547	1031	66.6451	1.09852	6.175
Untrimmed_ENCODE	19214773	11657215	1547	1031	66.6451	1.09852	6.175
Untrimmed_UMI_GEO	19214773	11657215	24	21	87.5	1.44228	2.374
Untrimmed_UMI_ENCODE	19214773	11657215	24	21	87.5	1.44228	2.374
trimmed_GEO	19214773	11657215	2	1	50	0.82416	0.073
trimmed_ENCODE	19214773	11657215	2	1	50	0.82416	0.073
trimmed_UMI_GEO	19214773	11657215	0	0	-	-	0
trimmed_UMI_ENCODE	19214773	11657215	0	0	-	-	0

Untrimmed_peak_ChIA	19214773	11657215	1547	1031	66.6451	1.09852	6.175
Untrimmed_UMI_peak_ChIA	19214773	11657215	24	21	87.5	1.44228	2.374
Untrimmed_peak_qbed_ChIA	19214773	11657215	949	654	68.9146	1.13593	7.103
Untrimmed_UMI_qbed_ChIA	19214773	11657215	26	18	69.2308	1.14114	0.607

8.2.6. qbed files with non-unique insertions show inflated overlap with ChIP-Seq and ChIA-PET peaks/anchors

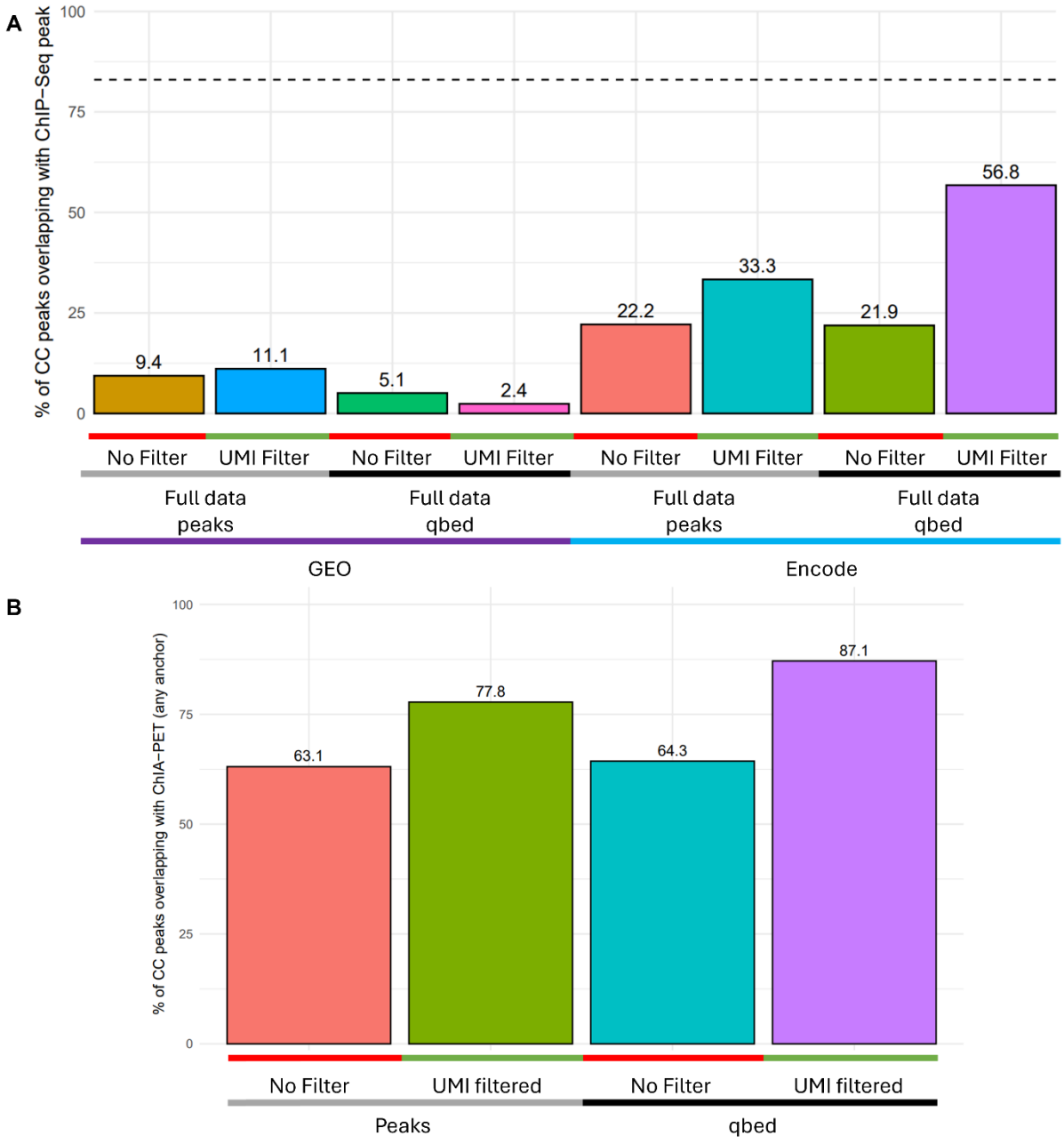


Figure 8.13: Without the removal of non-unique full long-read calling card insertions, overlap with ChIP-Seq peaks and ChIA-PET anchors
Bedtools window was used to determine whether insertions from full long read peaks and qbed files overlap with existing ChIP-Seq datasets **(A)** and ChIA-PET datasets **(B)**. There is not a large change in overlap with the GEO ChIP-Seq, but UMI-filtered qbeds show an additional 40% of overlap. ChIA pet anchors also show an additional 25% of overlap

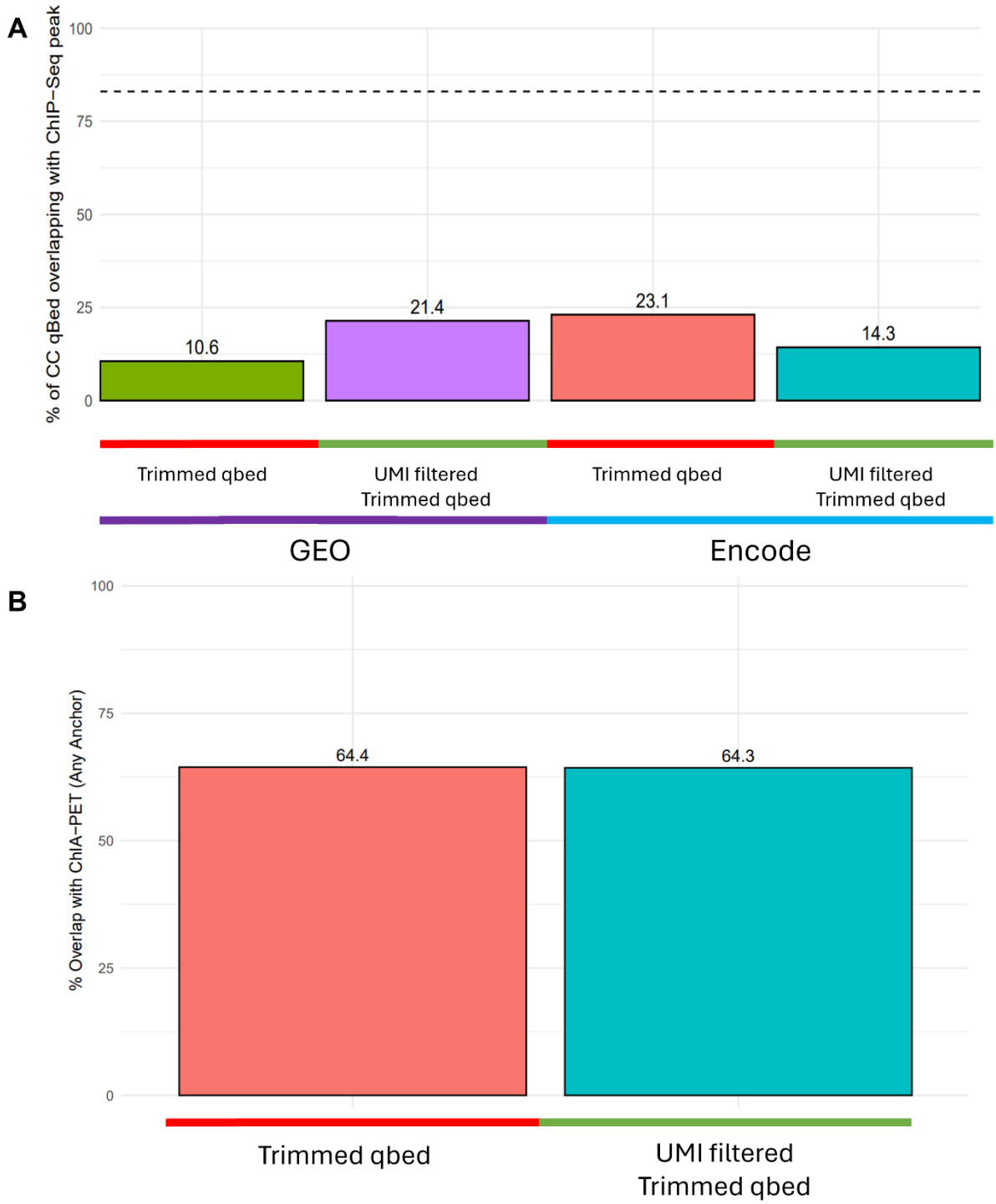
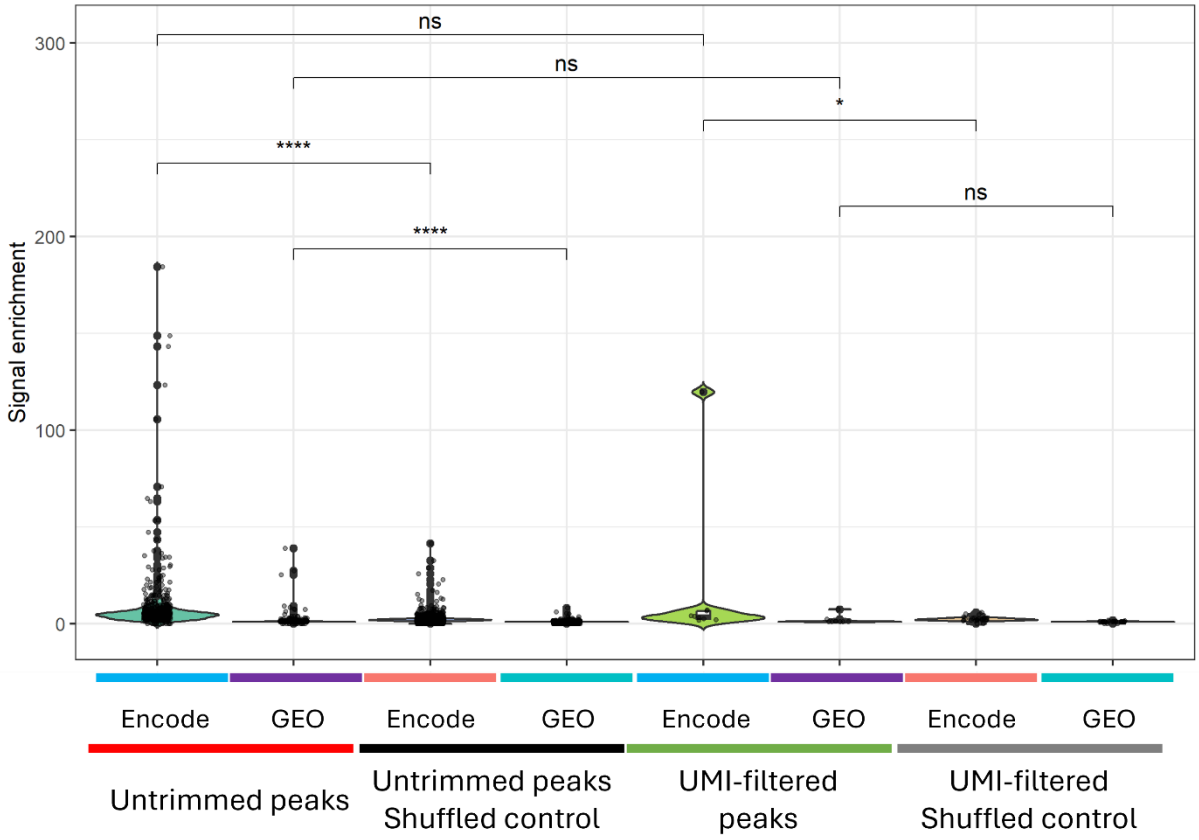
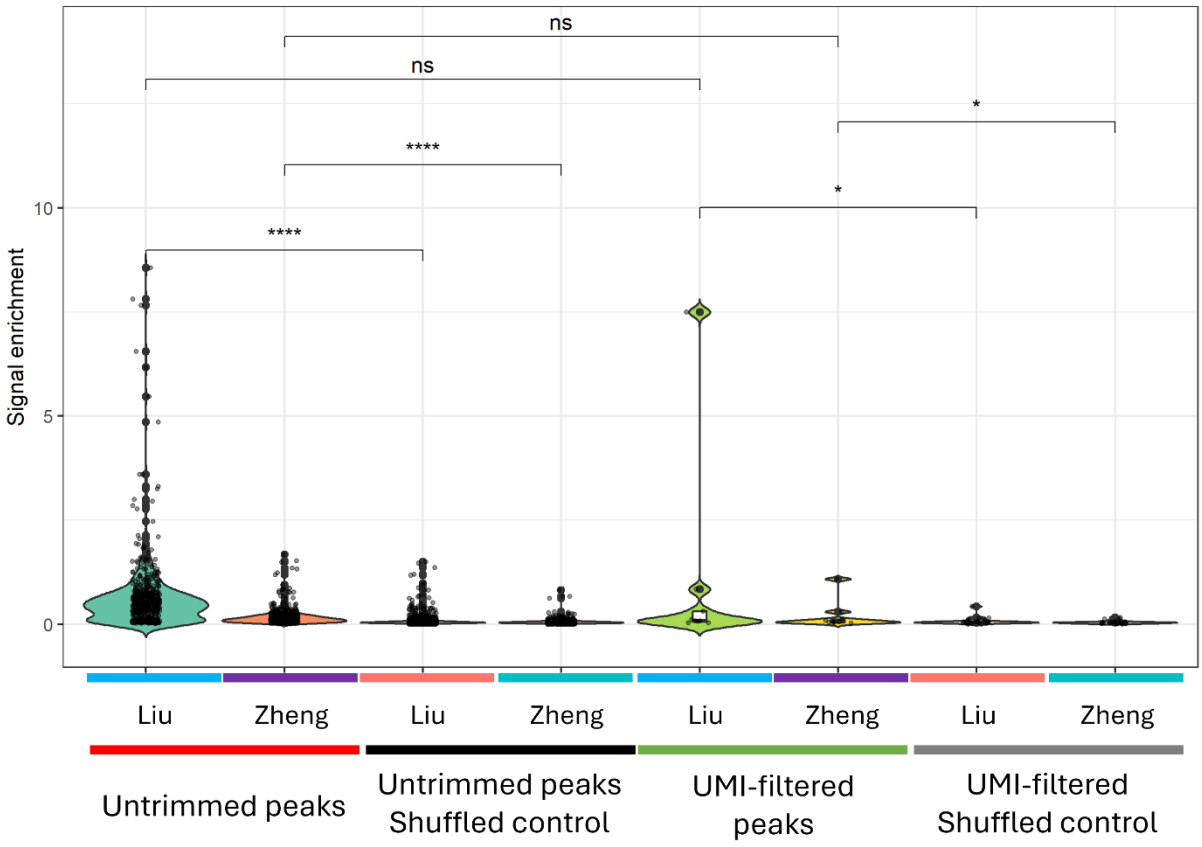


Figure 8.14: Without the removal of non-unique trimmed long-read calling card insertions, overlap with ChIP-Seq peaks and ChIA-PET anchors
Bedtools window was used to determine whether insertions from full long read peaks and qbed files overlap with existing ChIP-Seq datasets **(A)** and ChIA-PET datasets **(B)**. There is not a large change in overlap with any of the datasets, unlike full long-read calling card insertions.

8.2.7. Long-read calling cards are enriched for





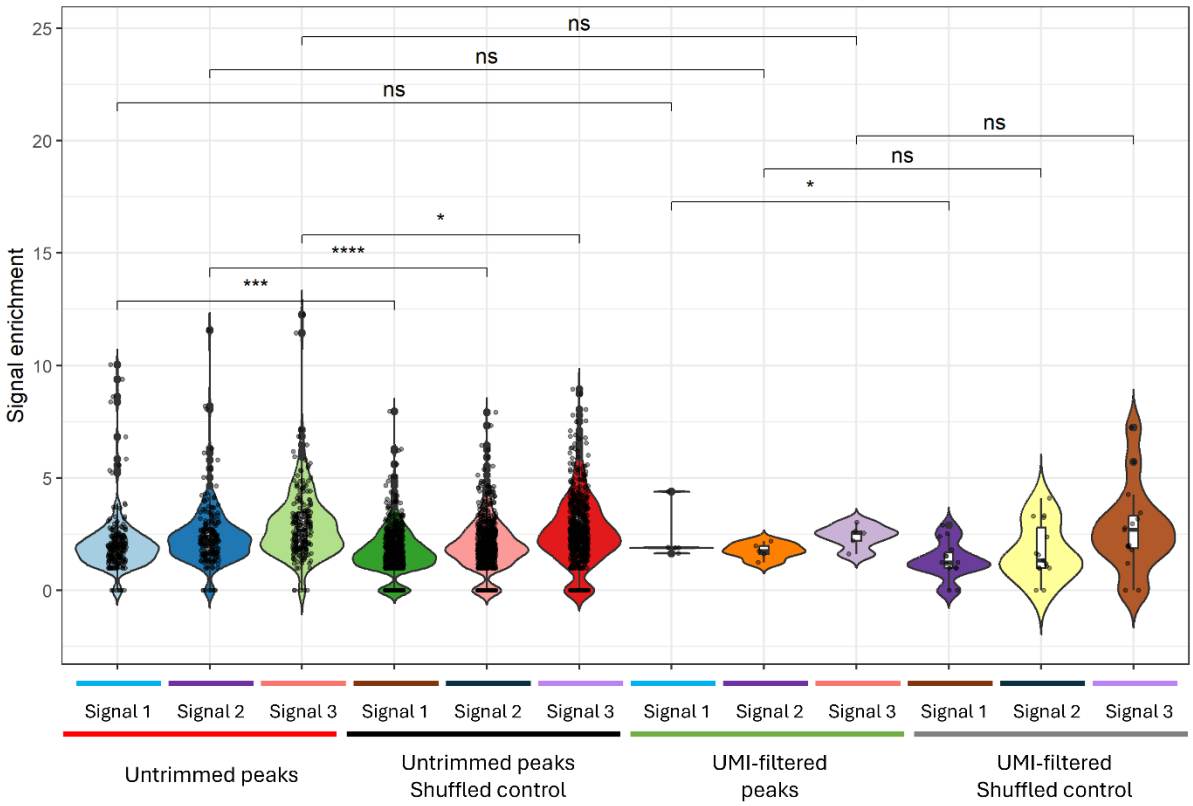


Figure 8.15: Longread HyPB-ERα calling card peaks show significant enrichment of ERα ChIP-Seq, BRD4 ChIP-Seq, and ERα ChIA-PET signal

Violin plots display mean signal intensity within 1 kb of high and low stringency long-read calling card peaks compared to matched shuffled controls. Calling card peak BED files were used as inputs for the Bed Tools shuffle function to generate random control regions of equivalent size and distribution. **(A)** ERα ChIP-Seq (ENCODE and GEO): Untrimmed peaks show significant enrichment of ENCODE signal regardless of UMI filtering, whereas GEO signal enrichment is only detected in non-filtered peaks. **(B)** BRD4 ChIP-Seq (Liu et al., 2022; Zheng et al., 2022): Long-read calling card peaks display robust enrichment of BRD4 signal, with stronger overlap observed for the Liu et al. dataset compared to Zheng et al., consistent with preferential BRD4 binding at calling card sites. **(C)** ERα ChIA-PET anchors: Both trimmed and untrimmed longread peaks show substantial enrichment of ChIA-PET signal, with UMI-filtered peaks reaching up to 77.8% overlap compared to 63.1% for non-filtered peaks. Across all datasets, Wilcoxon rank-sum tests confirm significant enrichment of longread calling card peaks relative to shuffled controls, with p-value thresholds denoted as follows: $p < 0.05$ = *, < 0.01 = **, < 0.001 = ***, < 0.0001 = ****.

8.2.8. Long read calling cards do not overlap with Bulk calling cards

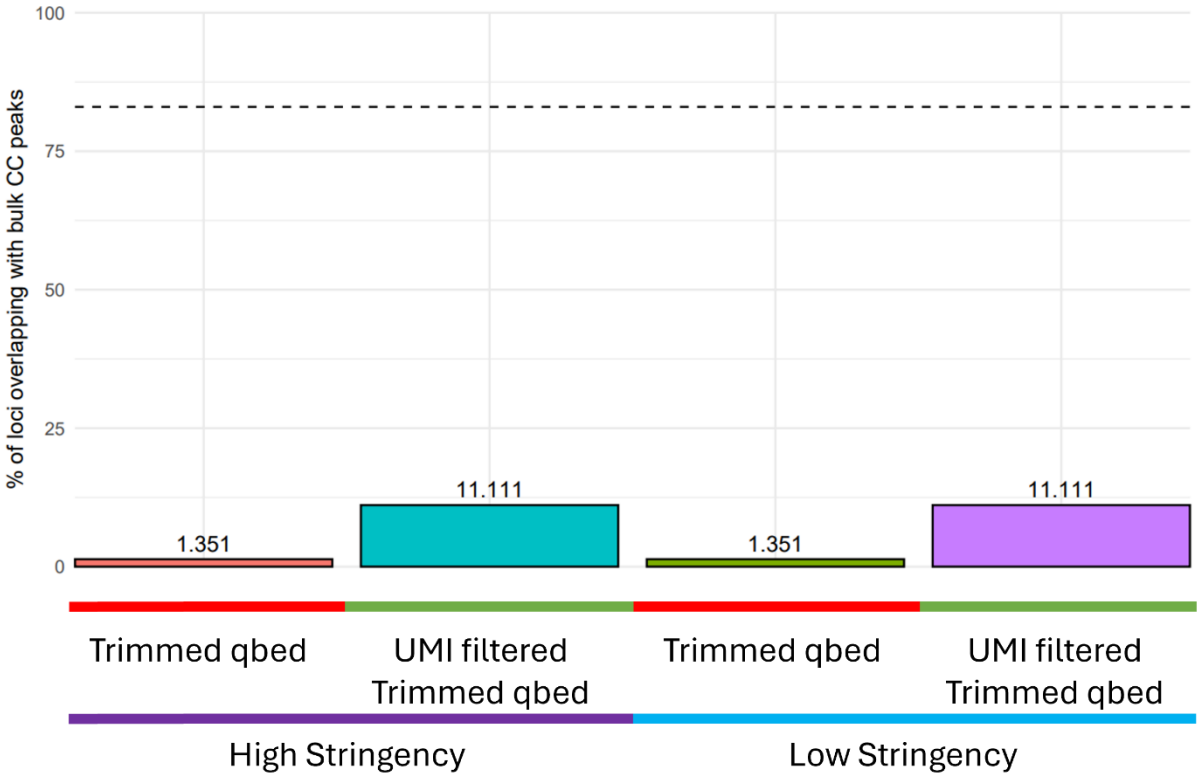


Figure 8.16: Trimmed long read HyPB-ERα directed calling card insertions only show one insertion that overlaps with bulk ERα calling card peaks.
Bedtools window was used to determine whether insertions from trimmed long read qbed files overlap with bulk calling card peaks. For the all the non-filtered insertions, across both high and low stringency peaks, only 1.35% of insertions overlapped with ChIP-Seq. UMI filtering removed insertions that didn't overlap with calling card peaks, increasing overlap to 11.11%, but this still only corresponds to one overlap event. This shows that long-read and bulk calling cards represent almost completely independent ERα binding events.

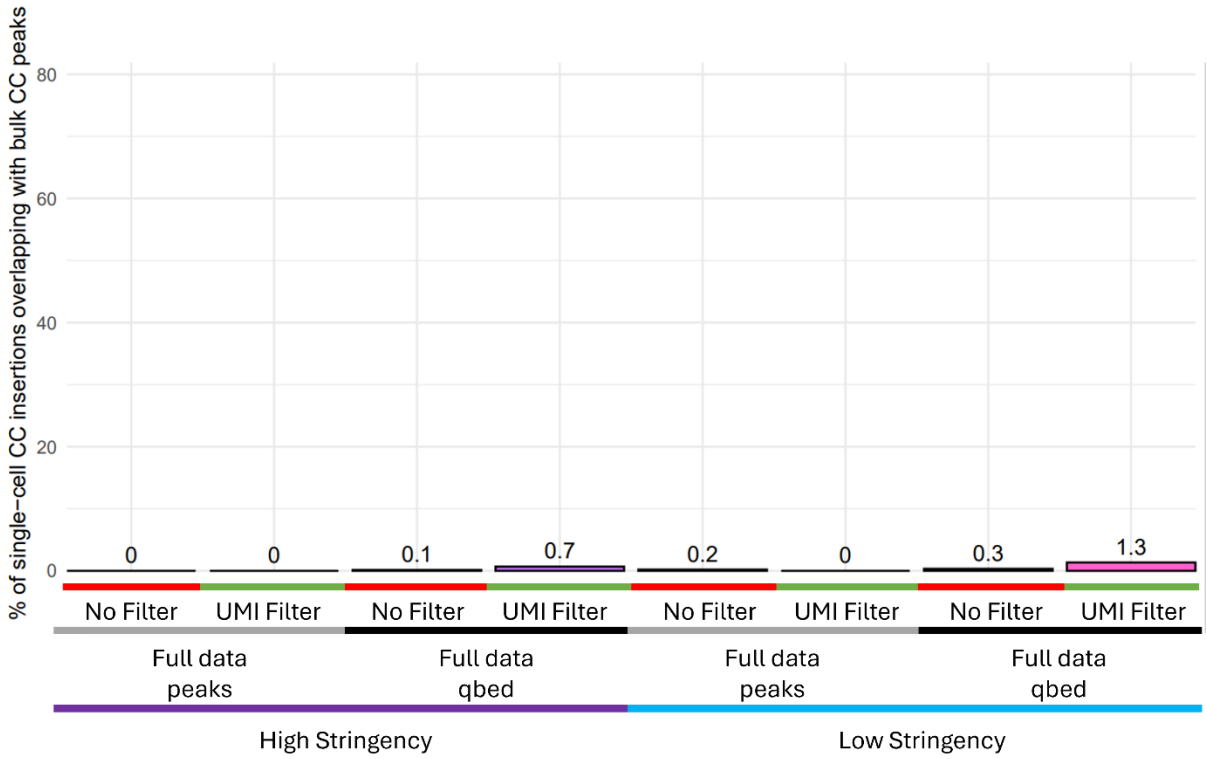


Figure 8.17: Full long read HyPB-ERα directed calling card insertions show less than 1.3 % overlap with bulk ERα calling card peals.
Bedtools window was used to determine whether insertions from full-length read qbed files overlap with bulk calling card peaks. All the files showed less than 1.3% of long read insertions overlap with existing bulk calling card peaks. This shows that long-read and bulk calling cards represent almost completely independent ERα binding events.

8.3. Chapter 5 results

8.3.1. Amiloride inhibition of EnAC in MCF7 spheroids assay plate layout

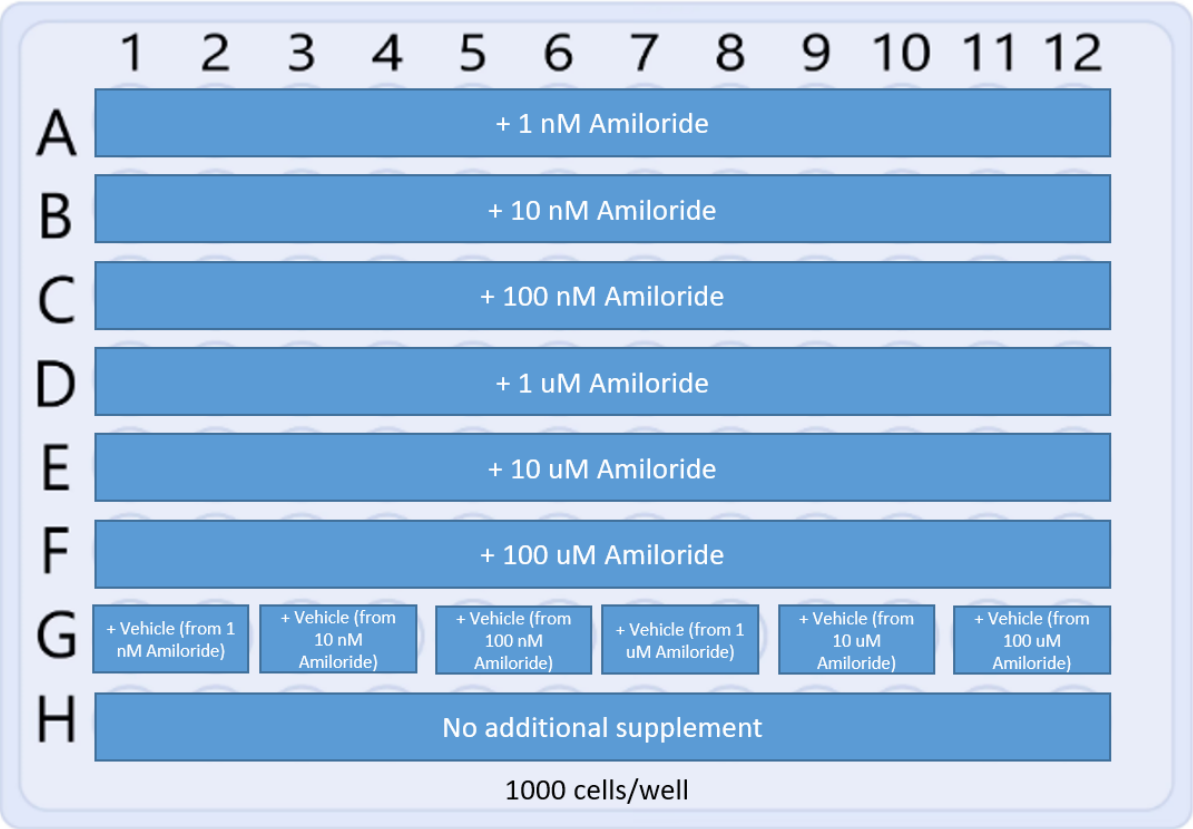


Figure 8.18: The ULA 96-well plate layout for the MCF7 spheroid amiloride treatment assay.

Spheroids grown in row A and H, as well as each sphere in columns 1 to 12, were excluded from analysis as these were found to be the smallest spheroids in the plate, and these wells exhibited more media evaporation than other wells.

8.3.2. Example images of spheroids segmented into bins by MeasureObjectIntensityDistribution in CellProfiler

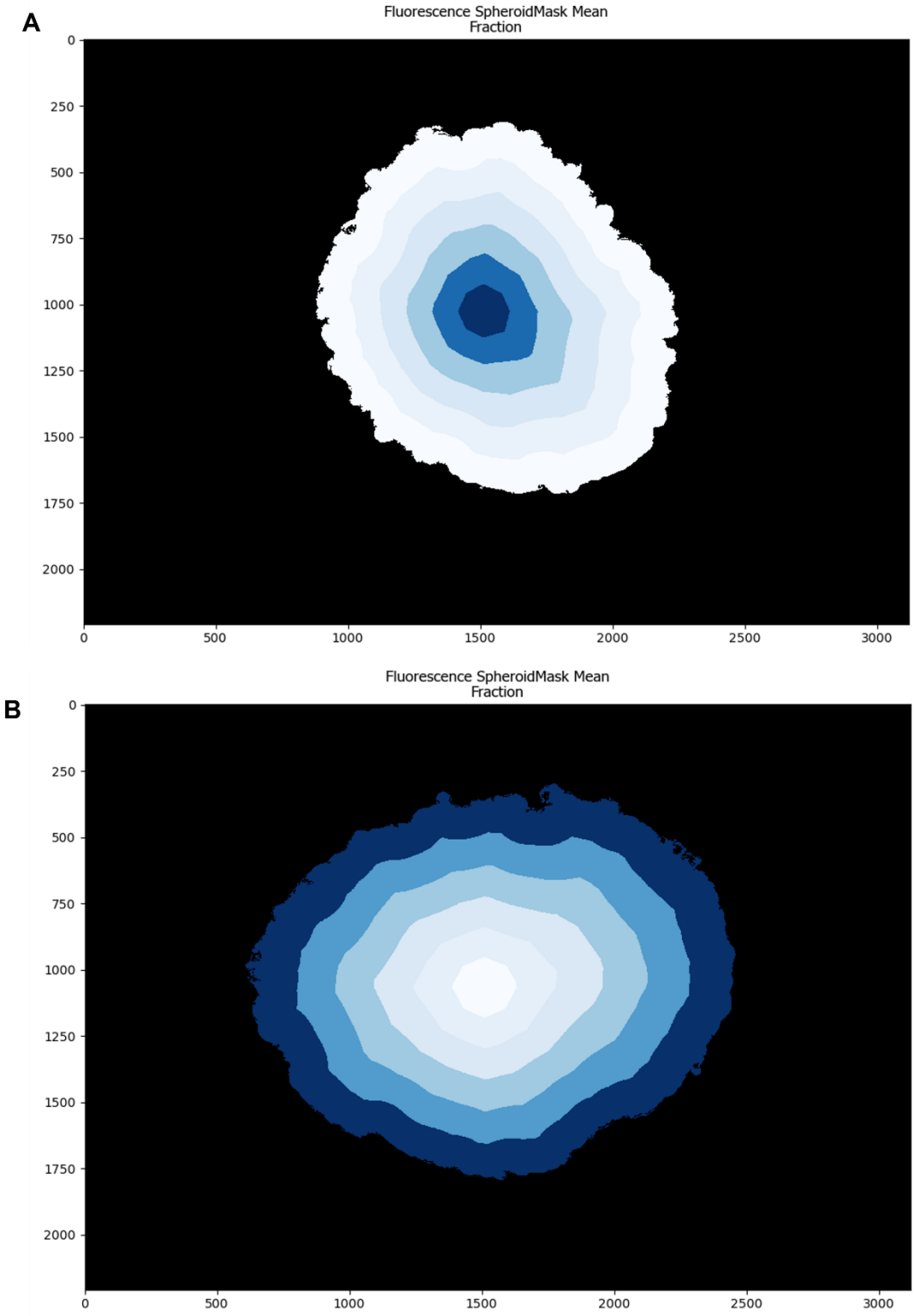
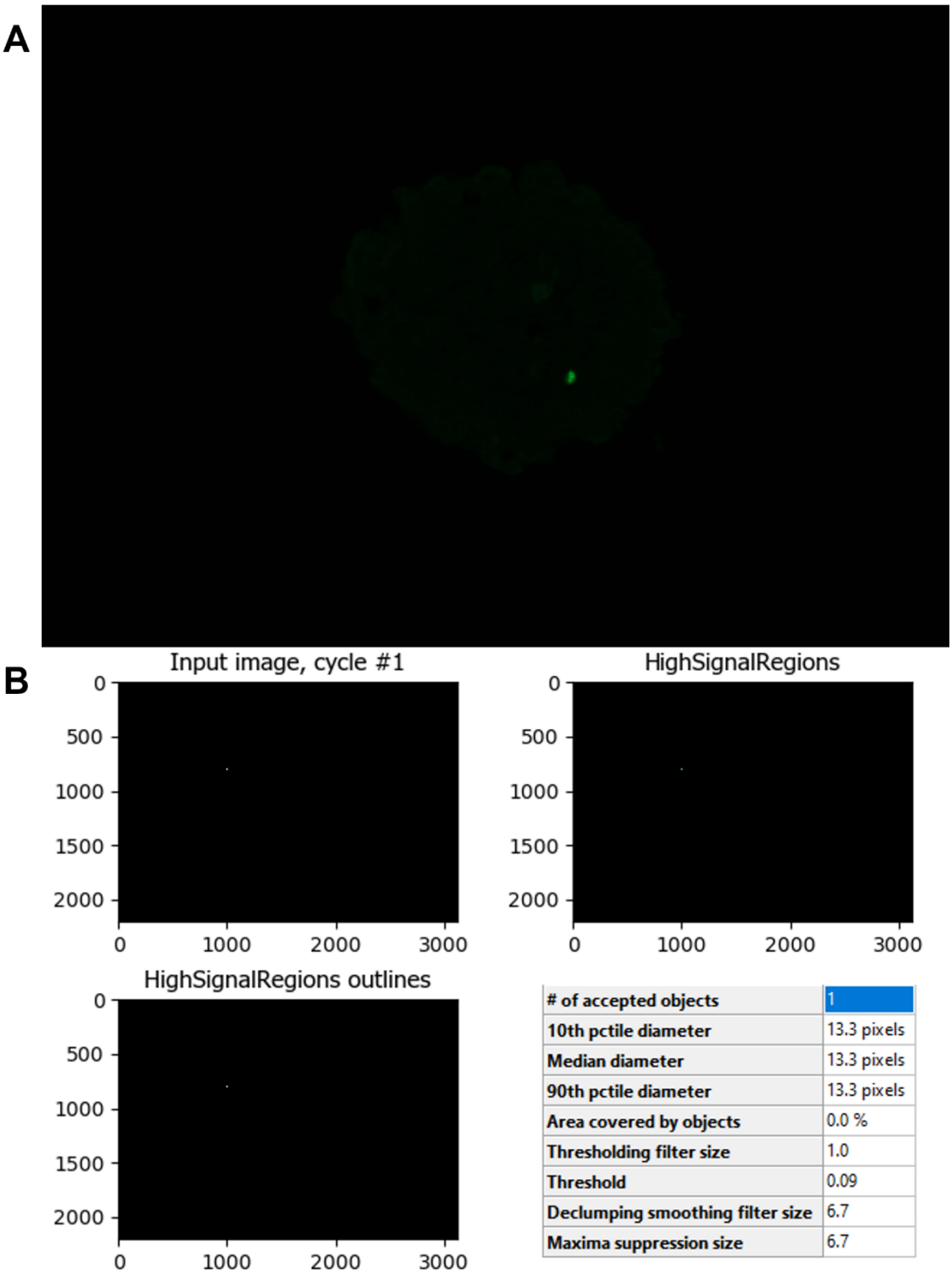


Figure 8.19: Representative example images of spheroids processed using CellProfiler's MeasureObjectIntensityDistribution function for EF5 and ERα immunostaining

CellProfiler analysis reveals EF5 **(A)** fluorescent signal originates from the core of spheroids and grows out, whereas ERα **(B)** is localised at the surface. Bins are numbered 1 to 6, with 1 in the centre of spheroids and 6 at the surface.

8.3.3. CellProfiler ‘High signal’ regions in day 7 spheroids are artefacts caused by high signal contamination



# of accepted objects	1
10th pctile diameter	13.3 pixels
Median diameter	13.3 pixels
90th pctile diameter	13.3 pixels
Area covered by objects	0.0 %
Thresholding filter size	1.0
Threshold	0.09
Declumping smoothing filter size	6.7
Maxima suppression size	6.7

Figure 8.20: Only 5 out of the 22 MCF7 spheroids harvested on day 7 show a high signal region, but each of them is caused by debris with high fluorescence in the AF488 channel.

A representative image of the anti-ERα signal of day 7 MCF7 spheroids **(A)** shows that the only high signal in the AF488 channel is derived from autofluorescence of a piece of debris. CellProfiler analysis **(B)** picks up this signal as it passes the 0.9 threshold, but does not represent genuine signal.

8.3.4. ERα CellProfiler MeanFrac plots reveal that staining is not quantitatively significant

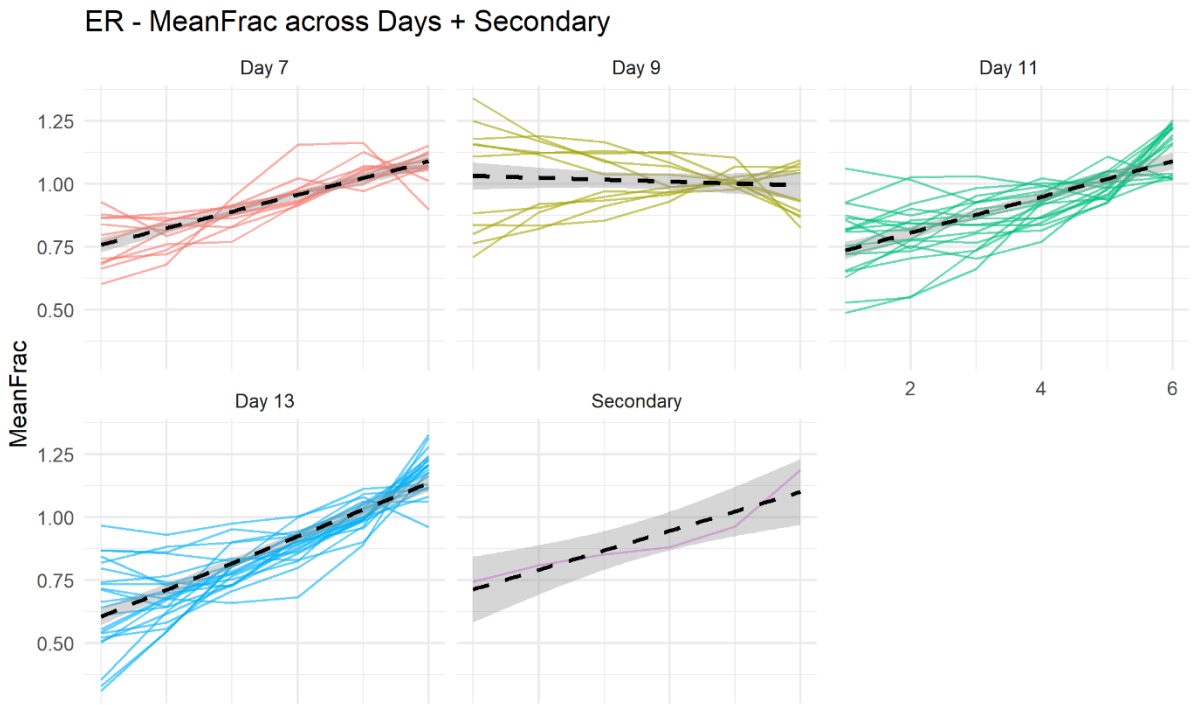


Figure 8.21: Radial distribution of ER staining in MCF7 spheroids shows surface-biased intensity but nonsignificant slopes due to high noise.

A single secondary-only control was used to assess the baseline signal. ERα-stained spheroids exhibit MeanFrac values that increase with bin size (from central bins 13 to peripheral bins 46), reflecting inherently higher edge brightness from imaging noise and background. Although qualitative staining suggests surface-localised ERα, the true signal is weak, and the resolution above noise is small; consequently, the fitted MeanFrac slopes are not significantly different from the control. Overall, ERα staining indicates a surface-biased pattern that cannot be quantified robustly under the current signal-to-noise constraints.

8.3.5. Early treatment with 100 µM amiloride affects spheroid formation from day 3 and 4 in culture

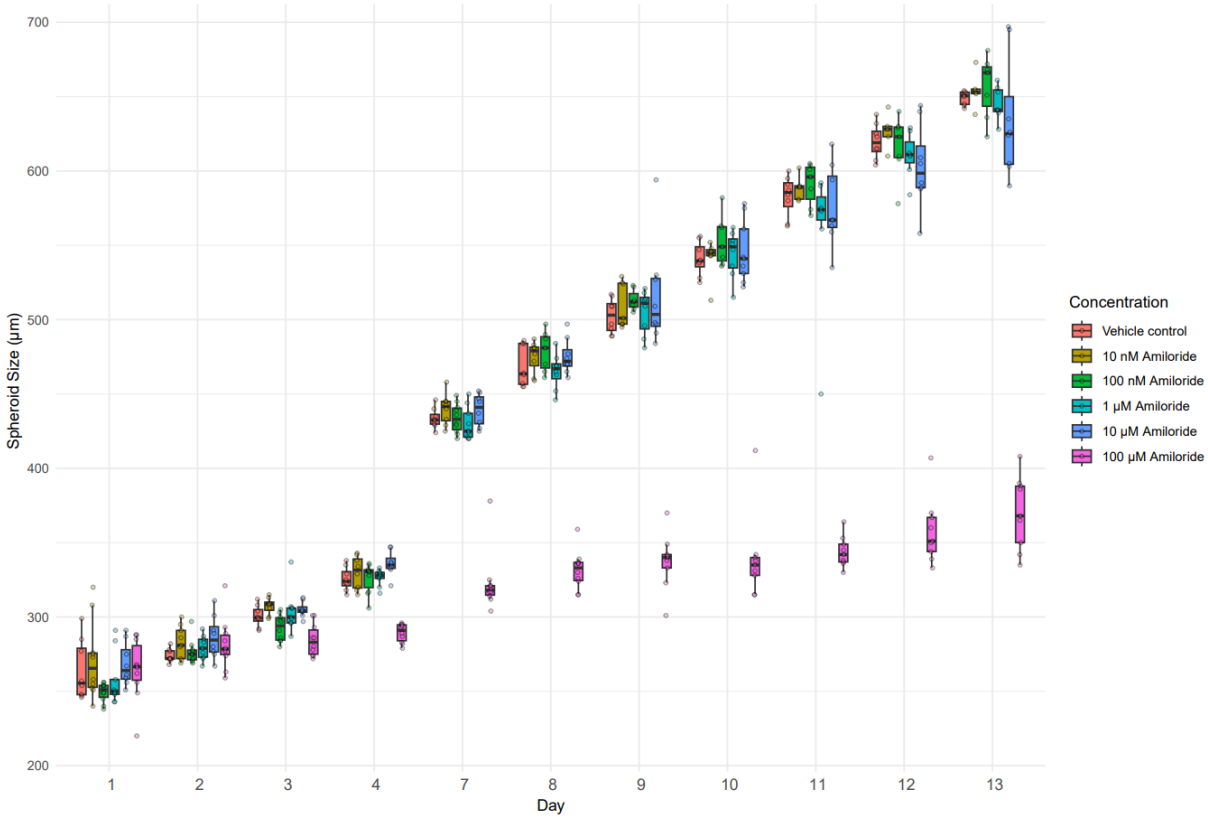


Figure 8.22: Treating MCF7 spheroids with a range of amiloride on day 2 in culture impairs spheroid formation above 10 µM.

Spheroids were treated on day 2 in culture with a range of amiloride concentrations to assess the effects of treatment on forming spheroids to mimic the treatment during tumour formation. Spheres treated with 100 µM amiloride were immediately affected by treatment as growth was arrested on day 3 and spheres diameter increased at a slower rate than other treatment conditions. The impaired growth of these spheres continued to day 13, where the large variance in diameter is due to the irregular shape of spheres. Other treatment conditions are very similar up to day 10. Here, the median of all the amiloride conditions other than 100 µM increases above the vehicle only control. At day 11, the median of both nanomolar treatments stays above the vehicle control, whereas 1 and 10 µM fall below the vehicle median. This trend continues through to day 13, although the difference between these conditions is not statistically significant. From day 7 onwards, the difference in Tukey p-value for 100 µM-treated spheres to every other condition was p = <0.0001

9. References

Abbasova, L. et al. (2025). CUT&Tag recovers up to half of ENCODE ChIP-seq histone acetylation peaks. *Nature Communications*, 16 (1), p.2993. [Accessed 13 March 2026].

Abd El-Sadek, I. et al. (2024). Label-free visualization and quantification of the drug-type-dependent response of tumor spheroids by dynamic optical coherence tomography. *Scientific reports*, 14 (1), p.3366. [Accessed 11 October 2025].

Abubakar, M. et al. (2019). Combined quantitative measures of ER, PR, HER2, and KI67 provide more prognostic information than categorical combinations in luminal breast cancer. *Modern pathology: an official journal of the United States and Canadian Academy of Pathology, Inc*, 32 (9), pp.1244–1256. [Accessed 20 October 2025].

Achilli, T.-M., Meyer, J. and Morgan, J. R. (2012). Advances in the formation, use and understanding of multi-cellular spheroids. *Expert opinion on biological therapy*, 12 (10), pp.1347–1360. [Accessed 24 October 2025].

Ahmed, S., Al-Saigh, S. and Matthews, J. (2012). FOXA1 is essential for aryl hydrocarbon receptor-dependent regulation of cyclin G2. *Molecular cancer research: MCR*, 10 (5), pp.636–648. [Accessed 21 October 2025].

Ahn, S. et al. (2020). HER2 status in breast cancer: changes in guidelines and complicating factors for interpretation. *Journal of pathology and translational medicine*, 54 (1), pp.34–44. [Accessed 26 November 2025].

Ahn, S. G. et al. (2022). Primary endocrine resistance of ER+ breast cancer with ESR1 mutations interrogated by droplet digital PCR. *npj Breast Cancer 2022 8:1*, 8 (1), pp.1–7.

Akhoundova, D. and Rubin, M. A. (2022). Clinical application of advanced multi-omics tumor profiling: Shaping precision oncology of the future. *Cancer cell*, 40 (9), pp.920–938. [Accessed 3 November 2025].

Akhtar, J. et al. (2019). TAF-ChIP: an ultra-low input approach for genome-wide chromatin immunoprecipitation assay. *Life science alliance*, 2 (4), p.e201900318. [Accessed 25 October 2025].

Al-Ani, A. et al. (2018). Oxygenation in cell culture: Critical parameters for reproducibility are routinely not reported. *PloS one*, 13 (10), p.e0204269. [Accessed 21 November 2025].

Albadari, N., Deng, S. and Li, W. (2019). The transcriptional factors HIF-1 and HIF-2 and their novel inhibitors in cancer therapy. *Expert opinion on drug discovery*, 14 (7), pp.667–682. [Accessed 23 October 2025].

Albergaria, A. et al. (2009). Expression of FOXA1 and GATA-3 in breast cancer: the prognostic significance in hormone receptor-negative tumours. *Breast cancer research: BCR*, 11 (3), p.R40. [Accessed 21 October 2025].

Alhawarat, F. M. et al. (2019). The effect of cycling hypoxia on MCF-7 cancer stem cells and the impact of their microenvironment on angiogenesis using human umbilical vein endothelial cells (HUVECs) as a model. *PeerJ*, 7 (e5990), p.e5990. [Accessed 8 October 2025].

Ali, H. R. et al. (2014). Genome-driven integrated classification of breast cancer validated in over 7,500 samples. *Genome biology*, 15 (8), p.431. [Accessed 15 November 2025].

Ali, S. et al. (1993). Modulation of transcriptional activation by ligand-dependent phosphorylation of the human oestrogen receptor A/B region. *The EMBO journal*, 12 (3), pp.1153–1160. [Accessed 21 October 2025].

Amemiya, H. M., Kundaje, A. and Boyle, A. P. (2019). The ENCODE blacklist: Identification of problematic regions of the genome. *Scientific reports*, 9 (1), p.9354. [Accessed 27 August 2025].

Amirfallah, A. et al. (2019). High expression of the vacuole membrane protein 1 (VMP1) is a potential marker of poor prognosis in HER2 positive breast cancer. *PloS one*, 14 (8), p.e0221413. [Accessed 14 September 2025].

Amiry, N. et al. (2009). Trefoil factor-1 (TFF1) enhances oncogenicity of mammary carcinoma cells. *Endocrinology*, 150 (10), pp.4473–4483. [Accessed 21 October 2025].

Anbalagan, M. et al. (2012). Post-translational modifications of nuclear receptors and human disease. *Nuclear receptor signaling*, 10, p.e001. [Accessed 26 November 2025].

André, F. et al. (2019). Alpelisib for PIK3CA-mutated, hormone receptor-positive advanced breast cancer. *The New England journal of medicine*, 380 (20), pp.1929–1940. [Accessed 23 October 2025].

Andreev, D. E. et al. (2022). Non-AUG translation initiation in mammals. *Genome biology*, 23 (1), p.111. [Accessed 24 November 2025].

Angus, S. P. et al. (2021). FOXA1 and adaptive response determinants to HER2 targeted therapy in TBCRC 036. *NPJ breast cancer*, 7 (1), p.51. [Accessed 6 November 2025].

Apweiler, R. et al. (2004). UniProt: The Universal Protein knowledgebase. *Nucleic acids research*, 32 (Database issue), pp.D115-9. [Accessed 18 November 2025].

Arai, R. et al. (2001). Design of the linkers which effectively separate domains of a bifunctional fusion protein. *Protein engineering*, 14 (8), pp.529–532. [Accessed 24 November 2025].

Arao, Y. et al. (2013). Estrogen receptor α L543A,L544A mutation changes antagonists to agonists, correlating with the ligand binding domain dimerization associated with DNA binding activity. *The journal of biological chemistry*, 288 (29), pp.21105–21116. [Accessed 26 August 2025].

Arao, Y. and Korach, K. S. (2018). The F domain of estrogen receptor α is involved in species-specific, tamoxifen-mediated transactivation. *The journal of biological chemistry*, 293 (22), pp.8495–8507. [Accessed 21 October 2025].

Arao, Y. and Korach, K. S. (2021). The physiological role of estrogen receptor functional domains. *Essays in biochemistry*, 65 (6), pp.867–875. [Accessed 21 October 2025].

Arnold, M. et al. (2022). Current and future burden of breast cancer: Global statistics for 2020 and 2040. *Breast (Edinburgh, Scotland)*, 66, pp.15–23.

Arora, S. et al. (2024). Spheroids in cancer research: Recent advances and opportunities. *Journal of drug delivery science and technology*, 100 (106033), p.106033. [Accessed 24 October 2025].

Arpino, G. et al. (2008). Crosstalk between the Estrogen Receptor and the HER Tyrosine Kinase Receptor Family: Molecular Mechanism and Clinical Implications for Endocrine Therapy Resistance. *Endocrine Reviews*, 29 (2), pp.217–233. [Accessed 25 November 2025].

Arrigoni, L. et al. (2018). RELACS nuclei barcoding enables high-throughput ChIP-seq. *Communications Biology 2018 1:1*, 1 (1), pp.1–12.

Aryan, L. et al. (2020). The role of estrogen receptors in cardiovascular disease. *International journal of molecular sciences*, 21 (12), p.4314. [Accessed 31 October 2025].

Asano, A. et al. (2018). Intracellular hypoxia measured by 18F-fluoromisonidazole positron emission tomography has prognostic impact in patients with estrogen receptor-positive breast cancer. *Breast cancer research: BCR*, 20 (1), p.78. [Accessed 11 October 2025].

Asbury, R. F. et al. (1981). Treatment of metastatic breast cancer with aminoglutethimide. *Cancer*, 47 (8), pp.1954–1958. [Accessed 22 October 2025].

Asselin-Labat, M.-L. et al. (2007). Gata-3 is an essential regulator of mammary-gland morphogenesis and luminal-cell differentiation. *Nature cell biology*, 9 (2), pp.201–209. [Accessed 21 October 2025].

de Azambuja, E. et al. (2007). Ki-67 as prognostic marker in early breast cancer: a meta-analysis of published studies involving 12,155 patients. *British journal of cancer*, 96 (10), pp.1504–1513. [Accessed 25 November 2025].

Azizi, E. et al. (2018). Single-cell map of diverse immune phenotypes in the breast tumor microenvironment. *Cell*, 174 (5), pp.1293-1308.e36. [Accessed 23 November 2025].

Bai, R. et al. (2022). The hypoxia-driven crosstalk between tumor and tumor-associated macrophages: mechanisms and clinical treatment strategies. *Molecular cancer*, 21 (1), p.177. [Accessed 23 October 2025].

Bailey, T. et al. (2013). Practical guidelines for the comprehensive analysis of ChIP-seq data. *PLoS computational biology*, 9 (11), p.e1003326. [Accessed 3 November 2025].

Bains, M., Cousins, J. C. and Roberts, J. L. (2007). Neuroprotection by estrogen against MPP+-induced dopamine neuron death is mediated by ERalpha in primary cultures of mouse mesencephalon. *Experimental neurology*, 204 (2), pp.767–776. [Accessed 31 October 2025].

Baloglu, E. et al. (2020). The role of hypoxia-induced modulation of alveolar epithelial Na+-transport in hypoxemia at high altitude. *Pulmonary circulation*, 10 (1 Suppl), pp.50–58. [Accessed 21 November 2025].

Banerji, S. et al. (2012). Sequence analysis of mutations and translocations across breast cancer subtypes. *Nature*, 486 (7403), pp.405–409. [Accessed 21 October 2025].

Baquero, A. F. and Gilbertson, T. A. (2011). Insulin activates epithelial sodium channel (ENaC) via phosphoinositide 3-kinase in mammalian taste receptor cells. *American journal of physiology. Cell physiology*, 300 (4), pp.C860-71. [Accessed 21 November 2025].

Bardet, A. F. et al. (2013). Identification of transcription factor binding sites from ChIP-seq data at high resolution. *Bioinformatics (Oxford, England)*, 29 (21), pp.2705–2713. [Accessed 25 October 2025].

Bardin, A. et al. (2004). Loss of ERbeta expression as a common step in estrogen-dependent tumor progression. *Endocrine-related cancer*, 11 (3), pp.537–551. [Accessed 21 October 2025].

Baron, S. et al. (2007). Estrogen receptor alpha and the activating protein-1 complex cooperate during insulin-like growth factor-I-induced transcriptional activation of the pS2/TFF1 gene. *The journal of biological chemistry*, 282 (16), pp.11732–11741. [Accessed 17 September 2025].

Barroso-Sousa, R. et al. (2016). Neoadjuvant endocrine therapy in breast cancer: current role and future perspectives. *ecancermedicalscience*, 10, p.609. [Accessed 22 October 2025].

Barshad, G. and Danko, C. G. (2025). Revisiting models of enhancer-promoter communication in gene regulation. *Genome Research*, 35 (6), pp.1277–1286. [Accessed 13 March 2026].

Barski, A. et al. (2007). High-resolution profiling of histone methylations in the human genome. *Cell*, 129 (4), pp.823–837. [Accessed 25 October 2025].

Bartlett, D. A. et al. (2021). High-throughput single-cell epigenomic profiling by targeted insertion of promoters (TIP-seq). *The journal of cell biology*, 220 (12). [Online]. Available at: doi:10.1083/jcb.202103078 [Accessed 26 October 2025].

Bartosovic, M. and Castelo-Branco, G. (2023). Multimodal chromatin profiling using nanobody-based single-cell CUT&Tag. *Nature biotechnology*, 41 (6), pp.794–805. [Accessed 25 November 2025].

Bartosovic, M., Kabbe, M. and Castelo-Branco, G. (2021). Single-cell CUT&Tag profiles histone modifications and transcription factors in complex tissues. *Nature biotechnology*, 39 (7), pp.825–835. [Accessed 24 October 2025].

Baselga, J. et al. (2012). Pertuzumab plus trastuzumab plus docetaxel for metastatic breast cancer. *The New England journal of medicine*, 366 (2), pp.109–119. [Accessed 26 November 2025].

Batie, M. et al. (2019). Hypoxia induces rapid changes to histone methylation and reprograms chromatin. *Science (New York, N.Y.)*, 363 (6432), pp.1222–1226. [Accessed 23 November 2025].

Batie, M. et al. (2022). Regulation of chromatin accessibility by hypoxia and HIF. *The Biochemical journal*, 479 (6), pp.767–786. [Accessed 25 November 2025].

Batie, M. and Rocha, S. (2019). JmjC histone demethylases act as chromatin oxygen sensors. *Molecular & cellular oncology*, 6 (4), p.1608501. [Accessed 23 November 2025].

Baum, M. et al. (2002). Anastrozole alone or in combination with tamoxifen versus tamoxifen alone for adjuvant treatment of postmenopausal women with early breast cancer: first results of the ATAC randomised trial. *Lancet*, 359 (9324), pp.2131–2139. [Accessed 22 October 2025].

Beckers, C. et al. (2025). Live-cell imaging and analysis of 3D spheroids in hypoxia- and radiotherapy-related research. *Clinical and translational radiation oncology*, 51 (100920), p.100920. [Accessed 8 October 2025].

Beishline, K. and Azizkhan-Clifford, J. (2015). Sp1 and the 'hallmarks of cancer'. *The FEBS journal*, 282 (2), pp.224–258. [Accessed 9 November 2025].

Belachew, E. B. and Sewasew, D. T. (2021). Molecular mechanisms of endocrine resistance in estrogen-positive breast cancer. *Frontiers in endocrinology*, 12, p.599586. [Accessed 23 October 2025].

Bell, H. S. et al. (2001). The development of necrosis and apoptosis in glioma: experimental findings using spheroid culture systems: Apoptosis and necrosis in glioma spheroids. *Neuropathology and applied neurobiology*, 27 (4), pp.291–304. [Accessed 24 November 2025].

Benecke, A., Chambon, P. and Gronemeyer, H. (2000). Synergy between estrogen receptor alpha activation functions AF1 and AF2 mediated by transcription intermediary factor TIF2. *EMBO reports*, 1 (2), pp.151–157. [Accessed 21 October 2025].

Benitez-Cantos, M. S. et al. (2020). Translation initiation downstream from annotated start codons in human mRNAs coevolves with the Kozak context. *Genome research*, 30 (7), pp.974–984. [Accessed 24 November 2025].

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society series b-methodological*, 57 (1), pp.289–300. [Accessed 4 November 2025].

Berg, A. L. et al. (2022). The cationic amphiphilic drug hexamethylene amiloride eradicates bulk breast cancer cells and therapy-resistant subpopulations with similar efficiencies. *Cancers*, 14 (4), p.949. [Accessed 21 November 2025].

Bernardo, G. M. et al. (2010). FOXA1 is an essential determinant of ERalpha expression and mammary ductal morphogenesis. *Development (Cambridge, England)*, 137 (12), pp.2045–2054. [Accessed 21 October 2025].

Bernardo, G. M. and Keri, R. A. (2012). FOXA1: a transcription factor with parallel functions in development and cancer. *Bioscience reports*, 32 (2), pp.113–130. [Accessed 26 August 2025].

Beumers, L. et al. (2023). Clonal heterogeneity in ER+ breast cancer reveals the proteasome and PKC as potential therapeutic targets. *NPJ breast cancer*, 9 (1), p.97. [Accessed 23 October 2025].

Bevan, C. L. et al. (1999). The AF1 and AF2 domains of the androgen receptor interact with distinct regions of SRC1. *Molecular and cellular biology*, 19 (12), pp.8383–8392. [Accessed 21 October 2025].

Bhattacharya, S., Calar, K. and de la Puente, P. (2020). Mimicking tumor hypoxia and tumor-immune interactions employing three-dimensional in vitro models. *Journal of experimental & clinical cancer research: CR*, 39 (1), p.75. [Accessed 24 October 2025].

Bidard, F.-C. et al. (2022). Elacestrant (oral selective estrogen receptor degrader) Versus Standard Endocrine Therapy for Estrogen Receptor–Positive, Human Epidermal Growth Factor Receptor 2–Negative Advanced Breast Cancer: Results From the Randomized Phase III EMERALD Trial. *Journal of Clinical Oncology*, 40 (28), pp.3246–3256. [Accessed 25 November 2025].

Biju, T. S., Priya, V. V. and Francis, A. P. (2023). Role of three-dimensional cell culture in therapeutics and diagnostics: an updated review. *Drug delivery and translational research*, 13 (9), pp.2239–2253. [Accessed 3 November 2025].

Billon-Galés, A. et al. (2011). Activation function 2 (AF2) of estrogen receptor-alpha is required for the atheroprotective action of estradiol but not to accelerate endothelial healing. *Proceedings of the National Academy of Sciences of the United States of America*, 108 (32), pp.13311–13316. [Accessed 21 October 2025].

Bischoff, H., Espié, M. and Petit, T. (2024). Neoadjuvant therapy: Current landscape and future horizons for ER-positive/HER2-negative and triple-negative early breast cancer. *Current treatment options in oncology*, 25 (9), pp.1210–1224. [Accessed 22 October 2025].

Björnström, L. and Sjöberg, M. (2005). Mechanisms of Estrogen Receptor Signaling: Convergence of Genomic and Nongenomic Actions on Target Genes. *Molecular endocrinology* , 19 (4), pp.833–842.

Blanco-Fernandez, B. et al. (2021). Proteinaceous hydrogels for bioengineering advanced 3D tumor models. *Advanced science (Weinheim, Baden-Wurttemberg, Germany)*, 8 (4), p.2003129. [Accessed 24 October 2025].

Boér, K. (2017). Fulvestrant in advanced breast cancer: evidence to date and place in therapy. *Therapeutic advances in medical oncology*, 9 (7), pp.465–479. [Accessed 22 October 2025].

Boyd, J. R. et al. (2023). peaksat: an R package for ChIP-seq peak saturation analysis. *BMC genomics*, 24 (1), p.43. [Accessed 23 November 2025].

Bray, F. et al. (2024). Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*, 74 (3), pp.229–263. [Accessed 20 October 2025].

Breast International Group (BIG) 1-98 Collaborative Group et al. (2005). A comparison of letrozole and tamoxifen in postmenopausal women with early breast cancer. *The New England journal of medicine*, 353 (26), pp.2747–2757. [Accessed 22 October 2025].

Bredell, M. G. et al. (2016). Current relevance of hypoxia in head and neck cancer. *Oncotarget*, 7 (31), pp.50781–50804. [Accessed 23 October 2025].

Breslin, S. and O'Driscoll, L. (2016). The relevance of using 3D cell cultures, in addition to 2D monolayer cultures, when evaluating breast cancer drug sensitivity and resistance. *Oncotarget*, 7 (29), pp.45745–45756. [Accessed 13 March 2026].

Brett, J. O. et al. (2021). ESR1 mutation as an emerging clinical biomarker in metastatic hormone receptor-positive breast cancer. *Breast cancer research: BCR*, 23 (1), p.85. [Accessed 23 October 2025].

Bridges, E. and Harris, A. L. (2015). Vascular-promoting therapy reduced tumor growth and progression by improving chemotherapy efficacy. *Cancer cell*, 27 (1), pp.7–9. [Accessed 24 October 2025].

Brind'Amour, J. et al. (2015). An ultra-low-input native ChIP-seq protocol for genome-wide profiling of rare cell populations. *Nature Communications 2015 6:1*, 6 (1), pp.1–8.

Brooks, S. C., Locke, E. R. and Soule, H. D. (1973). Estrogen receptor in a human cell line (MCF-7) from breast carcinoma. *The journal of biological chemistry*, 248 (17), pp.6251–6253. [Accessed 26 August 2025].

Bross, P. F. et al. (2002). FDA drug approval summaries: fulvestrant. *The oncologist*, 7 (6), pp.477–480. [Accessed 22 October 2025].

Bross, P. F. et al. (2003). Fulvestrant in postmenopausal women with advanced breast cancer. *Clinical cancer research: an official journal of the American Association for Cancer Research*, 9 (12), pp.4309–4317. [Accessed 22 October 2025].

Browne, A. L. et al. (2018). Network analysis of SRC-1 reveals a novel transcription factor hub which regulates endocrine resistant breast cancer. *Oncogene*, 37 (15), pp.2008–2021. [Accessed 23 October 2025].

Brzozowski, A. M. et al. (1997). Molecular basis of agonism and antagonism in the oestrogen receptor. *Nature*, 389 (6652), pp.753–758. [Accessed 20 October 2025].

Buenrostro, J. D. et al. (2015). Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature*, 523 (7561), pp.486–490. [Accessed 23 November 2025].

Burns, K. A. et al. (2014). Research resource: comparison of gene profiles from wild-type ERα and ERα hinge region mutants. *Molecular endocrinology (Baltimore, Md.)*, 28 (8), pp.1352–1361. [Accessed 21 October 2025].

Burstein, H. J. et al. (2014). Adjuvant endocrine therapy for women with hormone receptor-positive breast cancer: american society of clinical oncology clinical practice guideline focused update. *Journal of clinical oncology: official journal of the American Society of Clinical Oncology*, 32 (21), pp.2255–2269. [Accessed 22 October 2025].

Burstein, M. D. et al. (2015). Comprehensive Genomic Analysis Identifies Novel Subtypes and Targets of Triple-Negative Breast Cancer. *Clinical cancer research: an official journal of the American Association for Cancer Research*, 21 (7), pp.1688–1698.

Bushnell, G. G. et al. (2021). Breast cancer dormancy: need for clinically relevant models to address current gaps in knowledge. *NPJ breast cancer*, 7 (1), p.66. [Accessed 15 November 2025].

Cagnet, S. et al. (2018). Oestrogen receptor α AF-1 and AF-2 domains have cell population-specific functions in the mammary epithelium. *Nature communications*, 9 (1), p.4723. [Accessed 21 October 2025].

Cairns, R., Harris, I. and Mak, T. (2011). Regulation of cancer cell metabolism. *Nature reviews. Cancer*, 11, pp.85–95.

Cammack, A. J. et al. (2020). A viral toolkit for recording transcription factor–DNA interactions in live mouse tissues. *Proceedings of the National Academy of Sciences*, 117 (18), pp.10003–10014.

Campagne, A. et al. (2019). BAP1 complex promotes transcription by opposing PRC1-mediated H2A ubiquitylation. *Nature communications*, 10 (1), p.348. [Accessed 24 November 2025].

Cancer Genome Atlas Network. (2012). Comprehensive molecular portraits of human breast tumours. *Nature*, 490 (7418), pp.61–70. [Accessed 25 November 2025].

Canhasi, L., Tina, E. and Eremo, A. G. (2023). Hypoxia-mimetic by CoCl2 increases SLC7A5 expression in breast cancer cells in vitro. *BMC research notes*, 16 (1), p.366. [Accessed 23 November 2025].

Cao, G. et al. (2025). High-throughput optimization of peptide-linker for fusing function protein with GFP. *Protein expression and purification*, 231 (106718), p.106718. [Accessed 10 August 2025].

Cao, Z. et al. (2015). A microfluidic device for epigenomic profiling using 100 cells. *Nature methods*, 12 (10), pp.959–962.

Capatina, A. L. et al. (2024). Hypoxia-induced epigenetic regulation of breast cancer progression and the tumour microenvironment. *Frontiers in cell and developmental biology*, 12, p.1421629. [Accessed 24 October 2025].

Carmeci, C. et al. (1997). Identification of a gene (GPR30) with homology to the G-protein-coupled receptor superfamily associated with estrogen receptor expression in breast cancer. *Genomics*, 45 (3), pp.607–617. [Accessed 20 October 2025].

Carroll, J. S. et al. (2005). Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. *Cell*, 122 (1), pp.33–43. [Accessed 7 November 2025].

Carroll, J. S. et al. (2006). Genome-wide analysis of estrogen receptor binding sites. *Nature genetics*, 38 (11), pp.1289–1297. [Accessed 25 October 2025].

Carter, C. L., Allen, C. and Henson, D. E. (1989). Relation of tumor size, lymph node status, and survival in 24,740 breast cancer cases. *Cancer*, 63 (1), pp.181–187. [Accessed 25 November 2025].

Caswell-Jin, J. L., Lorenz, C. and Curtis, C. (2021). Molecular heterogeneity and evolution in breast cancer. *Annual review of cancer biology*, 5 (1), pp.79–94. [Accessed 20 October 2025].

Cavo, M. et al. (2018). A new cell-laden 3D Alginate-Matrigel hydrogel resembles human breast cancer cell malignant morphology, spread and invasion capability observed 'in vivo'. *Scientific reports*, 8 (1), p.5333. [Accessed 11 October 2025].

Chacón, R. D. and Costanzo, M. V. (2010). Triple-negative breast cancer. *Breast cancer research: BCR*, 12 Suppl 2 (S2), p.S3. [Accessed 26 November 2025].

Chandarlapaty, S. et al. (2016). Prevalence of ESR1 mutations in cell-free DNA and outcomes in metastatic breast cancer: A secondary analysis of the BOLERO-2 clinical trial: A secondary analysis of the BOLERO-2 clinical trial. *JAMA oncology*, 2 (10), pp.1310–1315. [Accessed 22 October 2025].

Chantalat, E. et al. (2016). The AF-1-deficient estrogen receptor ERα46 isoform is frequently expressed in human breast tumors. *Breast cancer research: BCR*, 18 (1), p.123. [Accessed 21 October 2025].

Chao, C., Ngo Le, P. and Engelward, B. P. (2020). SpheroidChip: Patterned agarose microwell compartments harboring HepG2 spheroids are compatible with genotoxicity testing. *ACS biomaterials science & engineering*, 6 (4), pp.2427–2439. [Accessed 24 October 2025].

Cheang, M. C. U. et al. (2009). Ki67 index, HER2 status, and prognosis of patients with luminal B breast cancer. *Journal of the National Cancer Institute*, 101 (10), pp.736–750. [Accessed 26 November 2025].

Chen, A. et al. (2018). Intermittent hypoxia induces a metastatic phenotype in breast cancer. *Oncogene*, 37 (31), pp.4214–4225. [Accessed 8 July 2025].

Chen, D. et al. (2000). Activation of estrogen receptor alpha by S118 phosphorylation involves a ligand-dependent interaction with TFIIH and participation of CDK7. *Molecular cell*, 6 (1), pp.127–137. [Accessed 26 November 2025].

Chen, D. et al. (2024). Stratification and prognostic evaluation of breast cancer subtypes defined by obesity-associated genes. *Discover. Oncology*, 15 (1), p.133. [Accessed 20 October 2025].

Chen, G., Liu, W. and Yan, B. (2022). Breast Cancer MCF-7 Cell Spheroid Culture for Drug Discovery and Development. *Journal of cancer therapy*, 13 (3), pp.117–130.

Chen, H. et al. (2019). Assessment of computational methods for the analysis of single-cell ATAC-seq data. *Genome biology*, 20 (1), p.241. [Accessed 23 November 2025].

Chen, P., Li, B. and Ou-Yang, L. (2022). Role of estrogen receptors in health and disease. *Frontiers in endocrinology*, 13, p.839005. [Accessed 31 October 2025].

Chen, Q. et al. (2020). Structural basis of seamless excision and specific targeting by piggyBac transposase. *Nature communications*, 11 (1), p.3446. [Accessed 15 October 2025].

Chen, S. et al. (2023). Application of single-cell sequencing to the research of tumor microenvironment. *Frontiers in immunology*, 14, p.1285540. [Accessed 15 November 2025].

Chen, X., Zaro, J. L. and Shen, W.-C. (2013). Fusion protein linkers: property, design and functionality. *Advanced drug delivery reviews*, 65 (10), pp.1357–1369. [Accessed 26 August 2025].

Cheng, G. J., Leung, E. Y. and Singleton, D. C. (2022). In vitro breast cancer models for studying mechanisms of resistance to endocrine therapy. *Exploration of targeted anti-tumor therapy*, 3 (3), pp.297–320. [Accessed 30 October 2025].

Cheng, M., Michalski, S. and Kommagani, R. (2018). Role for growth regulation by estrogen in breast cancer 1 (GREB1) in hormone-dependent cancers. *International journal of molecular sciences*, 19 (9), p.2543. [Accessed 26 August 2025].

Cho, J. et al. (2005). Cobalt chloride-induced estrogen receptor alpha down-regulation involves hypoxia-inducible factor-1alpha in MCF-7 human breast cancer cells. *Molecular endocrinology (Baltimore, Md.)*, 19 (5), pp.1191–1199. [Accessed 24 October 2025].

Choi, Y. H. and Kim, J. K. (2019). Dissecting cellular heterogeneity using single-cell RNA sequencing. *Molecules and cells*, 42 (3), pp.189–199. [Accessed 26 October 2025].

Christiansen, P. et al. (2022). Breast-conserving surgery or mastectomy?: Impact on survival. *Annals of surgery open: perspectives of surgical history, education, and clinical approaches*, 3 (4), p.e205. [Accessed 22 October 2025].

Chua, Y. L. et al. (2010). Stabilization of hypoxia-inducible factor-1alpha protein in hypoxia occurs independently of mitochondrial reactive oxygen species production. *The journal of biological chemistry*, 285 (41), pp.31277–31284. [Accessed 23 October 2025].

Cicatiello, L. et al. (2004). Estrogens and progesterone promote persistent CCND1 gene activation during G1 by inducing transcriptional derepression via c-Jun/c-Fos/estrogen receptor (progesterone receptor) complex assembly to a distal regulatory element and recruitment of cyclin D1 to its own gene promoter. *Molecular and cellular biology*, 24 (16), pp.7260–7274. [Accessed 21 October 2025].

Ciriello, G. et al. (2013). The molecular diversity of Luminal A breast tumors. *Breast cancer research and treatment*, 141 (3), pp.409–420. [Accessed 21 October 2025].

Cirillo, L. A. et al. (2002). Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Molecular cell*, 9 (2), pp.279–289. [Accessed 21 October 2025].

Clarke, R., Tyson, J. J. and Dixon, J. M. (2015). Endocrine resistance in breast cancer--An overview and update. *Molecular and cellular endocrinology*, 418 Pt 3 (0 3), pp.220–234. [Accessed 15 November 2025].

Colegio, O. R. et al. (2014). Functional polarization of tumour-associated macrophages by tumour-derived lactic acid. *Nature*, 513 (7519), pp.559–563. [Accessed 3 November 2025].

Collin, L. J. et al. (2021). Hypoxia-inducible factor-1α expression and breast cancer recurrence in a Danish population-based case control study. *Breast cancer research: BCR*, 23 (1), p.103. [Accessed 11 October 2025].

Coombes, R. C. et al. (2004). A randomized trial of exemestane after two to three years of tamoxifen therapy in postmenopausal women with primary breast cancer. *The New England journal of medicine*, 350 (11), pp.1081–1092. [Accessed 22 October 2025].

Çorbacıoğlu, Ş. K. and Aksel, G. (2023). Receiver operating characteristic curve analysis in diagnostic accuracy studies: A guide to interpreting the area under the curve value: A guide to interpreting the area under the curve value. *Turkish journal of emergency medicine*, 23 (4), pp.195–198. [Accessed 5 November 2025].

Corces, M. R. et al. (2018). The chromatin accessibility landscape of primary human cancers. *Science (New York, N.Y.)*, 362 (6413), p.eaav1898. [Accessed 24 November 2025].

Cordey, M. and Pike, C. J. (2005). Neuroprotective properties of selective estrogen receptor agonists in cultured neurons. *Brain research*, 1045 (1–2), pp.217–223. [Accessed 31 October 2025].

Cortes, J. et al. (2022). Pembrolizumab plus chemotherapy in advanced triple-negative breast cancer. *The New England journal of medicine*, 387 (3), pp.217–226. [Accessed 26 November 2025].

Costa, E. C. et al. (2016). 3D tumor spheroids: an overview on the tools and techniques used for their analysis. *Biotechnology advances*, 34 (8), pp.1427–1441.

Craig Jordan, V. et al. (2014). The evolution of nonsteroidal antiestrogens to become selective estrogen receptor modulators. *Steroids*, 90, pp.3–12. [Accessed 22 October 2025].

Curtis, C. et al. (2012). The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*, 486 (7403), pp.346–352.

Cuzick, J. et al. (2010). Effect of anastrozole and tamoxifen as adjuvant treatment for early-stage breast cancer: 10-year analysis of the ATAC trial. *The lancet oncology*, 11 (12), pp.1135–1141. [Accessed 22 October 2025].

Da Silva, L. and Lakhani, S. R. (2010). Pathology of hereditary breast cancer. *Modern pathology: an official journal of the United States and Canadian Academy of Pathology, Inc*, 23 Suppl 2 (S2), pp.S46-51. [Accessed 22 October 2025].

Dago, D. N. et al. (2015). Estrogen receptor beta impacts hormone-induced alternative mRNA splicing in breast cancer cells. *BMC genomics*, 16 (1), p.367. [Accessed 21 October 2025].

Dahl, J. A. and Gilfillan, G. D. (2018). How low can you go? Pushing the limits of low-input ChIP-seq. *Briefings in functional genomics*, 17 (2), pp.89–95. [Accessed 25 October 2025].

Dai, X. et al. (2015). Breast cancer intrinsic subtype classification, clinical use and future trends. *American journal of cancer research*, 5 (10), pp.2929–2943. [Accessed 26 November 2025].

Das, V. et al. (2016). Reproducibility of uniform spheroid formation in 384-well plates: The effect of medium evaporation: The effect of medium evaporation. *Journal of biomolecular screening*, 21 (9), pp.923–930. [Accessed 21 November 2025].

Däster, S. et al. (2017). Induction of hypoxia and necrosis in multicellular tumor spheroids is associated with resistance to chemotherapy treatment. *Oncotarget*, 8 (1), pp.1725–1736. [Accessed 24 October 2025].

Dauvois, S., White, R. and Parker, M. G. (1993). The antiestrogen ICI 182780 disrupts estrogen receptor nucleocytoplasmic shuttling. *Journal of cell science*, 106 (Pt 4) (4), pp.1377–1388. [Accessed 25 November 2025].

Davies, C. et al. (2013). Long-term effects of continuing adjuvant tamoxifen to 10 years versus stopping at 5 years after diagnosis of oestrogen receptor-positive breast cancer: ATLAS, a randomised trial. *Lancet*, 381 (9869), pp.805–816. [Accessed 22 October 2025].

De La Cruz, L. et al. (2016). Outcomes after oncoplastic breast-conserving surgery in breast cancer patients: A systematic literature review. *Annals of surgical oncology*, 23 (10), pp.3247–3258. [Accessed 22 October 2025].

De Simone, M. et al. (2025). A comprehensive analysis framework for evaluating commercial single-cell RNA sequencing technologies. *Nucleic Acids Research*, 53 (2), p.gkae1186. [Accessed 24 November 2025].

Dekker, J., Marti-Renom, M. A. and Mirny, L. A. (2013). Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. *Nature reviews. Genetics*, 14 (6), pp.390–403. [Accessed 24 November 2025].

Dengler, V. L., Galbraith, M. and Espinosa, J. M. (2014). Transcriptional regulation by hypoxia inducible factors. *Critical reviews in biochemistry and molecular biology*, 49 (1), pp.1–15. [Accessed 23 October 2025].

Denkert, C. et al. (2018). Tumour-infiltrating lymphocytes and prognosis in different subtypes of breast cancer: a pooled analysis of 3771 patients treated with neoadjuvant therapy. *The lancet oncology*, 19 (1), pp.40–50. [Accessed 23 November 2025].

Dent, R. et al. (2007). Triple-negative breast cancer: clinical features and patterns of recurrence. *Clinical cancer research: an official journal of the American Association for Cancer Research*, 13 (15 Pt 1), pp.4429–4434. [Accessed 26 November 2025].

Devailly, G. et al. (2015). Variable reproducibility in genome-scale public data: A case study using ENCODE ChIP sequencing resource. *FEBS letters*, 589 (24 Pt B), pp.3866–3870. [Accessed 16 September 2025].

Devi, S. (2025). Projected global rise in breast cancer incidence and mortality by 2050. *The lancet oncology*, 26 (4), p.417. [Accessed 20 October 2025].

Devin-Leclerc, J. et al. (1998). Interaction and dissociation by ligands of estrogen receptor and Hsp90: the antiestrogen RU 58668 induces a protein synthesis-dependent clustering of the receptor in the cytoplasm. *Molecular endocrinology (Baltimore, Md.)*, 12 (6), pp.842–854. [Accessed 21 October 2025].

Dhandapani, H. et al. (2023). In Vitro 3D Spheroid Model Preserves Tumor Microenvironment of Hot and Cold Breast Cancer Subtypes. *Advanced healthcare materials*, 12 (21), p.e2300164.

Di Giovannantonio, M. et al. (2025). Defining hypoxia in cancer: A landmark evaluation of hypoxia gene expression signatures. *Cell genomics*, 5 (2), p.100764. [Accessed 23 November 2025].

Di Tommaso, P. et al. (2017). Nextflow enables reproducible computational workflows. *Nature Biotechnology*, 35 (4), pp.316–319. [Accessed 24 November 2025].

Diaz Bessone, M. I. et al. (2019). The tumor microenvironment as a regulator of endocrine resistance in breast cancer. *Frontiers in endocrinology*, 10, p.547. [Accessed 23 October 2025].

Ding, S. et al. (2005). Efficient transposition of the piggyBac (PB) transposon in mammalian cells and mice. *Cell*, 122 (3), pp.473–483. [Accessed 30 October 2025].

Ditsch, N. et al. (2022). AGO recommendations for the Diagnosis and treatment of patients with early Breast Cancer: Update 2022. *Breast care (Basel, Switzerland)*, 17 (4), pp.403–420. [Accessed 22 October 2025].

Dizin, E. et al. (2021). Activation of the hypoxia-inducible factor pathway inhibits epithelial sodium channel-mediated sodium transport in collecting duct principal cells. *Journal of the American Society of Nephrology: JASN*, 32 (12), pp.3130–3145. [Accessed 21 November 2025].

Dong, C. et al. (2021). Activation of PI3K/AKT/mTOR pathway causes drug resistance in breast cancer. *Frontiers in pharmacology*, 12, p.628690. [Accessed 23 October 2025].

Doublier, S. et al. (2012). HIF-1 activation induces doxorubicin resistance in MCF7 3-D spheroids via P-glycoprotein expression: a potential model of the chemo-resistance of invasive micropapillary carcinoma of the breast. *BMC cancer*, 12, p.4.

Drabsch, Y. et al. (2007). Mechanism of and requirement for estrogen-regulated MYB expression in estrogen-receptor-positive breast cancer cells. *Proceedings of the National Academy of Sciences of the United States of America*, 104 (34), pp.13762–13767. [Accessed 7 November 2025].

Droog, M. et al. (2017). Estrogen receptor α wields treatment-specific enhancers between morphologically similar endometrial tumors. *Proceedings of the National Academy of Sciences of the United States of America*, 114 (8), pp.E1316–E1325. [Accessed 31 October 2025].

Dziadkowiec, K. N. et al. (2016). PARP inhibitors: review of mechanisms of action and BRCA1/2 mutation targeting. *Przeglad menopauzalny*, 15 (4), pp.215–219. [Accessed 22 October 2025].

Early Breast Cancer Trialists' Collaborative Group (EBCTCG) et al. (2012). Comparisons between different polychemotherapy regimens for early breast cancer: meta-analyses of long-term outcome among 100,000 women in 123 randomised trials. *Lancet*, 379 (9814), pp.432–444. [Accessed 22 October 2025].

Edmondson, R. et al. (2014). Three-dimensional cell culture systems and their applications in drug discovery and cell-based biosensors. *Assay and drug development technologies*, 12 (4), pp.207–218. [Accessed 3 November 2025].

Eeckhoute, J. et al. (2007). Positive cross-regulatory loop ties GATA-3 to estrogen receptor alpha expression in breast cancer. *Cancer research*, 67 (13), pp.6477–6483.

Ekert, J. E. et al. (2014). Three-dimensional lung tumor microenvironment modulates therapeutic compound responsiveness in vitro--implication for drug development. *PloS one*, 9 (3), p.e92248. [Accessed 3 November 2025].

Elston, C. W. and Ellis, I. O. (1991). Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. *Histopathology*, 19 (5), pp.403–410. [Accessed 25 November 2025].

ENCODE Project Consortium. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489 (7414), pp.57–74. [Accessed 22 August 2025].

Evans, D. M. and Sloan-Stakleff, K. (2000). Suppression of the invasive capacity of human breast cancer cells by inhibition of urokinase plasminogen activator via amiloride and B428. *The American surgeon*, 66 (5), pp.460–464. [Accessed 21 November 2025].

Fang, K. et al. (2024). Integrated single-cell analysis reveals distinct epigenetic-regulated cancer cell states and a heterogeneity-guided core signature in tamoxifen-resistant breast cancer. *Genome medicine*, 16 (1), p.134. [Accessed 3 November 2025].

Fang, S. H., Chen, Y. and Weigel, R. J. (2009). GATA-3 as a marker of hormone response in breast cancer. *The Journal of surgical research*, 157 (2), pp.290–295. [Accessed 26 August 2025].

Fanning, S. W. et al. (2016). Estrogen receptor alpha somatic mutations Y537S and D538G confer breast cancer endocrine resistance by stabilizing the activating function-2 binding conformation. *eLife*, 5, p.e12792. [Accessed 23 October 2025].

Fanning, S. W. and Greene, G. L. (2019). Next-generation ERα inhibitors for endocrine-resistant ER+ breast cancer. *Endocrinology*, 160 (4), pp.759–769. [Accessed 25 November 2025].

Farcas, A. M. et al. (2021). Genome-wide estrogen receptor activity in breast cancer. *Endocrinology*, 162 (2), p.bqaa224. [Accessed 25 November 2025].

Feeley, L. P. et al. (2014). Distinguishing luminal breast cancer subtypes by Ki67, progesterone receptor or TP53 status provides prognostic information. *Modern pathology: an official journal of the United States and Canadian Academy of Pathology, Inc*, 27 (4), pp.554–561. [Accessed 26 November 2025].

Feng, J. et al. (2012). Identifying ChIP-seq enrichment using MACS. *Nature protocols*, 7 (9), pp.1728–1740. [Accessed 15 October 2025].

Feng, Q. et al. (2014). An epigenomic approach to therapy for tamoxifen-resistant breast cancer. *Cell research*, 24 (7), pp.809–819. [Accessed 23 November 2025].

Ferreira Almeida, C. et al. (2020). Estrogen receptor-positive (ER+) breast cancer treatment: Are multi-target compounds the next promising approach? *Biochemical pharmacology*, 177 (113989), p.113989. [Accessed 24 October 2025].

Filardo, E. J. et al. (2000). Estrogen-induced activation of Erk-1 and Erk-2 requires the G protein-coupled receptor homolog, GPR30, and occurs via trans-activation of the epidermal growth factor receptor through release of HB-EGF. *Molecular endocrinology (Baltimore, Md.)*, 14 (10), pp.1649–1660. [Accessed 20 October 2025].

Filardo, E. J. et al. (2002). Estrogen action via the G protein-coupled receptor, GPR30: stimulation of adenylyl cyclase and cAMP-mediated attenuation of the epidermal growth factor receptor-to-MAPK signaling axis. *Molecular endocrinology (Baltimore, Md.)*, 16 (1), pp.70–84. [Accessed 21 October 2025].

Filice, E. et al. (2009). A new membrane G protein-coupled receptor (GPR30) is involved in the cardiac effects of 17beta-estradiol in the male rat. *Journal of physiology and pharmacology: an official journal of the Polish Physiological Society*, 60 (4), pp.3–10. [Accessed 21 October 2025].

Finlay-Schultz, J. et al. (2017). Breast cancer suppression by progesterone receptors is mediated by their modulation of estrogen receptors and RNA polymerase III. *Cancer research*, 77 (18), pp.4934–4946. [Accessed 3 November 2025].

Fiorito, E., Katika, M. R. and Hurtado, A. (2013). Cooperating transcription factors mediate the function of estrogen receptor. *Chromosoma*, 122 (1–2), pp.1–12. [Accessed 13 March 2026].

Fisher, M. F. and Rao, S. S. (2020). Three-dimensional culture models to study drug resistance in breast cancer. *Biotechnology and bioengineering*, 117 (7), pp.2262–2278. [Accessed 24 October 2025].

Fliss, A. E. et al. (2000). Control of estrogen receptor ligand binding by Hsp90. *The Journal of steroid biochemistry and molecular biology*, 72 (5), pp.223–230. [Accessed 21 October 2025].

Fong, G.-H. and Takeda, K. (2008). Role and regulation of prolyl hydroxylase domain proteins. *Cell death and differentiation*, 15 (4), pp.635–641. [Accessed 23 October 2025].

Foty, R. (2011). A simple hanging drop cell culture protocol for generation of 3D spheroids. *Journal of visualized experiments: JoVE*, (51), p.2720. [Accessed 24 October 2025].

Fountzilas, E. et al. (2020). Pathology of BRCA1- and BRCA2-associated breast cancers: Known and less known connections. *Clinical breast cancer*, 20 (2), pp.152–159. [Accessed 22 October 2025].

Fowble, B. et al. (1996). The impact of tamoxifen on breast recurrence, cosmesis, complications, and survival in estrogen receptor-positive early-stage breast cancer. *International journal of radiation oncology, biology, physics*, 35 (4), pp.669–677. [Accessed 22 October 2025].

Fox, S. et al. (2006). Hypoxia-inducible factor (HIF)-1α predicts a poor response to primary chemo-endocrine therapy and disease-free survival in primary human breast cancer. *Cancer Research*, 66 (8_Supplement), pp.441–442. [Accessed 11 October 2025].

Franchi-Mendes, T. et al. (2021). 3D cancer models: Depicting cellular crosstalk within the tumour microenvironment. *Cancers*, 13 (18), p.4610. [Accessed 3 November 2025].

Franchi-Mendes, T., Lopes, N. and Brito, C. (2021). Heterotypic tumor spheroids in agitation-based cultures: A scaffold-free cell model that sustains long-term survival of endothelial cells. *Frontiers in bioengineering and biotechnology*, 9, p.649949. [Accessed 3 November 2025].

Friedrich, J. et al. (2009). Spheroid-based drug screen: considerations and practical approach. *Nature protocols*, 4 (3), pp.309–324. [Accessed 30 October 2025].

Froehlich, K. et al. (2016). Generation of multicellular breast cancer tumor spheroids: Comparison of different protocols. *Journal of mammary gland biology and neoplasia*, 21 (3–4), pp.89–98. [Accessed 3 November 2025].

Fu, M. et al. (2025). Global, regional, and national time trends in mortality for breast cancer, 1992-2021: an age-period-cohort analysis for the global burden of disease 2021 study. *BMC public health*, 25 (1), p.1599. [Accessed 20 October 2025].

Fu, X. et al. (2016). FOXA1 overexpression mediates endocrine resistance by altering the ER transcriptome and IL-8 expression in ER-positive breast cancer. *Proceedings of the National Academy of Sciences of the United States of America*, 113 (43), pp.E6600–E6609. [Accessed 23 October 2025].

Fu, X. et al. (2019). FOXA1 upregulation promotes enhancer and transcriptional reprogramming in endocrine-resistant breast cancer. *Proceedings of the National Academy of Sciences of the United States of America*, 116 (52), pp.26823–26834. [Accessed 23 October 2025].

Fu, X. et al. (2023). High FOXA1 levels induce ER transcriptional reprogramming, a pro-metastatic secretome, and metastasis in endocrine-resistant breast cancer. *Cell reports*, 42 (8), p.112821. [Accessed 26 August 2025].

Fuller-Pace, F. V. and Ali, S. (2008). The DEAD box RNA helicases p68 (Ddx5) and p72 (Ddx17): novel transcriptional co-regulators. *Biochemical Society transactions*, 36 (Pt 4), pp.609–612. [Accessed 24 November 2025].

Fullwood, M. J. et al. (2009). An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature*, 462 (7269), pp.58–64. [Accessed 8 August 2025].

Furey, T. S. (2012). ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions. *Nature reviews. Genetics*, 13 (12), pp.840–852. [Accessed 26 October 2025].

Galas, D. J. and Schmitz, A. (1978). DNAse footprinting: a simple method for the detection of protein-DNA binding specificity. *Nucleic acids research*, 5 (9), pp.3157–3170.

Galea, M. H. et al. (1992). The Nottingham Prognostic Index in primary breast cancer. *Breast cancer research and treatment*, 22, pp.207–219.

van Galen, P. et al. (2016). A Multiplexed System for Quantitative Comparisons of Chromatin Landscapes. *Molecular cell*, 61 (1), pp.170–180.

Galvan, D. L. et al. (2009). Genome-wide mapping of PiggyBac transposon integrations in primary human T cells. *Journal of immunotherapy (Hagerstown, Md.: 1997)*, 32 (8), pp.837–844. [Accessed 10 November 2025].

Gao, Y. et al. (2024). Mechanisms of endocrine resistance in hormone receptor-positive breast cancer. *Frontiers in oncology*, 14, p.1448687. [Accessed 23 October 2025].

GBD 2021 Forecasting Collaborators. (2024). Burden of disease scenarios for 204 countries and territories, 2022-2050: a forecasting analysis for the Global Burden of Disease Study 2021. *Lancet*, 403 (10440), pp.2204–2256. [Accessed 3 November 2025].

GBD Collaborators. (2025). Global, regional, and national burden of breast cancer among females, 1990–2023, with forecasts to 2050: A systematic analysis for the global burden of disease study 2023. *Social Science Research Network*. [Online]. Available at: doi:10.2139/ssrn.5533924 [Accessed 3 November 2025].

Ge, Y. et al. (2023). Roles of estrogen receptor α in endometrial carcinoma (Review). *Oncology letters*, 26 (6), p.530. [Accessed 31 October 2025].

Generali, D. et al. (2006). Hypoxia-inducible factor-1alpha expression predicts a poor response to primary chemoendocrine therapy and disease-free survival in primary human breast cancer. *Clinical cancer research: an official journal of the American Association for Cancer Research*, 12 (15), pp.4562–4568. [Accessed 24 October 2025].

Gilfillan, G. D. et al. (2012). Limitations and possibilities of low cell number ChIP-seq. *BMC genomics*, 13, p.645. [Accessed 25 October 2025].

Gilmour, D. S. and Lis, J. T. (1985). In vivo interactions of RNA polymerase II with genes of Drosophila melanogaster. *Molecular and cellular biology*, 5 (8), pp.2009–2018. [Accessed 25 October 2025].

Giorgetti, M. et al. (2021). New generation ENaC inhibitors detach cystic fibrosis airway mucus bundles via sodium/hydrogen exchanger inhibition. *European journal of pharmacology*, 904 (174123), p.174123. [Accessed 21 November 2025].

Gisselsson, D. et al. (2019). Clonal evolution through genetic bottlenecks and telomere attrition: Potential threats to in vitro data reproducibility. *Genes, chromosomes & cancer*, 58 (7), pp.452–461. [Accessed 24 November 2025].

Giuliano, M. et al. (2011). Biological mechanisms and clinical implications of endocrine resistance in breast cancer. *Breast (Edinburgh, Scotland)*, 20 Suppl 3, pp.S42-9.

Glibetic, N. et al. (2024). The use of patient-derived organoids in the study of molecular metabolic adaptation in breast cancer. *International journal of molecular sciences*, 25 (19), p.10503. [Accessed 24 October 2025].

Goetz, M. P. et al. (2017). MONARCH 3: Abemaciclib as initial therapy for advanced breast cancer. *Journal of clinical oncology: official journal of the American Society of Clinical Oncology*, 35 (32), pp.3638–3646. [Accessed 22 October 2025].

Gogol-Doring, A. et al. (2016). Genome-wide Profiling Reveals Remarkable Parallels Between Insertion Site Selection Properties of the MLV Retrovirus and the piggyBac Transposon in Primary Human CD4+ T Cells. *Molecular therapy: the journal of the American Society of Gene Therapy*, 24 (3), pp.592–606.

Goldrick, C. et al. (2023). 3D multicellular systems in disease modelling: From organoids to organ-on-chip. *Frontiers in cell and developmental biology*, 11, p.1083175. [Accessed 24 November 2025].

Gong, X. et al. (2015). Generation of multicellular tumor spheroids with microwell-based agarose scaffolds for drug testing. *PloS one*, 10 (6), p.e0130348. [Accessed 24 October 2025].

Gooz, M. and Maldonado, E. N. (2023). Fluorescence microscopy imaging of mitochondrial metabolism in cancer cells. *Frontiers in oncology*, 13, p.1152553. [Accessed 10 October 2025].

Granja, J. M. et al. (2021). ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nature genetics*, 53 (3), pp.403–411. [Accessed 23 November 2025].

Grimes, D. R. et al. (2014). A method for estimating the oxygen consumption rate in multicellular tumour spheroids. *Journal of the Royal Society, Interface*, 11 (92), p.20131124.

Grinda, T. et al. (2021). Phenotypic discordance between primary and metastatic breast cancer in the large-scale real-life multicenter French ESME cohort. *NPJ breast cancer*, 7 (1), p.41. [Accessed 22 October 2025].

Grist, S. M. et al. (2019). Long-term monitoring in a microfluidic system to study tumour spheroid response to chronic and cycling hypoxia. *Scientific reports*, 9 (1), p.17782. [Accessed 24 October 2025].

Grosch, J.-H. et al. (2016). Enzyme activity deviates due to spatial and temporal temperature profiles in commercial microtiter plate readers. *Biotechnology journal*, 11 (4), pp.519–529. [Accessed 21 November 2025].

Grosselin, K. et al. (2019). High-throughput single-cell ChIP-seq identifies heterogeneity of chromatin states in breast cancer. *Nature Genetics 2019 51:6*, 51 (6), pp.1060–1066.

Guertin, M. J. et al. (2018). Parallel factor ChIP provides essential internal control for quantitative differential ChIP-seq. *Nucleic acids research*, 46 (12), p.e75. [Accessed 25 November 2025].

Guillen, K. P. et al. (2022). A human breast cancer-derived xenograft and organoid platform for drug discovery and precision oncology. *Nature cancer*, 3 (2), pp.232–250. [Accessed 11 October 2025].

Guiu, S. et al. (2012). Molecular subclasses of breast cancer: how do we define them? The IMPAKT 2012 Working Group Statement. *Annals of oncology*, 23 (12), pp.2997–3006. [Accessed 20 October 2025].

Gunti, S. et al. (2021). Organoid and spheroid tumor models: Techniques and applications. *Cancers*, 13 (4), p.874. [Accessed 24 October 2025].

Guo, J. et al. (2024). Pycallingcards: an integrated environment for visualizing, analyzing, and interpreting Calling Cards data. *Bioinformatics (Oxford, England)*, 40 (2), p.btae070. [Accessed 6 May 2025].

Guo, L. et al. (2023). Breast cancer heterogeneity and its implication in personalized precision therapy. *Experimental hematology & oncology*, 12 (1), p.3. [Accessed 20 October 2025].

Guo, Y. et al. (2020). Different breast cancer subtypes show different metastatic patterns: A study from A large public database. *Asian Pacific journal of cancer prevention: APJCP*, 21 (12), pp.3587–3593. [Accessed 15 November 2025].

Gust, L. and D'journo, X. B. (2015). The use of correlation functions in thoracic surgery research. *Journal of thoracic disease*, 7 (3), pp.E11-5. [Accessed 15 October 2025].

Ha, S.-A. et al. (2011). The prognostic potential of keratin 18 in breast cancer associated with tumor dedifferentiation, and the loss of estrogen and progesterone receptors. *Cancer biomarkers: section A of Disease markers*, 10 (5), pp.219–231. [Accessed 24 November 2025].

Haase, V. H. (2009). The VHL tumor suppressor: master regulator of HIF. *Current pharmaceutical design*, 15 (33), pp.3895–3903. [Accessed 23 October 2025].

Habara, M. and Shimada, M. (2022). Estrogen receptor α revised: Expression, structure, function, and stability. *BioEssays: news and reviews in molecular, cellular and developmental biology*, 44 (12), p.e2200148. [Accessed 20 October 2025].

Hah, N. et al. (2011). A Rapid, Extensive, and Transient Transcriptional Response to Estrogen Signaling in Breast Cancer Cells. *Cell*, 145 (4), pp.622–634.

Hah, N. et al. (2013). Enhancer transcripts mark active estrogen receptor binding sites. *Genome research*, 23 (8), pp.1210–1223. [Accessed 24 November 2025].

Hainer, S. J. et al. (2019). Profiling of Pluripotency Factors in Single Cells and Early Embryos. *Cell*, 177 (5), pp.1319-1329.e11.

Haines, C. N. et al. (2020). GREB1 regulates PI3K/Akt signaling to control hormone-sensitive breast cancer proliferation. *Carcinogenesis*, 41 (12), pp.1660–1670. [Accessed 26 August 2025].

Ham, S. L. et al. (2016). Engineered breast cancer cell spheroids reproduce biologic properties of solid tumors. *Advanced healthcare materials*, 5 (21), pp.2788–2798. [Accessed 24 October 2025].

Hammond, M. E. H. et al. (2010). American Society of Clinical Oncology/College Of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer. *Journal of clinical oncology: official journal of the American Society of Clinical Oncology*, 28 (16), pp.2784–2795. [Accessed 25 November 2025].

Han, S. J., Kwon, S. and Kim, K. S. (2021). Challenges of applying multicellular tumor spheroids in preclinical phase. *Cancer cell international*, 21 (1), p.152. [Accessed 3 November 2025].

Hanahan, D. and Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. *Cell*, 144 (5), pp.646–674. [Accessed 23 October 2025].

Hanker, A. B., Sudhan, D. R. and Arteaga, C. L. (2020). Overcoming endocrine resistance in breast cancer. *Cancer cell*, 37 (4), pp.496–513. [Accessed 22 October 2025].

Hansen, P. et al. (2015). Saturation analysis of ChIP-seq data for reproducible identification of binding peaks. *Genome research*, 25 (9), pp.1391–1400. [Accessed 23 November 2025].

Haque, A. et al. (2017). A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. *Genome medicine*, 9 (1), p.75. [Accessed 26 October 2025].

Harada, A. et al. (2018). A chromatin integration labelling method enables epigenomic profiling with lower input. *Nature Cell Biology 2018 21:2*, 21 (2), pp.287–296.

Hartmaier, R. J. et al. (2018). Recurrent hyperactive ESR1 fusion proteins in endocrine therapy-resistant breast cancer. *Annals of oncology*, 29 (4), pp.872–880. [Accessed 25 November 2025].

Hashimoto, Y. et al. (2023). Detection of ultra-rare ESR1 mutations in primary breast cancer using LNA-clamp ddPCR. *Cancers*, 15 (9), p.2632. [Accessed 15 November 2025].

Hass, D. T. et al. (2023). Medium depth influences O2 availability and metabolism in human RPE cultures. *Investigative ophthalmology & visual science*, 64 (14), p.4. [Accessed 21 November 2025].

Hayashi, S. and McMahon, A. P. (2002). Efficient recombination in diverse tissues by a tamoxifen-inducible form of Cre: a tool for temporally regulated gene activation/inactivation in the mouse. *Developmental biology*, 244 (2), pp.305–318. [Accessed 22 August 2025].

Haynes, M. P. et al. (2000). Membrane estrogen receptor engagement activates endothelial nitric oxide synthase via the PI3-kinase-Akt pathway in human endothelial cells. *Circulation research*, 87 (8), pp.677–682. [Accessed 21 October 2025].

He, C., Zhang, M. Q. and Wang, X. (2015). MICC: an R package for identifying chromatin interactions from ChIA-PET data. *Bioinformatics (Oxford, England)*, 31 (23), pp.3832–3834. [Accessed 16 September 2025].

He, L. et al. (2019). The prognosis comparison of different molecular subtypes of breast tumors after radiotherapy and the intrinsic reasons for their distinct radiosensitivity. *Cancer management and research*, 11, pp.5765–5775. [Accessed 22 October 2025].

He, Z. and Zhang, S. (2021). Tumor-associated macrophages and their functional transformation in the hypoxic tumor microenvironment. *Frontiers in immunology*, 12, p.741305. [Accessed 3 November 2025].

Heddleston, J. M. et al. (2010). Hypoxia inducible factors in cancer stem cells. *British journal of cancer*, 102 (5), pp.789–795. [Accessed 24 October 2025].

Heinz, S. et al. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular cell*, 38 (4), pp.576–589. [Accessed 7 November 2025].

Heldring, N. et al. (2007). Estrogen receptors: how do they signal and what are their targets. *Physiological reviews*, 87 (3), pp.905–931. [Accessed 21 October 2025].

Helleday, T. (2011). The underlying mechanism for the PARP and BRCA synthetic lethality: clearing up the misunderstandings. *Molecular oncology*, 5 (4), pp.387–393. [Accessed 22 October 2025].

Hernández, G., Osnaya, V. G. and Pérez-Martínez, X. (2019). Conservation and variability of the AUG initiation Codon context in eukaryotes. *Trends in biochemical sciences*, 44 (12), pp.1009–1021. [Accessed 26 August 2025].

Hewitt, S. C. and Korach, K. S. (2018). Estrogen receptors: New directions in the new millennium. *Endocrine reviews*, 39 (5), pp.664–675. [Accessed 20 October 2025].

Hirschhaeuser, F. et al. (2010). Multicellular tumor spheroids: an underestimated tool is catching up again. *Journal of biotechnology*, 148 (1), pp.3–15.

Hoch, R. V. et al. (1999). GATA-3 is expressed in association with estrogen receptor in breast cancer. *International journal of cancer. Journal international du cancer*, 84 (2), pp.122–128. [Accessed 26 August 2025].

Hodgkinson, K. et al. (2018). GREB1 is an estrogen receptor-regulated tumour promoter that is frequently expressed in ovarian cancer. *Oncogene*, 37 (44), pp.5873–5886. [Accessed 26 August 2025].

Hoefflin, R. et al. (2020). HIF-1α and HIF-2α differently regulate tumour development and inflammation of clear cell renal cell carcinoma in mice. *Nature communications*, 11 (1), p.4111. [Accessed 23 October 2025].

Holding, A. N. et al. (2019). VULCAN integrates ChIP-seq with patient-derived co-expression networks to identify GRHL2 as a key co-regulator of ERa at enhancers in breast cancer. *Genome biology*, 20 (1), p.91. [Accessed 22 August 2025].

Hongisto, V. et al. (2013). High-throughput 3D screening reveals differences in drug sensitivities between culture models of JIMT1 breast cancer cells. *PloS one*, 8 (10), p.e77232. [Accessed 3 November 2025].

Horigome, T. et al. (1988). Estradiol-Stimulated Proteolytic Cleavage of the Estrogen Receptor in Mouse Uterus. *Endocrinology*, 123 (5), pp.2540–2548.

Hou, J. et al. (2020). The role of amplified in breast cancer 1 in breast cancer: A meta-analysis: A meta-analysis. *Medicine*, 99 (46), p.e23248. [Accessed 23 October 2025].

Hsu, P.-Y. et al. (2013). Amplification of distant estrogen response elements deregulates target genes associated with tamoxifen resistance in breast cancer. *Cancer cell*, 24 (2), pp.197–212. [Accessed 10 November 2025].

Hu, D. et al. (2022). Cancer-associated fibroblasts in breast cancer: Challenges and opportunities. *Cancer communications (London, England)*, 42 (5), pp.401–434. [Accessed 3 November 2025].

Hu, K., Zhong, G. and He, F. (2005). Expression of estrogen receptors ERalpha and ERbeta in endometrial hyperplasia and adenocarcinoma. *International journal of gynecological cancer: official journal of the International Gynecological Cancer Society*, 15 (3), pp.537–541. [Accessed 31 October 2025].

Hu, M. et al. (2010). On the detection and refinement of transcription factor binding sites using ChIP-Seq data. *Nucleic acids research*, 38 (7), pp.2154–2167. [Accessed 25 October 2025].

Hu, Y. et al. (2023). Single-cell multi-scale footprinting reveals the modular organization of DNA regulatory elements. *bioRxivorg*, p.2023.03.28.533945. [Online]. Available at: doi:10.1101/2023.03.28.533945 [Accessed 23 November 2025].

Hua, H. et al. (2018). Mechanisms for estrogen receptor expression in human cancer. *Experimental hematology & oncology*, 7 (1), p.24. [Accessed 21 October 2025].

Huang, Z., Yu, P. and Tang, J. (2020). Characterization of triple-negative breast cancer MDA-MB-231 cell spheroid model. *OncoTargets and therapy*, 13, pp.5395–5405. [Accessed 24 November 2025].

Hughes, P. et al. (2007). The costs of using unauthenticated, over-passaged cell lines: how much more data do we need? *BioTechniques*, 43 (5), pp.575, 577–578, 581-2 passim. [Accessed 24 November 2025].

Hurtado, A. et al. (2011). FOXA1 is a key determinant of estrogen receptor function and endocrine response. *Nature genetics*, 43 (1), pp.27–33.

Hynds, R. E., Vladimirou, E. and Janes, S. M. (2018). The secret lives of cancer cell lines. *Disease models & mechanisms*, 11 (11), p.dmm037366. [Accessed 24 November 2025].

Idrisova, K. F., Simon, H.-U. and Gomzikova, M. O. (2022). Role of patient-derived models of cancer in translational oncology. *Cancers*, 15 (1), p.139. [Accessed 24 November 2025].

Imamura, Y. et al. (2015). Comparison of 2D- and 3D-culture models as drug-testing platforms in breast cancer. *Oncology reports*, 33 (4), pp.1837–1843.

Inaki, K. et al. (2011). Transcriptional consequences of genomic structural aberrations in breast cancer. *Genome research*, 21 (5), pp.676–687. [Accessed 24 November 2025].

Inwald, E. C. et al. (2013). Ki-67 is a prognostic parameter in breast cancer patients: results of a large population-based cohort of a cancer registry. *Breast cancer research and treatment*, 139 (2), pp.539–552. [Accessed 25 November 2025].

Iwafuchi-Doi, M. et al. (2016). The pioneer transcription factor FoxA maintains an accessible nucleosome configuration at enhancers for tissue-specific gene activation. *Molecular cell*, 62 (1), pp.79–91. [Accessed 9 November 2025].

Iwafuchi-Doi, M. and Zaret, K. S. (2014). Pioneer transcription factors in cell reprogramming. *Genes & development*, 28 (24), pp.2679–2692. [Accessed 24 November 2025].

Iyer, S. V. et al. (2013). Understanding the role of keratins 8 and 18 in neoplastic potential of breast cancer derived cell lines. *PloS one*, 8 (1), p.e53532. [Accessed 24 November 2025].

Jahin, I. et al. (2023). Extracellular matrix stiffness activates mechanosensitive signals but limits breast cancer cell spheroid proliferation and invasion. *Frontiers in Cell and Developmental Biology*, 11, p.1292775. [Accessed 13 March 2026].

Jaini, S. et al. (2014). Transcription factor binding site mapping using ChIP-Seq. *Microbiology spectrum*, 2 (2). [Online]. Available at: doi:10.1128/microbiolspec.MGM2-0035-2013 [Accessed 25 October 2025].

Jakesz, R. et al. (1984). Influence of cell proliferation and cell cycle phase on expression of estrogen receptor in MCF-7 breast cancer cells. *Cancer research*, 44 (2), pp.619–625. [Accessed 24 November 2025].

Jalili, V. et al. (2015). Using combined evidence from replicates to evaluate ChIP-seq peaks. *Bioinformatics (Oxford, England)*, 31 (17), pp.2761–2769. [Accessed 24 November 2025].

Jardin, I. et al. (2023). Orai1α and Orai1β support calcium entry and mammosphere formation in breast cancer stem cells. *Scientific reports*, 13 (1), p.19471. [Accessed 8 October 2025].

Jaśkiewicz, M. et al. (2022a). The transition from HIF-1 to HIF-2 during prolonged hypoxia results from reactivation of PHDs and HIF1A mRNA instability. *Cellular & molecular biology letters*, 27 (1), p.109.

Jaśkiewicz, M. et al. (2022b). The transition from HIF-1 to HIF-2 during prolonged hypoxia results from reactivation of PHDs and HIF1A mRNA instability. *Research Square*. [Online]. Available at: doi:10.21203/rs.3.rs-1607563/v1.

Jeffreys, S. A. et al. (2022). Prognostic and predictive value of CCND1/cyclin D1 amplification in breast cancer with a focus on postmenopausal patients: A systematic review and meta-analysis. *Frontiers in endocrinology*, 13, p.895729. [Accessed 21 October 2025].

Jehanno, C. et al. (2022). Hypoxia and ERα transcriptional crosstalk is associated with endocrine resistance in breast cancer. *Cancers*, 14 (19), p.4934. [Accessed 14 July 2025].

Jeon, H. et al. (2020). Comparative analysis of commonly used peak calling programs for ChIP-Seq analysis. *Genomics & informatics*, 18 (4), p.e42. [Accessed 15 October 2025].

Jeong, Y., Tin, A. and Irudayaraj, J. (2022). Flipped well-plate hanging-drop technique for growing three-dimensional tumors. *Frontiers in bioengineering and biotechnology*, 10, p.898699. [Accessed 24 October 2025].

Jeselsohn, R. et al. (2015). ESR1 mutations—a mechanism for acquired endocrine resistance in breast cancer. *Nature reviews. Clinical oncology*, 12 (10), pp.573–583. [Accessed 22 October 2025].

Ji, X. et al. (2019). Chemoresistance mechanisms of breast cancer and their countermeasures. *Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie*, 114, p.108800.

Ji, Z. et al. (2020). Single-cell ATAC-seq signal extraction and enhancement with SCATE. *Genome biology*, 21 (1), p.161. [Accessed 23 November 2025].

Jia, W. et al. (2016). Direct 3D bioprinting of perfusable vascular constructs using a blend bioink. *Biomaterials*, 106, pp.58–68. [Accessed 24 October 2025].

Jia, X. et al. (2015). Basal and therapy-driven hypoxia-inducible factor-1α confers resistance to endocrine therapy in estrogen receptor-positive breast cancer. *Oncotarget*, 6 (11), pp.8648–8662. [Accessed 11 October 2025].

Jiang, Y.-Z. et al. (2014). GATA3mutations define a unique subtype of luminal-like breast cancer with improved survival: GATA3Mutations in Breast Cancer. *Cancer*, 120 (9), pp.1329–1337. [Accessed 21 October 2025].

Joffroy, C. M. et al. (2010). Antiestrogens induce transforming growth factor beta-mediated immunosuppression in breast cancer. *Cancer research*, 70 (4), pp.1314–1322. [Accessed 23 October 2025].

Johansson, A. et al. (2025). Differential long-term tamoxifen therapy benefit by menopausal status in breast cancer patients: secondary analysis of a controlled randomized clinical trial. *Journal of the National Cancer Institute*, 117 (5), pp.868–878. [Accessed 22 October 2025].

Johansson, H. J. et al. (2013). Retinoic acid receptor alpha is associated with tamoxifen resistance in breast cancer. *Nature communications*, 4 (1), p.2175. [Accessed 7 November 2025].

Johnson, D. S. et al. (2007). Genome-wide mapping of in vivo protein-DNA interactions. *Science (New York, N.Y.)*, 316 (5830), pp.1497–1502. [Accessed 25 October 2025].

Johnson, K. A. (2006). Editorial: The SERM of my dreams. *The journal of clinical endocrinology and metabolism*, 91 (10), pp.3754–3756. [Accessed 22 October 2025].

Johnson, K. S., Conant, E. F. and Soo, M. S. (2021). Molecular subtypes of breast cancer: A review for breast radiologists. *Journal of breast imaging*, 3 (1), pp.12–24. [Accessed 22 October 2025].

Johnston, S. R. D. et al. (2020). Abemaciclib combined with endocrine therapy for the adjuvant treatment of HR+, HER2-, node-positive, high-risk, early breast cancer (monarchE). *Journal of clinical oncology: official journal of the American Society of Clinical Oncology*, 38 (34), pp.3987–3998. [Accessed 22 October 2025].

Jothi, R. et al. (2008). Genome-wide identification of in vivo protein-DNA binding sites from ChIP-Seq data. *Nucleic acids research*, 36 (16), pp.5221–5231. [Accessed 25 October 2025].

Jovic, D. et al. (2022). Single-cell RNA sequencing technologies and applications: A brief overview. *Clinical and translational medicine*, 12 (3), p.e694. [Accessed 24 October 2025].

Jubelin, C. et al. (2023). Technical report: liquid overlay technique allows the generation of homogeneous osteosarcoma, glioblastoma, lung and prostate adenocarcinoma spheroids that can be used for drug cytotoxicity measurements. *Frontiers in bioengineering and biotechnology*, 11, p.1260049. [Accessed 24 October 2025].

Jung, Y. L. et al. (2014). Impact of sequencing depth in ChIP-seq experiments. *Nucleic acids research*, 42 (9), p.e74. [Accessed 4 November 2025].

Kahiapo, J. K. and Monahan, K. (2023). Chromatin immunoprecipitation from formaldehyde cross-linked olfactory sensory neurons. *Methods in molecular biology (Clifton, N.J.)*, 2710, pp.71–82. [Accessed 25 October 2025].

Kalaitzidis, D. et al. (2004). Characterization of a human REL-estrogen receptor fusion protein with a reverse conditional transforming activity in chicken spleen cells. *Oncogene*, 23 (45), pp.7580–7587. [Accessed 22 August 2025].

Kalisky, T. et al. (2018). A brief review of single-cell transcriptomic technologies. *Briefings in functional genomics*, 17 (1), pp.64–76. [Accessed 25 October 2025].

Kaluz, S., Kaluzová, M. and Stanbridge, E. J. (2008). Regulation of gene expression by hypoxia: integration of the HIF-transduced hypoxic signal at the hypoxia-responsive element. *Clinica chimica acta; international journal of clinical chemistry*, 395 (1–2), pp.6–13. [Accessed 23 October 2025].

Kammerer, S. et al. (2016). KCNJ3 is a new independent prognostic marker for estrogen receptor positive breast cancer patients. *Oncotarget*, 7 (51), pp.84705–84717. [Accessed 14 September 2025].

Kamoya, T. et al. (2016). An oxygen-permeable spheroid culture chip (Oxy chip) promotes osteoblastic differentiation of mesenchymal stem cells. *Sensors and actuators. B, Chemical*, 232, pp.75–83. [Accessed 21 November 2025].

Kang, Y. P. et al. (2019). Spheroid-induced epithelial-mesenchymal transition provokes global alterations of breast cancer lipidome: A multi-layered omics analysis. *Frontiers in oncology*, 9, p.145. [Accessed 8 October 2025].

Kao, T.-W. et al. (2023). Novel cancer treatment paradigm targeting hypoxia-induced factor in conjunction with current therapies to overcome resistance. *Journal of experimental & clinical cancer research: CR*, 42 (1), p.171. [Accessed 23 October 2025].

Kapałczyńska, M. et al. (2018). 2D and 3D cell cultures - a comparison of different types of cancer cell cultures. *Archives of Medical Science*, 14 (4), pp.910–919. [Accessed 13 March 2026].

Kapara, A. et al. (2021). Detection of estrogen receptor alpha and assessment of fulvestrant activity in MCF-7 tumor spheroids using microfluidics and SERS. *Analytical chemistry*, 93 (14), pp.5862–5871. [Accessed 10 October 2025].

Kato, S. et al. (1995). Activation of the estrogen receptor through phosphorylation by mitogen-activated protein kinase. *Science (New York, N.Y.)*, 270 (5241), pp.1491–1494. [Accessed 26 November 2025].

Kaya-Okur, H. S. et al. (2019). CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nature Communications 2019 10:1*, 10 (1), pp.1–10.

Kaye, S. B. et al. (1981). The treatment of metastatic breast cancer with aminoglutethimide. *Australian and New Zealand journal of medicine*, 11 (3), pp.370–374.

Kelm, J. M. et al. (2003). Method for generation of homogeneous multicellular tumor spheroids applicable to a wide variety of cell types. *Biotechnology and bioengineering*, 83 (2), pp.173–180.

Kenny, P. A. et al. (2007). The morphologies of breast cancer cell lines in three-dimensional assays correlate with their profiles of gene expression. *Molecular oncology*, 1 (1), pp.84–96. [Accessed 11 October 2025].

Kent, W. J. et al. (2002). The human genome browser at UCSC. *Genome research*, 12 (6), pp.996–1006. [Accessed 24 November 2025].

Kharchenko, P. V., Tolstorukov, M. Y. and Park, P. J. (2008). Design and analysis of ChIP-seq experiments for DNA-binding proteins. *Nature biotechnology*, 26 (12), p.1351. [Accessed 3 November 2025].

Kim, H. et al. (2018). The hypoxic tumor microenvironment in vivo selects the cancer stem cell fate of breast cancer cells. *Breast cancer research: BCR*, 20 (1), p.16. [Accessed 25 November 2025].

Kim, I. et al. (2022). Cancer-associated fibroblasts in the hypoxic tumor microenvironment. *Cancers*, 14 (14), p.3321. [Accessed 3 November 2025].

Kim, J. et al. (2025). Global patterns and trends in breast cancer incidence and mortality across 185 countries. *Nature medicine*, 31 (4), pp.1154–1162.

Kinnel, B. et al. (2023). Targeted Therapy and Mechanisms of Drug Resistance in Breast Cancer. *Cancers*, 15 (4). [Online]. Available at: doi:10.3390/cancers15041320.

Kinslow, C. J. et al. (2022). Prevalence of estrogen receptor alpha (ESR1) somatic mutations in breast cancer. *JNCI cancer spectrum*, 6 (5), p.pkac060. [Accessed 20 October 2025].

Klein, J. S. et al. (2014). Design and characterization of structured protein linkers with differing flexibilities. *Protein Engineering Design and Selection*, 27 (10), pp.325–330. [Accessed 24 November 2025].

Klopper, A. and Hall, M. (1971). New synthetic agent for the induction of ovulation: preliminary trials in women. *British medical journal*, 1 (5741), pp.152–154. [Accessed 22 October 2025].

Kobayashi, Y. et al. (2000). p300 mediates functional synergism between AF-1 and AF-2 of estrogen receptor alpha and beta by interacting directly with the N-terminal A/B domains. *The journal of biological chemistry*, 275 (21), pp.15645–15651. [Accessed 21 October 2025].

Koedoot, E. et al. (2021). Differential reprogramming of breast cancer subtypes in 3D cultures and implications for sensitivity to targeted therapy. *Scientific reports*, 11 (1), p.7259. [Accessed 11 October 2025].

Koh, M. Y. and Powis, G. (2012). Passing the baton: the HIF switch. *Trends in biochemical sciences*, 37 (9), pp.364–372. [Accessed 23 October 2025].

Kohn, A. D. et al. (1998). Construction and characterization of a conditionally active version of the serine/threonine kinase Akt. *The journal of biological chemistry*, 273 (19), pp.11937–11943. [Accessed 26 August 2025].

Koide, A. et al. (2007). Identification of regions within the F domain of the human estrogen receptor alpha that are important for modulating transactivation and protein-protein interactions. *Molecular endocrinology (Baltimore, Md.)*, 21 (4), pp.829–842. [Accessed 26 August 2025].

Koletsis, N. E. et al. (2025). Development, functional characterization, and matrix effectors dynamics in 3D spheroids of triple-negative breast cancer cells. *Cells (Basel, Switzerland)*, 14 (17), p.1351. [Accessed 13 March 2026].

Kolligs, F. T. et al. (2002). ITF-2, a downstream target of the Wnt/TCF pathway, is activated in human cancers with beta-catenin defects and promotes neoplastic transformation. *Cancer cell*, 1 (2), pp.145–155. [Accessed 22 August 2025].

Kong, S. L. et al. (2011). Cellular reprogramming by the conjoint action of ERα, FOXA1, and GATA3 to a ligand-inducible growth state. *Molecular systems biology*, 7 (1), p.526. [Accessed 21 October 2025].

Kozak, M. (2002). Pushing the limits of the scanning mechanism for initiation of translation. *Gene*, 299 (1–2), pp.1–34. [Accessed 24 November 2025].

Kozal, K. and Krześlak, A. (2022). The role of hypoxia-inducible factor isoforms in breast cancer and perspectives on their inhibition in therapy. *Cancers*, 14 (18), p.4518. [Accessed 23 October 2025].

Kuiper, G. G. et al. (1996). Cloning of a novel receptor expressed in rat prostate and ovary. *Proceedings of the National Academy of Sciences of the United States of America*, 93 (12), pp.5925–5930. [Accessed 20 October 2025].

Kumar, R. et al. (2011). The dynamic structure of the estrogen receptor. *Journal of amino acids*, 2011, p.812540. [Accessed 9 November 2025].

Kuo, C.-T. et al. (2017). Three-dimensional spheroid culture targeting versatile tissue bioassays using a PDMS-based hanging drop array. *Scientific reports*, 7 (1), p.4363. [Accessed 24 October 2025].

Kurata, H. et al. (1999). Ectopic expression of activated Stat6 induces the expression of Th2-specific cytokines and transcription factors in developing Th1 cells. *Immunity*, 11 (6), pp.677–688. [Accessed 22 August 2025].

Kurebayashi, J. et al. (2001). Hypoxia reduces hormone responsiveness of human breast cancer cells. *Japanese journal of cancer research: Gann*, 92 (10), pp.1093–1101. [Accessed 24 November 2025].

Kwon, J., Lee, D. and Lee, S.-A. (2023). BAP1 as a guardian of genome stability: implications in human cancer. *Experimental & molecular medicine*, 55 (4), pp.745–754. [Accessed 9 November 2025].

Laganière, J. et al. (2005). From the Cover: Location analysis of estrogen receptor alpha target promoters reveals that FOXA1 defines a domain of the estrogen response. *Proceedings of the National Academy of Sciences of the United States of America*, 102 (33), pp.11651–11656.

Lakhani, S. R. et al. (1998). Multifactorial analysis of differences between sporadic breast cancers and cancers involving BRCA1 and BRCA2 mutations. *Journal of the National Cancer Institute*, 90 (15), pp.1138–1145. [Accessed 22 October 2025].

Lalli, M. et al. (2022). Measuring transcription factor binding and gene expression using barcoded self-reporting transposon calling cards and transcriptomes. *NAR genomics and bioinformatics*, 4 (3), p.lqac061. [Accessed 6 May 2025].

Lamichhane, A. and Tavana, H. (2024). Three-dimensional tumor models to study cancer stemness-mediated drug resistance. *Cellular and molecular bioengineering*, 17 (2), pp.107–119. [Accessed 24 October 2025].

Landt, S. G. et al. (2012). ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome research*, 22 (9), pp.1813–1831. [Accessed 15 October 2025].

Lao, Z. et al. (2015). Improved methods to generate spheroid cultures from tumor cells, tumor cells & fibroblasts or tumor-fragments: Microenvironment, microvesicles and MiRNA. *PloS one*, 10 (7), p.e0133895. [Accessed 8 October 2025].

Lardo, S. M. and Hainer, S. J. (2022). Single-cell factor localization on chromatin using ultra-low input cleavage Under Targets and release using nuclease. *Journal of visualized experiments: JoVE*, (180), p.e63536. [Accessed 26 October 2025].

Larson, H. G. et al. (2021). A genome-edited ERα-HiBiT fusion reporter cell line for the identification of ERα modulators via high-throughput screening and CETSA. *Assay and drug development technologies*, 19 (8), pp.539–549. [Accessed 22 August 2025].

Larsson, P. et al. (2020). Optimization of cell viability assays to improve replicability and reproducibility of cancer drug sensitivity screens. *Scientific reports*, 10 (1), p.5798. [Accessed 15 September 2025].

Latysheva, N. S. and Babu, M. M. (2019). Molecular signatures of fusion proteins in cancer. *ACS pharmacology & translational science*, 2 (2), pp.122–133. [Accessed 23 October 2025].

Lawrence, M. et al. (2013). Software for computing and annotating genomic ranges. *PLoS computational biology*, 9 (8), p.e1003118. [Accessed 24 November 2025].

Lawrence, M., Gentleman, R. and Carey, V. (2009). rtracklayer: an R package for interfacing with genome browsers. *Bioinformatics (Oxford, England)*, 25 (14), pp.1841–1842. [Accessed 24 November 2025].

Lebedeva, G. et al. (2012). A model of estrogen-related gene expression reveals non-linear effects in transcriptional response to tamoxifen. *BMC systems biology*, 6 (1), p.138. [Accessed 26 August 2025].

Lee, H.-R., Kim, T.-H. and Choi, K.-C. (2012). Functions and physiological roles of two types of estrogen receptors, ERα and ERβ, identified by estrogen receptor knockout mouse. *Laboratory animal research*, 28 (2), pp.71–76. [Accessed 21 October 2025].

Lee, J. J.-K. et al. (2023a). ERα-associated translocations underlie oncogene amplifications in breast cancer. *Nature*, 618 (7967), pp.1024–1032. [Accessed 10 November 2025].

Lee, J. Y. et al. (2017a). A transcriptional complex composed of ER(α), GATA3, FOXA1 and ELL3 regulates IL-20 expression in breast cancer cells. *Oncotarget*, 8 (26), pp.42752–42760. [Accessed 26 August 2025].

Lee, J.-E. et al. (2017b). Brd4 binds to active enhancers to control cell identity gene induction in adipogenesis and myogenesis. *Nature communications*, 8 (1), p.2217. [Accessed 31 October 2025].

Lee, M. J. et al. (2025). Comparison of oncological outcomes of premenopausal with ovarian function suppression versus postmenopausal women in ER+/HER2- breast cancer. *Breast (Edinburgh, Scotland)*, 81 (104449), p.104449. [Accessed 22 October 2025].

Lee, S. Y. et al. (2010). Implication of necrosis-linked p53 aggregation in acquired apoptotic resistance to 5-FU in MCF-7 multicellular tumour spheroids. *Oncology reports*, 24 (1), pp.73–79. [Accessed 26 November 2025].

Lee, S.-Y. et al. (2023b). In Vitro three-dimensional (3D) cell culture tools for spheroid and organoid models. *SLAS discovery*, 28 (4), pp.119–137. [Accessed 24 October 2025].

Lee, Y. J. et al. (2023c). Prognosis according to the timing of recurrence in breast cancer. *Annals of surgical treatment and research*, 104 (1), pp.1–9. [Accessed 22 October 2025].

Lee, Y.-J. et al. (2023d). Survival outcomes of breast cancer patients with recurrence after surgery according to period and subtype. *PloS one*, 18 (7), p.e0284460. [Accessed 22 October 2025].

Lehmann, B. D. et al. (2011). Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *The journal of clinical investigation*, 121 (7), pp.2750–2767. [Accessed 15 November 2025].

Lei, J. T. et al. (2018). Functional Annotation of ESR1 Gene Fusions in Estrogen Receptor-Positive Breast Cancer. *Cell reports*, 24 (6), pp.1434-1444.e7.

Leon, L. J. et al. (2013). A cell-permeant amiloride derivative induces caspase-independent, AIF-mediated programmed necrotic death of breast cancer cells. *PloS one*, 8 (4), p.e63038. [Accessed 10 July 2025].

Levin, E. R. (2002). Cellular functions of plasma membrane estrogen receptors. *Steroids*, 67 (6), pp.471–475. [Accessed 21 October 2025].

Li, D. et al. (2020a). Tumor-associated macrophages secrete CC-chemokine ligand 2 and induce tamoxifen resistance by activating PI3K/Akt/mTOR in breast cancer. *Cancer science*, 111 (1), pp.47–58. [Accessed 23 October 2025].

Li, G. et al. (2014). Chromatin Interaction Analysis with Paired-End Tag (ChIA-PET) sequencing technology and application. *BMC genomics*, 15 Suppl 12 (S12), p.S11. [Accessed 24 November 2025].

Li, G. et al. (2017a). ChIA-PET2: a versatile and flexible pipeline for ChIA-PET data analysis. *Nucleic acids research*, 45 (1), p.e4. [Accessed 18 November 2025].

Li, H. et al. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25 (16), pp.2078–2079. [Accessed 24 November 2025].

Li, J. J., Tsang, J. Y. and Tse, G. M. (2021). Tumor microenvironment in breast cancer-updates on therapeutic implications and pathologic assessment. *Cancers*, 13 (16), p.4233. [Accessed 23 November 2025].

Li, M. A. et al. (2013a). The piggyBac transposon displays local and distant reintegration preferences and can cause mutations at noncanonical integration sites. *Molecular and cellular biology*, 33 (7), pp.1317–1330. [Accessed 10 November 2025].

Li, Q. et al. (2011). Measuring reproducibility of high-throughput experiments. *The Annals of Applied Statistics*, 5 (3), pp.1752–1779. [Accessed 15 October 2025].

Li, S. et al. (2013b). Endocrine-therapy-resistant ESR1 variants revealed by genomic characterization of breast-cancer-derived xenografts. *Cell reports*, 4 (6), pp.1116–1130. [Accessed 25 November 2025].

Li, S. et al. (2024). A meta-analysis of randomized controlled trials comparing breast-conserving surgery and mastectomy in terms of patient survival rate and quality of life in breast cancer. *International journal for quality in health care*, 36 (2), p.mzae043. [Accessed 22 October 2025].

Li, X. et al. (2017b). Long-read ChIA-PET for base-pair-resolution mapping of haplotype-specific chromatin interactions. *Nature protocols*, 12 (5), pp.899–915.

Li, Y. et al. (2020b). Dynamic surveillance of tamoxifen-resistance in ER-positive breast cancer by CAIX-targeted ultrasound imaging. *Cancer medicine*, 9 (7), pp.2414–2426. [Accessed 11 October 2025].

Li, Z. et al. (2013c). Simple piggyBac transposon-based mammalian cell expression system for inducible protein production. *Proceedings of the National Academy of Sciences of the United States of America*, 110 (13), pp.5004–5009. [Accessed 10 November 2025].

Li, Z. et al. (2022a). ESR1 mutant breast cancers show elevated basal cytokeratins and immune activation. *Nature communications*, 13 (1), p.2011. [Accessed 21 October 2025].

Li, Z. et al. (2022b). Hotspot ESR1 mutations are multimodal and contextual modulators of breast cancer metastasis. *Cancer Research*, 82 (7), pp.1321–1339. [Accessed 13 March 2026].

Li, Z. et al. (2023). The EstroGene database reveals diverse temporal, context-dependent, and bidirectional estrogen receptor regulomes in breast cancer. *Cancer research*, 83 (16), pp.2656–2674. [Accessed 23 November 2025].

Liao, C. et al. (2023). Tumor hypoxia: From basic knowledge to therapeutic implications. *Seminars in cancer biology*, 88, pp.172–186. [Accessed 23 October 2025].

Liao, X.-H. et al. (2014). Estrogen receptor α mediates proliferation of breast cancer MCF-7 cells via a p21/PCNA/E2F1-dependent pathway. *The FEBS journal*, 281 (3), pp.927–942. [Accessed 26 August 2025].

Lin, C.-Y. et al. (2007). Whole-genome cartography of estrogen receptor alpha binding sites. *PLoS Genetics*, 3 (6), p.e87. [Accessed 13 March 2026].

Lin, N. U. et al. (2023). Tucatinib vs placebo, both in combination with trastuzumab and capecitabine, for previously treated ERBB2 (HER2)-positive metastatic breast cancer in patients with brain metastases: Updated exploratory analysis of the HER2CLIMB randomized clinical trial: Updated exploratory analysis of the HER2CLIMB randomized clinical trial. *JAMA oncology*, 9 (2), pp.197–205. [Accessed 20 October 2025].

Lindberg, M. K. et al. (2003). Estrogen receptor (ER)-beta reduces ERalpha-regulated gene transcription, supporting a 'ying yang' relationship between ERalpha and ERbeta in mice. *Molecular endocrinology (Baltimore, Md.)*, 17 (2), pp.203–208. [Accessed 21 October 2025].

Lindström, L. S. et al. (2018). Intratumor Heterogeneity of the Estrogen Receptor and the Long-term Risk of Fatal Breast Cancer. *JNCI Journal of the National Cancer Institute*, 110 (7), p.726.

Lippman, S. M. and Brown, P. H. (1999). Tamoxifen prevention of breast cancer: an instance of the fingerpost. *Journal of the National Cancer Institute*, 91 (21), pp.1809–1819. [Accessed 22 October 2025].

Liu, B. et al. (2022a). BRD4-directed super-enhancer organization of transcription repression programs links to chemotherapeutic efficacy in breast cancer. *Proceedings of the National Academy of Sciences of the United States of America*, 119 (6), p.e2109133119. [Accessed 22 August 2025].

Liu, M. H. and Cheung, E. (2014). Estrogen receptor-mediated long-range chromatin interactions and transcription in breast cancer. *Molecular and cellular endocrinology*, 382 (1), pp.624–632. [Accessed 22 August 2025].

Liu, N. et al. (2024a). FOXA1 and FOXA2: the regulatory mechanisms and therapeutic implications in cancer. *Cell death discovery*, 10 (1), p.172. [Accessed 9 November 2025].

Liu, Q. et al. (2022b). Acute vs. chronic vs. intermittent hypoxia in breast Cancer: a review on its application in in vitro research. *Molecular biology reports*, 49 (11), pp.10961–10973. [Accessed 8 July 2025].

Liu, S. V. et al. (2025). Oncogenic gene fusions in cancer: from biology to therapy. *Signal transduction and targeted therapy*, 10 (1), p.111. [Accessed 23 October 2025].

Liu, X. et al. (2024b). The N-terminal activation function AF-1 domain of ERα interacts directly with the C-terminal AF-2-holding ligand-binding domain to recruit the coactivator proteins. *PloS one*, 19 (10), p.e0312276. [Accessed 21 October 2025].

Liu, Y. et al. (2021). FOXA2-interacting FOXP2 prevents epithelial-mesenchymal transition of breast cancer cells by stimulating E-cadherin and PHF2 transcription. *Frontiers in oncology*, 11, p.605025. [Accessed 24 November 2025].

Liu, Y.-R. et al. (2016). Comprehensive transcriptome analysis identifies novel molecular subtypes and subtype-specific RNAs of triple-negative breast cancer. *Breast cancer research: BCR*, 18 (1), p.33. [Accessed 20 October 2025].

Liu, Z. et al. (2017). *Characteristic of ER + / PR-and Ki 67 value with breast cancer*. [Online]. Available at: https://www.semanticscholar.org/paper/Characteristic-of-ER-%2B-PR-and-Ki-67-value-with-Liu-Zhang/27f8c941d4497ffcf1c2be34d48409ba463a8c81 [Accessed 20 October 2025].

Liu, Z.-J., Semenza, G. L. and Zhang, H.-F. (2015). Hypoxia-inducible factor 1 and breast cancer metastasis. *Journal of Zhejiang University. Science. B*, 16 (1), pp.32–43. [Accessed 23 October 2025].

Liverani, C. et al. (2019). A biomimetic 3D model of hypoxia-driven cancer progression. *Scientific reports*, 9 (1), p.12263. [Accessed 24 October 2025].

Loboda, A., Jozkowicz, A. and Dulak, J. (2010). HIF-1 and HIF-2 transcription factors--similar but not identical. *Molecules and cells*, 29 (5), pp.435–442. [Accessed 23 October 2025].

Loibl, S. et al. (2021). Breast cancer. *Lancet*, 397 (10286), pp.1750–1769. [Accessed 15 November 2025].

Lonkwic, K. M., Zajdel, R. and Kaczka, K. (2025). Unlocking the potential of spheroids in personalized medicine: A systematic review of seeding methodologies. *International journal of molecular sciences*, 26 (13), p.6478. [Accessed 24 October 2025].

Lopez, G. et al. (2019). Molecular insights into the classification of luminal breast cancers: The genomic heterogeneity of progesterone-negative tumors. *International journal of molecular sciences*, 20 (3), p.510. [Accessed 21 October 2025].

Louie, E. et al. (2010). Identification of a stem-like cell population by exposing metastatic breast cancer cell lines to repetitive cycles of hypoxia and reoxygenation. *Breast cancer research: BCR*, 12 (6), p.R94. [Accessed 25 November 2025].

Lovitt, C. J., Shelper, T. B. and Avery, V. M. (2014). Advanced cell culture techniques for cancer drug discovery. *Biology*, 3 (2), pp.345–367. [Accessed 3 November 2025].

Lu, M. et al. (2008). Small molecule activator of the human epithelial sodium channel. *The journal of biological chemistry*, 283 (18), pp.11981–11994.

Ludwig, C. H. and Bintu, L. (2019). Mapping chromatin modifications at the single cell level. *Development (Cambridge, England)*, 146 (12), p.dev170217. [Accessed 11 November 2025].

Ludwik, K. A. et al. (2024). Identifying the effectiveness of 3D culture systems to recapitulate breast tumor tissue in situ. *Cellular oncology*, 47 (2), pp.481–496. [Accessed 11 October 2025].

Luecken, M. D. and Theis, F. J. (2019). Current best practices in single-cell RNA-seq analysis: a tutorial. *Molecular systems biology*, 15 (6), p.e8746. [Accessed 24 November 2025].

Lumachi, F., Santeufemia, D. A. and Basso, S. M. (2015). Current medical treatment of estrogen receptor-positive breast cancer. *World journal of biological chemistry*, 6 (3), pp.231–239. [Accessed 21 October 2025].

Lutz, C. et al. (2025). The complex landscape of luminal breast cancer. *Endocrine-related cancer*, 32 (1). [Online]. Available at: doi:10.1530/ERC-24-0201 [Accessed 25 November 2025].

Lv, D. et al. (2025). Exploration of the clonal evolution and construction of the tumor clonal evolution rate as a prognostic indicator in metastatic breast cancer. *BMC medicine*, 23 (1), p.122. [Accessed 15 November 2025].

M, S., T, D. and Uk, L. (2004). ChIC and ChEC; genomic mapping of chromatin proteins. *Molecular cell*, 16 (1), pp.147–157.

Ma, S. et al. (2022). Hypoxia induces HIF1α-dependent epigenetic vulnerability in triple negative breast cancer to confer immune effector dysfunction and resistance to anti-PD-1 immunotherapy. *Nature Communications 2022 13:1*, 13 (1), pp.1–18.

Maaroufi, Y. et al. (2000). Estrogen receptor of primary breast cancers: evidence for intracellular proteolysis. *Breast cancer research: BCR*, 2 (6), pp.444–454. [Accessed 26 August 2025].

Maehara, K. et al. (2021). Modeling population size independent tissue epigenomes by ChIL-seq with single thin sections. *Molecular systems biology*, 17 (11), p.e10323. [Accessed 26 October 2025].

Magnani, L. et al. (2011). PBX1 genomic pioneer function drives ERα signaling underlying progression in breast cancer. *PLoS genetics*, 7 (11), p.e1002368. [Accessed 25 October 2025].

Maier, K. E. et al. (2023). Automated cut&run brings scalable epigenomic profiling to hematology. *Blood*, 142 (Supplement 1), pp.7150–7150. [Accessed 26 October 2025].

Maioli, S. et al. (2021). Estrogen receptors and the aging brain. *Essays in biochemistry*, 65 (6), pp.913–925. [Accessed 31 October 2025].

Maisonneuve, P. et al. (2014). Proposed new clinicopathological surrogate definitions of luminal A and luminal B (HER2-negative) intrinsic breast cancer subtypes. *Breast cancer research: BCR*, 16 (3), p.R65. [Accessed 26 November 2025].

Major, G. et al. (2024). Programming temporal stiffness cues within extracellular matrix hydrogels for modelling cancer niches. *Materials today. Bio*, 25 (101004), p.101004. [Accessed 24 October 2025].

Makhlouf, S. et al. (2024). Quantitative expression of oestrogen receptor in breast cancer: Clinical and molecular significance. *European Journal of Cancer (Oxford, England: 1990)*, 197 (113473), p.113473. [Accessed 13 March 2026].

Malcolm, J. R. et al. (2025). An ERα-dependent hypoxia response defines EMT-adjacent tumour regions and suppresses the pro-survival effects of amiloride in estrogen receptor-positive breast cancer. *bioRxiv*, p.2025.10.03.680339. [Online]. Available at: doi:10.1101/2025.10.03.680339 [Accessed 8 October 2025].

Mamnoon, B. et al. (2021). Targeting estrogen receptor-positive breast microtumors with endoxifen-conjugated, hypoxia-sensitive polymersomes. *ACS omega*, 6 (42), pp.27654–27667. [Accessed 24 October 2025].

Mangani, S. et al. (2025). Spheroid-based 3D models to decode cell function and matrix effectors in breast cancer. *Cancers*, 17 (21), p.3512. [Accessed 24 November 2025].

Mansoury, M. et al. (2021). The edge effect: A global problem. The trouble with culturing cells in 96-well plates. *Biochemistry and biophysics reports*, 26 (100987), p.100987. [Accessed 9 October 2025].

Marinov, G. K. et al. (2014). Large-scale quality analysis of published ChIP-seq data. *G3 (Bethesda, Md.)*, 4 (2), pp.209–223. [Accessed 4 November 2025].

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, 17 (1), p.10. [Accessed 24 November 2025].

Martínez-Pérez, C. et al. (2023). Neoadjuvant endocrine therapy in postmenopausal women with HR+/HER2- breast cancer. *Expert review of anticancer therapy*, 23 (1), pp.67–86. [Accessed 22 October 2025].

Masclef, L. et al. (2021). Roles and mechanisms of BAP1 deubiquitinase in tumor suppression. *Cell death and differentiation*, 28 (2), pp.606–625. [Accessed 9 November 2025].

Masoud, G. N. and Li, W. (2015). HIF-1α pathway: role, regulation and intervention for cancer therapy. *Acta pharmaceutica Sinica. B*, 5 (5), pp.378–389. [Accessed 23 October 2025].

Matei, M. B. et al. (2024). Cost-effective optimized method to process 3D tumoral spheroids in microwell arrays for immunohistochemistry analysis. *Journal of medicine and life*, 17 (6), pp.601–609.

Mavrommati, I. et al. (2021). Subclonal heterogeneity and evolution in breast cancer. *NPJ breast cancer*, 7 (1), p.155. [Accessed 26 November 2025].

Maxwell, C. B. et al. (2023). The edge effect in high-throughput proteomics: A cautionary tale. *Journal of the American Society for Mass Spectrometry*, 34 (6), pp.1065–1072. [Accessed 9 October 2025].

Maynard, K. R. et al. (2021). Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. *Nature neuroscience*, 24 (3), pp.425–436. [Accessed 23 November 2025].

Mazzucchelli, S. et al. (2024). Breast cancer patient-derived organoids for the investigation of patient-specific tumour evolution. *Cancer cell international*, 24 (1), p.220. [Accessed 24 October 2025].

McArthur, H. L. and Morris, P. G. (2010). Aromatase inhibitor strategies in metastatic breast cancer. *International journal of women's health*, 1, pp.67–72. [Accessed 22 October 2025].

McHugh, M. L. (2013). The chi-square test of independence. *Biochemia medica*, 23 (2), pp.143–149. [Accessed 22 August 2025].

McQueen, S. R. A. et al. (2025). Stable overexpression of the epithelial sodium channel alpha subunit reduces migration and proliferation in breast cancer cells. *Breast cancer research and treatment*, 211 (3), pp.595–604. [Accessed 9 July 2025].

Meers, M. P. et al. (2019). Improved CUT&RUN chromatin profiling tools. *eLife*, 8. [Online]. Available at: doi:10.7554/eLife.46314 [Accessed 26 October 2025].

Mendez, F. M. et al. (2018). Native chromatin immunoprecipitation using Murine brain tumor neurospheres. *Journal of visualized experiments: JoVE*, (131). [Online]. Available at: doi:10.3791/57016 [Accessed 25 October 2025].

Meng, Y. et al. (2009). Keratin 18 attenuates estrogen receptor alpha-mediated signaling by sequestering LRP16 in cytoplasm. *BMC cell biology*, 10 (1), p.96. [Accessed 24 November 2025].

Métivier, R. et al. (2001). Synergism between ERalpha transactivation function 1 (AF-1) and AF-2 mediated by steroid receptor coactivator protein-1: requirement for the AF-1 alpha-helical core and for a direct interaction between the N- and C-terminal domains. *Molecular endocrinology (Baltimore, Md.)*, 15 (11), pp.1953–1970. [Accessed 21 October 2025].

Métivier, R. et al. (2003). Estrogen receptor-alpha directs ordered, cyclical, and combinatorial recruitment of cofactors on a natural target promoter. *Cell*, 115 (6), pp.751–763. [Accessed 26 August 2025].

Métivier, R. et al. (2008). Dynamics of estrogen receptor-mediated transcriptional activation of responsive genes in vivo: apprehending transcription in four dimensions. *Advances in experimental medicine and biology*, 617, pp.129–138. [Accessed 26 August 2025].

Miao, R. Y. et al. (2011). MYB is essential for mammary tumorigenesis. *Cancer research*, 71 (22), pp.7029–7037. [Accessed 7 November 2025].

Mihara, H. et al. (2019). Improved oxygen supply to multicellular spheroids using A gas-permeable plate and embedded hydrogel beads. *Cells (Basel, Switzerland)*, 8 (6), p.525. [Accessed 21 November 2025].

Mikhail, A. S., Eetezadi, S. and Allen, C. (2013). Multicellular tumor spheroids for evaluation of cytotoxicity and tumor growth inhibitory effects of nanomedicines in vitro: a comparison of docetaxel-loaded block copolymer micelles and Taxotere®. *PloS one*, 8 (4), p.e62630. [Accessed 8 October 2025].

Miki, Y. et al. (1994). A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science (New York, N.Y.)*, 266 (5182), pp.66–71. [Accessed 22 October 2025].

Min, J. Y. et al. (2022). Upregulation of CYP1B1 by hypoxia is mediated by ERα activation in breast cancer cells. *American journal of cancer research*, 12 (6), pp.2798–2816. [Accessed 24 October 2025].

Mitra, P. et al. (2012). Estrogen receptor-α recruits P-TEFb to overcome transcriptional pausing in intron 1 of the MYB gene. *Nucleic acids research*, 40 (13), pp.5988–6000. [Accessed 7 November 2025].

Mitra, R., Fain-Thornton, J. and Craig, N. L. (2008). piggyBac can bypass DNA synthesis during cut and paste transposition. *The EMBO journal*, 27 (7), pp.1097–1109. [Accessed 15 October 2025].

Mittal, S., Brown, N. J. and Holen, I. (2018). The breast tumor microenvironment: role in cancer development, progression and response to therapy. *Expert review of molecular diagnostics*, 18 (3), pp.227–243. [Accessed 23 November 2025].

Miziak, P. et al. (2023). Estrogen receptor signaling in breast cancer. *Cancers*, 15 (19), p.4689. [Accessed 25 November 2025].

Mohammed, H. et al. (2015). Progesterone receptor modulates ERα action in breast cancer. *Nature*, 523 (7560), pp.313–317.

Mokhtari, R. B. et al. (2021). 3D multicellular stem-like human breast tumor spheroids enhance tumorigenicity of orthotopic xenografts in athymic nude rat model. *Cancers*, 13 (11), p.2784. [Accessed 8 October 2025].

Monje, P. et al. (2001). Differential cellular localization of estrogen receptor alpha in uterine and mammary cells. *Molecular and cellular endocrinology*, 181 (1–2), pp.117–129. [Accessed 25 November 2025].

Montani, C. et al. (2014). Fibroblasts maintained in 3 dimensions show a better differentiation state and higher sensitivity to estrogens. *Toxicology and Applied Pharmacology*, 280 (3), pp.421–433. [Accessed 13 March 2026].

Montano, M. M. et al. (1995). The carboxy-terminal F domain of the human estrogen receptor: role in the transcriptional activity of the receptor and the effectiveness of antiestrogens as estrogen antagonists. *Molecular endocrinology (Baltimore, Md.)*, 9 (7), pp.814–825. [Accessed 26 August 2025].

Monterrey, D. T. et al. (2022). Design and biocatalytic applications of genetically fused multifunctional enzymes. *Biotechnology advances*, 60 (108016), p.108016. [Accessed 26 August 2025].

Morden, J. P. et al. (2017). Long-term follow-up of the Intergroup Exemestane Study. *Journal of clinical oncology: official journal of the American Society of Clinical Oncology*, 35 (22), pp.2507–2514. [Accessed 22 October 2025].

Morgan, R. G. et al. (2018). Optimized delivery of siRNA into 3D tumor spheroid cultures in situ. *Scientific reports*, 8 (1), p.7952. [Accessed 8 October 2025].

Morita, M. et al. (2019). Fluorescence-based discrimination of breast cancer cells by direct exposure to 5-aminolevulinic acid. *Cancer medicine*, 8 (12), pp.5524–5533. [Accessed 10 October 2025].

Morotti, M. et al. (2019). Hypoxia-induced switch in SNAT2/SLC38A2 regulation generates endocrine resistance in breast cancer. *Proceedings of the National Academy of Sciences of the United States of America*, 116 (25), pp.12452–12461. [Accessed 8 July 2025].

Mou, J. et al. (2024). Research progress in tumor angiogenesis and drug resistance in breast cancer. *Cancer biology & medicine*, 21 (7), pp.1–15. [Accessed 24 October 2025].

Moudgil, A. et al. (2020). Self-Reporting Transposons Enable Simultaneous Readout of Gene Expression and Transcription Factor Binding in Single Cells. *Cell*, 182 (4), pp.992-1008.e21.

Moudgil, A. et al. (2021). The qBED track: a novel genome browser visualization for point processes. *Bioinformatics (Oxford, England)*, 37 (8), pp.1168–1170. [Accessed 6 August 2025].

Muguruma, M. et al. (2020). Differences in drug sensitivity between two-dimensional and three-dimensional culture systems in triple-negative breast cancer cell lines. *Biochemical and biophysical research communications*, 533 (3), pp.268–274.

Mukaka, M. M. (2012). Statistics corner: A guide to appropriate use of correlation coefficient in medical research. *Malawi medical journal: the journal of Medical Association of Malawi*, 24 (3), pp.69–71. [Accessed 15 October 2025].

Mukherjee, A. et al. (2020). Engineered mutant α-ENaC subunit mRNA delivered by lipid nanoparticles reduces amiloride currents in cystic fibrosis-based cell and mice models. *Science advances*, 6 (47), p.eabc5911. [Accessed 21 November 2025].

Mukomoto, R. et al. (2020). Oxygen consumption rate of tumour spheroids during necrotic-like core formation. *The Analyst*, 145 (19), pp.6342–6348. [Accessed 10 October 2025].

Mumbach, M. R. et al. (2016). HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nature methods*, 13 (11), pp.919–922. [Accessed 18 November 2025].

Mundade, R. et al. (2014). Role of ChIP-seq in the discovery of transcription factor binding sites, differential gene regulation mechanism, epigenetic marks and beyond. *Cell cycle (Georgetown, Tex.)*, 13 (18), pp.2847–2852. [Accessed 25 October 2025].

Munne, P. M. et al. (2021). Compressive stress-mediated p38 activation required for ERα + phenotype in breast cancer. *Nature Communications*, 12 (1), p.6967. [Accessed 13 March 2026].

Muñoz, L. et al. (2010). Paradoxial changes in the expression of estrogen receptor alpha in breast cancer multicellular spheroids. *Tissue & cell*, 42 (5), pp.334–337. [Accessed 8 July 2025].

Muñoz-Galindo, L. et al. (2019). Changes in the transcriptome profile of breast cancer cells grown as spheroids. *Biochemical and biophysical research communications*, 516 (4), pp.1258–1264. [Accessed 10 October 2025].

Murray, J. I. et al. (2015). Intratumoural inflammation and endocrine resistance in breast cancer. *Endocrine-related cancer*, 22 (1), pp.R51-67. [Accessed 23 October 2025].

Muz, B. et al. (2015). The role of hypoxia in cancer progression, angiogenesis, metastasis, and resistance to therapy. *Hypoxia (Auckland, N.Z.)*, 3, pp.83–92. [Accessed 23 October 2025].

Nag, S. and Resnick, A. (2017). Stabilization of hypoxia inducible factor by cobalt chloride can alter renal epithelial transport. *Physiological Reports*, 5 (24), p.e13531. [Accessed 21 November 2025].

Nagarajan, S. et al. (2014). Bromodomain protein BRD4 is required for estrogen receptor-dependent enhancer activation and gene transcription. *Cell reports*, 8 (2), pp.460–469. [Accessed 26 August 2025].

Nagarajan, S. et al. (2015). H4K12ac is regulated by estrogen receptor-alpha and is associated with BRD4 function and inducible transcription. *Oncotarget*, 6 (9), pp.7305–7317.

Nagelkerke, A. et al. (2013). Hypoxia stimulates migration of breast cancer cells via the PERK/ATF4/LAMP3-arm of the unfolded protein response. *Breast cancer research: BCR*, 15 (1), p.R2. [Accessed 24 October 2025].

Nagpal, N. et al. (2015). HIF-inducible miR-191 promotes migration in breast cancer through complex regulation of TGFβ-signaling in hypoxic microenvironment. *Scientific reports*, 5 (1), p.9650. [Accessed 8 October 2025].

Nagy, Z. and Jeselsohn, R. (2022). ESR1 fusions and therapeutic resistance in metastatic breast cancer. *Frontiers in oncology*, 12, p.1037531. [Accessed 23 October 2025].

Nahm, F. S. (2022). Receiver operating characteristic curve: overview and practical use for clinicians. *Korean journal of anesthesiology*, 75 (1), pp.25–36. [Accessed 5 November 2025].

Nakato, R. and Sakata, T. (2021). Methods for ChIP-seq analysis: A practical workflow and advanced applications. *Methods (San Diego, Calif.)*, 187, pp.44–53. [Accessed 15 October 2025].

Nakato, R. and Shirahige, K. (2017). Recent advances in ChIP-seq analysis: from quality management to whole-genome annotation. *Briefings in bioinformatics*, 18 (2), pp.279–290. [Accessed 9 November 2025].

Narlikar, L. and Jothi, R. (2012). ChIP-Seq data analysis: identification of protein-DNA binding sites with SISSRs peak-finder. *Methods in molecular biology (Clifton, N.J.)*, 802, pp.305–322. [Accessed 25 October 2025].

Naughton, C. et al. (2007). Progressive loss of estrogen receptor alpha cofactor recruitment in endocrine resistance. *Molecular endocrinology (Baltimore, Md.)*, 21 (11), pp.2615–2626. [Accessed 26 August 2025].

Nelson, J. D., Denisenko, O. and Bomsztyk, K. (2006). Protocol for the fast chromatin immunoprecipitation (ChIP) method. *Nature protocols*, 1 (1), pp.179–185. [Accessed 25 November 2025].

Neven, P. et al. (2023). Updated overall survival from the MONALEESA-3 trial in postmenopausal women with HR+/HER2- advanced breast cancer receiving first-line ribociclib plus fulvestrant. *Breast cancer research: BCR*, 25 (1), p.103. [Accessed 22 October 2025].

Nguyen, Q. H. et al. (2018). Profiling human breast epithelial cells using single cell RNA sequencing identifies cell diversity. *Nature communications*, 9 (1), p.2028. [Accessed 24 November 2025].

Nicol, S. M. and Fuller-Pace, F. V. (2010). Analysis of the RNA helicase p68 (Ddx5) as a transcriptional regulator. *Methods in molecular biology (Clifton, N.J.)*, 587, pp.265–279. [Accessed 24 November 2025].

Nik-Zainal, S. et al. (2016). Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature*, 534 (7605), pp.47–54. [Accessed 25 November 2025].

Nishide, G. et al. (2025). Zooming into gene activation: Estrogen receptor α dimerization and DNA binding visualized by high-speed atomic force microscopy. *ACS nano*, 19 (16), pp.15395–15410. [Accessed 21 October 2025].

Nishizawa-Yokoi, A. et al. (2014). Precise marker excision system using an animal-derived piggyBac transposon in plants. *The Plant journal: for cell and molecular biology*, 77 (3), pp.454–463. [Accessed 15 October 2025].

Nitsch, S. and Schneider, R. (2024). Native ChIP: Studying the genome-wide distribution of histone modifications in cells and tissue. *Methods in molecular biology (Clifton, N.J.)*, 2846, pp.1–16. [Accessed 25 October 2025].

Niu, Y., Xu, J. and Sun, T. (2019). Cyclin-dependent kinases 4/6 inhibitors in breast cancer: Current status, resistance, and combination strategies. *Journal of cancer*, 10 (22), pp.5504–5517. [Accessed 22 October 2025].

Noreng, S. et al. (2020). Molecular principles of assembly, activation, and inhibition in epithelial sodium channel. *eLife*, 9, p.e59038. [Accessed 11 June 2025].

Norris, J. L. and Hughes, R. M. (2018). protaTETHER - a method for the incorporation of variable linkers in protein fusions reveals impacts of linker flexibility in a PKAc-GFP fusion protein. *FEBS open bio*, 8 (6), pp.1029–1042. [Accessed 22 August 2025].

Nunes, A. S. et al. (2019). 3D tumor spheroids as in vitro models to mimic in vivo human solid tumors resistance to therapeutic drugs: NUNESet al. *Biotechnology and bioengineering*, 116 (1), pp.206–226. [Accessed 25 November 2025].

Obidiro, O., Battogtokh, G. and Akala, E. O. (2023). Triple negative breast cancer treatment options and limitations: Future outlook. *Pharmaceutics*, 15 (7), p.1796. [Accessed 26 November 2025].

O'Connor, L., Gilmour, J. and Bonifer, C. (2016). The role of the ubiquitously expressed transcription factor Sp1 in tissue-specific transcriptional regulation and in disease. *The Yale journal of biology and medicine*, 89 (4), pp.513–525. [Accessed 9 November 2025].

O'Donovan, P. J. and Livingston, D. M. (2010). BRCA1 and BRCA2: breast/ovarian cancer susceptibility gene products and participants in DNA double-strand break repair. *Carcinogenesis*, 31 (6), pp.961–967. [Accessed 22 October 2025].

O'Leary, B. et al. (2018). The genetic landscape and clonal evolution of breast cancer resistance to palbociclib plus fulvestrant in the PALOMA-3 trial. *Cancer discovery*, 8 (11), pp.1390–1403. [Accessed 22 October 2025].

O'Lone, R. et al. (2007). Estrogen receptors alpha and beta mediate distinct pathways of vascular gene expression, including genes involved in mitochondrial electron transport and generation of reactive oxygen species. *Molecular endocrinology (Baltimore, Md.)*, 21 (6), pp.1281–1296. [Accessed 31 October 2025].

Onitilo, A. A. et al. (2009). Breast cancer subtypes based on ER/PR and Her2 expression: comparison of clinicopathologic features and survival. *Clinical medicine & research*, 7 (1–2), pp.4–13. [Accessed 22 October 2025].

Orlando, D. A. et al. (2014). Quantitative ChIP-Seq normalization reveals global modulation of the epigenome. *Cell Reports*, 9 (3), pp.1163–1170.

Osborne, C. K. et al. (2005). Crosstalk between estrogen receptor and growth factor receptor pathways as a cause for endocrine therapy resistance in breast cancer. *Clinical cancer research: an official journal of the American Association for Cancer Research*, 11 (2 Pt 2), pp.865s–870s. [Accessed 25 November 2025].

Osborne, C. K. and Schiff, R. (2011). Mechanisms of endocrine resistance in breast cancer. *Annual review of medicine*, 62 (1), pp.233–247. [Accessed 22 October 2025].

Ostrander, J. H. et al. (2010). Optical redox ratio differentiates breast cancer cell lines based on estrogen receptor status. *Cancer research*, 70 (11), pp.4759–4766. [Accessed 10 October 2025].

Ottenbourgs, T. and Van Nieuwenhuysen, E. (2024). Novel endocrine therapeutic opportunities for estrogen receptor-positive ovarian cancer-what can we learn from breast cancer? *Cancers*, 16 (10), p.1862. [Accessed 15 November 2025].

Ovadia, E. M. et al. (2020). Understanding ER+ breast cancer dormancy using bioinspired synthetic matrices for long-term 3D culture and insights into late recurrence. *Advanced biosystems*, 4 (9), p.e2000119. [Accessed 15 November 2025].

Ozyurt, R. and Ozpolat, B. (2022). Molecular mechanisms of anti-estrogen therapy resistance and novel targeted therapies. *Cancers*, 14 (21), p.5206. [Accessed 24 October 2025].

Paakinaho, V. and Palvimo, J. J. (2021). Genome-wide crosstalk between steroid receptors in breast and prostate cancers. *Endocrine-related cancer*, 28 (9), pp.R231–R250. [Accessed 3 November 2025].

Paakkola, N.-M. et al. (2021). The prognostic and predictive impact of low estrogen receptor expression in early breast cancer: a systematic review and meta-analysis. *ESMO open*, 6 (6), p.100289. [Accessed 21 October 2025].

Padró, M. et al. (2017). Genome-independent hypoxic repression of estrogen receptor alpha in breast cancer cells. *BMC cancer*, 17 (1), p.203. [Accessed 14 July 2025].

Palaniappan, M. et al. (2019). The genomic landscape of estrogen receptor α binding sites in mouse mammary gland. *PloS one*, 14 (8), p.e0220311. [Accessed 10 November 2025].

Palma, A. et al. (2016). Metabolic study of breast MCF-7 tumor spheroids after gamma irradiation by (1)H NMR spectroscopy and microimaging. *Frontiers in oncology*, 6, p.105. [Accessed 10 October 2025].

Palmi, S. et al. (2025). The association between intrinsic breast cancer subtypes, mammography screening and prognosis: a large population-based real world cohort study. *Breast (Edinburgh, Scotland)*, 82 (104507), p.104507. [Accessed 22 October 2025].

Palstra, R.-J. and Grosveld, F. (2012). Transcription factor binding at enhancers: shaping a genomic regulatory landscape in flux. *Frontiers in Genetics*, 3, p.195. [Accessed 13 March 2026].

Pan, D. and Jia, D. (2021). Application of single-cell multi-omics in dissecting cancer cell plasticity and tumor heterogeneity. *Frontiers in molecular biosciences*, 8, p.757024. [Accessed 15 November 2025].

Pan, H. et al. (2017). 20-year risks of breast-cancer recurrence after stopping endocrine therapy at 5 years. *The New England journal of medicine*, 377 (19), pp.1836–1846. [Accessed 22 October 2025].

Pan, Y. et al. (2020). Tumor-associated macrophages in tumor immunity. *Frontiers in immunology*, 11, p.583084. [Accessed 3 November 2025].

Pan, Y. et al. (2023). TRPML1 as a potential therapeutic target for triple-negative breast cancer: a review. *Frontiers in Oncology*, 13, p.1326023. [Accessed 20 March 2026].

Pan, Y. F. et al. (2008). Regulation of estrogen receptor-mediated long range transcription via evolutionarily conserved distal response elements. *The journal of biological chemistry*, 283 (47), pp.32977–32988. [Accessed 22 August 2025].

Panter, G. and Jerala, R. (2011). The ectodomain of the Toll-like receptor 4 prevents constitutive receptor activation. *The journal of biological chemistry*, 286 (26), pp.23334–23344. [Accessed 9 November 2025].

Paolillo, M. et al. (2019). Stem-like cancer cells in a dynamic 3D culture system: A model to study metastatic cell adhesion and anti-cancer drugs. *Cells (Basel, Switzerland)*, 8 (11), p.1434. [Accessed 11 October 2025].

Pare, G. et al. (2002). Estrogen receptor-alpha mediates the protective effects of estrogen against vascular injury. *Circulation research*, 90 (10), pp.1087–1092. [Accessed 31 October 2025].

Park, J. and Lee, Y. (2017). Hypoxia induced phosphorylation of estrogen receptor at serine 118 in the absence of ligand. *The Journal of steroid biochemistry and molecular biology*, 174, pp.146–152. [Accessed 11 October 2025].

Park, M. et al. (2022). Breast cancer metastasis: Mechanisms and therapeutic implications. *International journal of molecular sciences*, 23 (12), p.6806. [Accessed 15 November 2025].

Park, P. J. (2009). ChIP-seq: advantages and challenges of a maturing technology. *Nature reviews. Genetics*, 10 (10), pp.669–680. [Accessed 11 November 2025].

Parker, J. S. et al. (2009). Supervised risk predictor of breast cancer based on intrinsic subtypes. *Journal of clinical oncology: official journal of the American Society of Clinical Oncology*, 27 (8), pp.1160–1167. [Accessed 20 October 2025].

Park-Simon, T.-W. et al. (2023). Arbeitsgemeinschaft Gynäkologische Onkologie recommendations for the Diagnosis and treatment of patients with early Breast Cancer: Update 2023. *Breast care (Basel, Switzerland)*, 18 (4), pp.289–305. [Accessed 22 October 2025].

Partridge, A. H. (2015). Chemotherapy in premenopausal breast cancer patients. *Breast care (Basel, Switzerland)*, 10 (5), pp.307–310. [Accessed 22 October 2025].

Patty, B. J. and Hainer, S. J. (2021). Transcription factor chromatin profiling genome-wide using uliCUT&RUN in single cells and individual blastocysts. *Nature Protocols 2021 16:5*, 16 (5), pp.2633–2666.

Pedersen, R. N. et al. (2022). The incidence of breast cancer recurrence 10-32 years after primary diagnosis. *Journal of the National Cancer Institute*, 114 (3), pp.391–399. [Accessed 23 October 2025].

Pejerrey, S. M. et al. (2018). The impact of ESR1 mutations on the treatment of metastatic breast cancer. *Hormones & cancer*, 9 (4), pp.215–228. [Accessed 15 November 2025].

Pernas, S. et al. (2018). CDK4/6 inhibition in breast cancer: current practice and future directions. *Therapeutic advances in medical oncology*, 10, p.1758835918786451. [Accessed 22 October 2025].

Persson, H. et al. (2017). Frequent miRNA-convergent fusion gene events in breast cancer. *Nature communications*, 8 (1), p.788. [Accessed 24 November 2025].

Piggott, L. et al. (2018). Acquired resistance of ER-positive breast cancer to endocrine treatment confers an adaptive sensitivity to TRAIL through posttranslational downregulation

of c-FLIP. *Clinical cancer research: an official journal of the American Association for Cancer Research*, 24 (10), pp.2452–2463. [Accessed 15 November 2025].

Pires, I. M. et al. (2012). Targeting radiation-resistant hypoxic tumour cells through ATR inhibition. *British journal of cancer*, 107 (2), pp.291–299. [Accessed 8 October 2025].

Pittet, M. J., Michielin, O. and Migliorini, D. (2022). Clinical relevance of tumour-associated macrophages. *Nature reviews. Clinical oncology*, 19 (6), pp.402–421. [Accessed 3 November 2025].

Ponglikitmongkol, M., Green, S. and Chambon, P. (1988). Genomic organization of the human oestrogen receptor gene. *The EMBO journal*, 7 (11), pp.3385–3388. [Accessed 20 October 2025].

Ponnusamy, S. et al. (2019). Androgen receptor is a non-canonical inhibitor of wild-type and mutant estrogen receptors in hormone receptor-positive breast cancers. *iScience*, 21, pp.341–358. [Accessed 3 November 2025].

Poon, E., Harris, A. L. and Ashcroft, M. (2009). Targeting the hypoxia-inducible factor (HIF) pathway in cancer. *Expert reviews in molecular medicine*, 11 (e26), p.e26. [Accessed 23 October 2025].

Porras, L., Ismail, H. and Mader, S. (2021). Positive regulation of estrogen receptor alpha in breast tumorigenesis. *Cells (Basel, Switzerland)*, 10 (11), p.2966. [Accessed 7 November 2025].

Powell, E. and Xu, W. (2008). Intermolecular interactions identify ligand-selective activity of estrogen receptor alpha/beta dimers. *Proceedings of the National Academy of Sciences of the United States of America*, 105 (48), pp.19012–19017. [Accessed 22 August 2025].

Pradella, D. et al. (2025). Engineered extrachromosomal oncogene amplifications promote tumorigenesis. *Nature*, 637 (8047), pp.955–964. [Accessed 23 November 2025].

Presti, D. and Quaquarini, E. (2019). The PI3K/AKT/mTOR and CDK4/6 pathways in endocrine resistant HR+/HER2- metastatic breast cancer: Biological mechanisms and new treatments. *Cancers*, 11 (9), p.1242. [Accessed 23 October 2025].

Pritchard, C. A. et al. (1995). Conditionally oncogenic forms of the A-Raf and B-Raf protein kinases display different biological and biochemical properties in NIH 3T3 cells. *Molecular and cellular biology*, 15 (11), pp.6430–6442. [Accessed 26 August 2025].

Prossnitz, E. R. and Barton, M. (2014). Estrogen biology: new insights into GPER function and clinical opportunities. *Molecular and cellular endocrinology*, 389 (1–2), pp.71–83. [Accessed 25 November 2025].

Pugh-Toole, M. et al. (2022). Tumor Spheroids Layered in an Imageable Cancer Environment (T-SLICE): a novel *in vitro* platform to study tumor biology. *bioRxiv*, p.2022.10.08.511443. [Online]. Available at: doi:10.1101/2022.10.08.511443 [Accessed 8 October 2025].

Pulze, L. et al. (2020). MCF7 Spheroid Development: New Insight about Spatio/Temporal Arrangements of TNTs, Amyloid Fibrils, Cell Connections, and Cellular Bridges. *International journal of molecular sciences*, 21 (15). [Online]. Available at: doi:10.3390/ijms21155400.

Pyne, E. et al. (2024). Investigating the impact of the interstitial fluid flow and hypoxia interface on cancer transcriptomes using a spheroid-on-chip perfusion system. *Lab on a chip*, 24 (19), pp.4609–4622. [Accessed 24 October 2025].

Quail, D. F. and Joyce, J. A. (2013). Microenvironmental regulation of tumor progression and metastasis. *Nature medicine*, 19 (11), pp.1423–1437. [Accessed 23 November 2025].

Quinlan, A. R. and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics (Oxford, England)*, 26 (6), pp.841–842. [Accessed 18 November 2025].

Raghavan, S. et al. (2016). Comparative analysis of tumor spheroid generation techniques for differential in vitro drug toxicity. *Oncotarget*, 7 (13), pp.16948–16961. [Accessed 24 October 2025].

Ragnum, H. B. et al. (2015). The tumour hypoxia marker pimonidazole reflects a transcriptional programme associated with aggressive prostate cancer. *British journal of cancer*, 112 (2), pp.382–390. [Accessed 8 October 2025].

Rajbhandari, P. et al. (2012). Regulation of estrogen receptor α N-terminus conformation and function by peptidyl prolyl isomerase Pin1. *Molecular and cellular biology*, 32 (2), pp.445–457. [Accessed 21 October 2025].

Rana, N. K., Singh, P. and Koch, B. (2019). CoCl2 simulated hypoxia induce cell proliferation and alter the expression pattern of hypoxia associated genes involved in angiogenesis and apoptosis. *Biological research*, 52 (1), p.12. [Accessed 30 October 2025].

Rangel, L. B. A. and Huang, T. H.-M. (2013). Estrogen response in luminal breast cancer. *Oncotarget*, 4 (10), pp.1548–1549. [Accessed 25 November 2025].

Rao, S. S. P. et al. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, 159 (7), pp.1665–1680. [Accessed 24 November 2025].

Raza, A. et al. (2017). Oxygen mapping of melanoma spheroids using small molecule platinum probe and phosphorescence lifetime imaging microscopy. *Scientific reports*, 7 (1), p.10743. [Accessed 24 October 2025].

Razavi, P. et al. (2018). The genomic landscape of endocrine-resistant advanced breast cancers. *Cancer cell*, 34 (3), pp.427-438.e6. [Accessed 26 November 2025].

Redig, A. J. and McAllister, S. S. (2013). Breast cancer as a systemic disease: a view of metastasis. *Journal of internal medicine*, 274 (2), pp.113–126. [Accessed 15 November 2025].

Redmond, J. et al. (2021). Advances in biofabrication techniques for collagen-based 3D in vitro culture models for breast cancer research. *Materials science & engineering. C, Materials for biological applications*, 122 (111944), p.111944. [Accessed 24 October 2025].

Reed, A. D. et al. (2024). A single-cell atlas enables mapping of homeostatic cellular shifts in the adult human breast. *Nature genetics*, 56 (4), pp.652–662. [Accessed 22 October 2025].

Regan, M. M. et al. (2011). Assessment of letrozole and tamoxifen alone and in sequence for postmenopausal women with steroid hormone receptor-positive breast cancer: the BIG 1-98 randomised clinical trial at 8·1 years median follow-up. *The lancet oncology*, 12 (12), pp.1101–1108.

Reuten, R. et al. (2016). Maltose-binding protein (MBP), a secretion-enhancing tag for mammalian protein expression systems. *PloS one*, 11 (3), p.e0152386. [Accessed 9 November 2025].

Revankar, C. M. et al. (2005). A transmembrane intracellular estrogen receptor mediates rapid cell signaling. *Science (New York, N.Y.)*, 307 (5715), pp.1625–1630. [Accessed 21 October 2025].

Rezania, S. et al. (2016). Overexpression of KCNJ3 gene splice variants affects vital parameters of the malignant breast cancer cell line MCF-7 in an opposing manner. *BMC cancer*, 16 (1), p.628. [Accessed 14 September 2025].

Riggio, A. I., Varley, K. E. and Welm, A. L. (2021). The lingering mysteries of metastatic recurrence in breast cancer. *British journal of cancer*, 124 (1), pp.13–26. [Accessed 18 September 2025].

Robertson, G. et al. (2007). Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nature methods*, 4 (8), pp.651–657. [Accessed 25 October 2025].

Robinson, D. R. et al. (2013). Activating ESR1 mutations in hormone-resistant metastatic breast cancer. *Nature genetics*, 45 (12), pp.1446–1451. [Accessed 22 October 2025].

Robinson, J. L. L. and Carroll, J. S. (2012). FoxA1 is a key mediator of hormonal response in breast and prostate cancer. *Frontiers in endocrinology*, 3, p.68. [Accessed 26 August 2025].

Rofstad, E. K. et al. (1996). Apoptosis, energy metabolism, and fraction of radiobiologically hypoxic cells: a study of human melanoma multicellular spheroids. *International journal of radiation biology*, 70 (3), pp.241–249. [Accessed 11 October 2025].

Romond, E. H. et al. (2005). Trastuzumab plus adjuvant chemotherapy for operable HER2-positive breast cancer. *The New England journal of medicine*, 353 (16), pp.1673–1684. [Accessed 22 October 2025].

Ross-Innes, C. S. et al. (2010). Cooperative interaction between retinoic acid receptor-alpha and estrogen receptor in breast cancer. *Genes & development*, 24 (2), pp.171–182. [Accessed 7 November 2025].

Ross-Innes, C. S. et al. (2012). Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature*, 481 (7381), pp.389–393.

Rotem, A. et al. (2015). Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state. *Nature biotechnology*, 33 (11), pp.1165–1172.

Roth, C. et al. (2023). Improved quality metrics for association and reproducibility in chromatin accessibility data using mutual information. *BMC bioinformatics*, 24 (1), p.441. [Accessed 15 October 2025].

Russo, J. et al. (1976). Reexpression of the original tumor pattern by a human breast carcinoma cell line (MCF-7) in sponge culture. *Journal of the National Cancer Institute*, 56 (2), pp.279–282. [Accessed 10 November 2025].

Ryu, K., Park, C. and Lee, Y. (2011). Hypoxia-inducible factor 1 alpha represses the transcription of the estrogen receptor alpha gene in human breast cancer cells. *Biochemical and biophysical research communications*, 407 (4), pp.831–836. [Accessed 24 October 2025].

Ş. Comşa, Cîmpean, A. and Raica, M. (2015). The story of MCF-7 breast cancer cell line: 40 years of experience in research. *Anticancer research*, 35 (6), pp.3147–3154. [Accessed 26 August 2025].

Sachs, N. et al. (2018). A living biobank of breast cancer organoids captures disease heterogeneity. *Cell*, 172 (1–2), pp.373-386.e10. [Accessed 24 October 2025].

Sadaf et al. (2023). Role of fork-head box genes in breast cancer: From drug resistance to therapeutic targets. *Biomedicines*, 11 (8), p.2159. [Accessed 20 March 2026].

Safe, S. (2001). Transcriptional activation of genes by 17 beta-estradiol through estrogen receptor-Sp1 interactions. *Vitamins and hormones*, 62, pp.231–252. [Accessed 26 August 2025].

Saha, A. et al. (2018). Intra-tumor molecular heterogeneity in breast cancer: definitions of measures and association with distant recurrence-free survival. *Breast cancer research and treatment*, 172 (1), pp.123–132. [Accessed 23 October 2025].

Sahtoe, D. D. et al. (2016). BAP1/ASXL1 recruitment and activation for H2A deubiquitination. *Nature communications*, 7 (1), p.10292. [Accessed 24 November 2025].

Sakalem, M. E. et al. (2021). Historical evolution of spheroids and organoids, and possibilities of use in life sciences and medicine. *Biotechnology journal*, 16 (5), p.e2000463. [Accessed 24 October 2025].

Salma, M. et al. (2023). High-throughput methods for the analysis of transcription factors and chromatin modifications: Low input, single cell and spatial genomic technologies. *Blood cells, molecules & diseases*, 101 (102745), p.102745. [Accessed 26 October 2025].

Salminen, A., Kaarniranta, K. and Kauppinen, A. (2016). Hypoxia-inducible histone lysine demethylases: Impact on the aging process and age-related diseases. *Aging and disease*, 7 (2), pp.180–200. [Accessed 25 November 2025].

Sandström, J. et al. (2024). GATA3 and markers of epithelial-mesenchymal transition predict long-term benefit from tamoxifen in ER-positive breast cancer. *NPJ breast cancer*, 10 (1), p.78. [Accessed 26 August 2025].

Santen, R. J. et al. (1977). Kinetic, hormonal and clinical studies with aminoglutethimide in breast cancer. *Cancer*, 39 (6 Suppl), pp.2948–2958. [Accessed 22 October 2025].

Santen, R. J. et al. (2017). Managing menopausal symptoms and associated clinical issues in breast cancer survivors. *The journal of clinical endocrinology and metabolism*, 102 (10), pp.3647–3661. [Accessed 22 October 2025].

Scherrer, L. C. et al. (1993). Evidence that the hormone binding domain of steroid receptors confers hormonal control on chimeric proteins by determining their hormone-regulated binding to heat-shock protein 90. *Biochemistry*, 32 (20), pp.5381–5386. [Accessed 26 August 2025].

Schiavon, G. and Smith, I. E. (2014). Status of adjuvant endocrine therapy for breast cancer. *Breast cancer research: BCR*, 16 (2), p.206. [Accessed 22 October 2025].

Schiff, R. et al. (2004). Cross-talk between estrogen receptor and growth factor pathways as a molecular target for overcoming endocrine resistance. *Clinical cancer research: an official journal of the American Association for Cancer Research*, 10 (1), pp.331s–336s. [Accessed 25 November 2025].

Schindelin, J. et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nature methods*, 9 (7), pp.676–682. [Accessed 24 November 2025].

Schneider, C. A., Rasband, W. S. and Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature methods*, 9 (7), pp.671–675. [Accessed 24 November 2025].

Schneider, R. et al. (2011). Aromatase inhibitors in the treatment of breast cancer in post-menopausal female patients: an update. *Breast cancer (Dove Medical Press)*, 3, pp.113–125. [Accessed 22 October 2025].

Schwabe, J. W. et al. (1993). The crystal structure of the estrogen receptor DNA-binding domain bound to DNA: how receptors discriminate between their response elements. *Cell*, 75 (3), pp.567–578. [Accessed 20 October 2025].

Sebestyén, A. et al. (2021). Hypoxia signaling in cancer: From basics to clinical practice. *Pathology oncology research: POR*, 27, p.1609802. [Accessed 23 October 2025].

Selich, A. et al. (2016). Massive clonal selection and transiently contributing clones during expansion of mesenchymal stem cell cultures revealed by Lentiviral RGB-barcode technology: Clonal selection during expansion of MSC cultures. *Stem cells translational medicine*, 5 (5), pp.591–601. [Accessed 24 November 2025].

Sella, T. et al. (2022). Optimal endocrine therapy in premenopausal women: A pragmatic approach to unanswered questions. *JCO oncology practice*, 18 (3), pp.211–216. [Accessed 22 October 2025].

Semenza, G. L. (2004). Hydroxylation of HIF-1: oxygen sensing at the molecular level. *Physiology (Bethesda, Md.)*, 19 (4), pp.176–182. [Accessed 23 October 2025].

Semenza, G. L. (2012). Hypoxia-inducible factors in physiology and medicine. *Cell*, 148 (3), pp.399–408. [Accessed 24 October 2025].

Senthebane, D. A. et al. (2017). The role of tumor microenvironment in chemoresistance: To survive, keep your enemies closer. *International journal of molecular sciences*, 18 (7), p.1586. [Accessed 15 November 2025].

Sermet-Gaudelus, I. et al. (2007). In vitroprediction of stop-codon suppression by intravenous gentamicin in patients with cystic fibrosis: a pilot study. *BMC Medicine*, 5 (1), p.5. [Accessed 21 November 2025].

Sever, R. and Glass, C. K. (2013). Signaling by nuclear receptors. *Cold Spring Harbor perspectives in biology*, 5 (3), p.a016709. [Accessed 21 October 2025].

Sha, R. K. et al. (Eds). (2024). *Global burden of breast cancer and attributable risk factors in 204 countries and territories, from 1990 to 2021: results from the Global Burden of Disease Study 2021*. Biomarker Research. [Online]. Available at: doi:10.1186/s40364-024-00631-8.

Shabani, N. et al. (2007). Prognostic significance of oestrogen receptor alpha (ERalpha) and beta (ERbeta), progesterone receptor A (PR-A) and B (PR-B) in endometrial carcinomas. *European journal of cancer (Oxford, England: 1990)*, 43 (16), pp.2434–2444. [Accessed 31 October 2025].

Shah, A. T. et al. (2017). Autofluorescence flow sorting of breast cancer cell metabolism. *Journal of biophotonics*, 10 (8), pp.1026–1033. [Accessed 10 October 2025].

Shamis, S. A. K. et al. (2022). The relationship between the tumor cell expression of hypoxic markers and survival in patients with ER-positive invasive ductal breast cancer. *The journal of histochemistry and cytochemistry: official journal of the Histochemistry Society*, 70 (7), pp.479–494. [Accessed 11 October 2025].

Shanabag, A. et al. (2025). Targeting CDK4/6 in breast cancer. *Experimental & molecular medicine*, 57 (2), pp.312–322. [Accessed 22 October 2025].

Shang, Y. et al. (2000). Cofactor dynamics and sufficiency in estrogen receptor-regulated transcription. *Cell*, 103 (6), pp.843–852. [Accessed 26 August 2025].

Shao, Z. et al. (1994). Expression of the retinoic Acid nuclear receptors (rars) and retinoid x-receptor (rxr) genes in estrogen-receptor positive and negative breast-cancer. *International journal of oncology*, 4 (4), pp.859–863. [Accessed 7 November 2025].

Sharaf, B. et al. (2024). Next generation selective estrogen receptor degraders in postmenopausal women with advanced-stage hormone receptors-positive, HER2-negative breast cancer. *Frontiers in oncology*, 14, p.1385577. [Accessed 22 October 2025].

Sherr, C. J., Beach, D. and Shapiro, G. I. (2016). Targeting CDK4 and CDK6: From discovery to therapy. *Cancer discovery*, 6 (4), pp.353–367. [Accessed 22 October 2025].

Shi, P. et al. (2022). Fundamental and practical approaches for single-cell ATAC-seq analysis. *aBIOTECH*, 3 (3), pp.212–223. [Accessed 23 November 2025].

Shi, W. et al. (2023). Embedded bioprinting of breast tumor cells and organoids using low-concentration collagen-based bioinks. *Advanced healthcare materials*, 12 (26), p.e2300905. [Accessed 24 October 2025].

Shi, Y. and Gilkes, D. M. (2025). HIF-1 and HIF-2 in cancer: structure, regulation, and therapeutic prospects. *Cellular and molecular life sciences: CMLS*, 82 (1), p.44. [Accessed 23 October 2025].

Shiau, A. K. et al. (1998). The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen. *Cell*, 95 (7), pp.927–937. [Accessed 20 October 2025].

Shiino, S. et al. (2016). Prognostic impact of discordance in hormone receptor status between primary and recurrent sites in patients with recurrent breast cancer. *Clinical breast cancer*, 16 (4), pp.e133-40. [Accessed 23 October 2025].

Shiino, S. et al. (2022). Prognostic significance of receptor expression discordance between primary and recurrent breast cancers: a meta-analysis. *Breast cancer research and treatment*, 191 (1), pp.1–14. [Accessed 22 October 2025].

Shim, H. et al. (1997). c-Myc transactivation of LDH-A: implications for tumor metabolism and growth. *Proceedings of the National Academy of Sciences of the United States of America*, 94 (13), pp.6658–6663. [Accessed 22 August 2025].

Simoncini, T. et al. (2000). Interaction of oestrogen receptor with the regulatory subunit of phosphatidylinositol-3-OH kinase. *Nature*, 407 (6803), pp.538–541. [Accessed 21 October 2025].

Singh, M. et al. (2016). Three-dimensional breast cancer models mimic hallmarks of size-induced tumor progression. *Cancer research*, 76 (13), pp.3732–3743. [Accessed 8 October 2025].

Skene, P. J. and Henikoff, S. (2017). An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *eLife*, 6. [Online]. Available at: doi:10.7554/ELIFE.21856.

Skolariki, A. et al. (2022). Role of PI3K/Akt/mTOR pathway in mediating endocrine resistance: concept to clinic. *Exploration of targeted anti-tumor therapy*, 3 (2), pp.172–199. [Accessed 23 October 2025].

Slattery, M. et al. (2014). Absence of a simple code: how transcription factors read the genome. *Trends in Biochemical Sciences*, 39 (9), pp.381–399. [Accessed 26 March 2026].

Smith, I. E. et al. (1978). Aminoglutethimide in treatment of metastatic breast carcinoma. *Lancet*, 2 (8091), pp.646–649. [Accessed 22 October 2025].

Smits, N. and Faulkner, G. J. (2023). Nanopore sequencing to identify transposable element insertions and their epigenetic modifications. *Methods in molecular biology (Clifton, N.J.)*, 2607, pp.151–171. [Accessed 24 November 2025].

Solomon, M. J., Larsen, P. L. and Varshavsky, A. (1988). Mapping proteinDNA interactions in vivo with formaldehyde: Evidence that histone H4 is retained on a highly transcribed gene. *Cell*, 53 (6), pp.937–947.

Somasundaram, K. (2010). BRCA1 and BRCA1 genes and inherited breast and/or ovarian cancer: Benefits of genetic testing. *Indian journal of surgical oncology*, 1 (3), pp.245–249. [Accessed 22 October 2025].

Spadazzi, C. et al. (2021). Trefoil factor-1 upregulation in estrogen-receptor positive breast cancer correlates with an increased risk of bone metastasis. *Bone*, 144 (115775), p.115775.

Sparano, J. A. et al. (2024). Clinical and genomic risk for late breast cancer recurrence and survival. *NEJM evidence*, 3 (8), p.EVIDoa2300267. [Accessed 22 October 2025].

Spence, R. D. et al. (2011). Neuroprotection mediated through estrogen receptor-alpha in astrocytes. *Proceedings of the National Academy of Sciences of the United States of America*, 108 (21), pp.8867–8872. [Accessed 31 October 2025].

Spoerke, J. M. et al. (2016). Heterogeneity and clinical significance of ESR1 mutations in ER-positive metastatic breast cancer patients receiving fulvestrant. *Nature communications*, 7, p.11579. [Accessed 15 November 2025].

Spring, L. M. et al. (2016). Neoadjuvant endocrine therapy for estrogen receptor-positive breast cancer: A systematic review and meta-analysis. *JAMA oncology*, 2 (11), pp.1477–1486. [Accessed 22 October 2025].

Squair, J. W. et al. (2021). Confronting false discoveries in single-cell differential expression. *Nature communications*, 12 (1), p.5692. [Accessed 23 November 2025].

Ståhl, P. L. et al. (2016). Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science (New York, N.Y.)*, 353 (6294), pp.78–82. [Accessed 23 November 2025].

Steiner, H. R. et al. (2022). An extended DNA binding domain of the estrogen receptor alpha directly interacts with RNAs in vitro. *Biochemistry*, 61 (22), pp.2490–2494. [Accessed 21 October 2025].

Stenning, J. (2026). *JackStenning/Thesis: Corrected Theis Reference*. Zenodo. [Online]. Available at: doi:10.5281/ZENODO.19235973 [Accessed 26 March 2026].

Stergachis, A. B. et al. (2013). Exonic transcription factor binding directs codon choice and affects protein evolution. *Science (New York, N.Y.)*, 342 (6164), pp.1367–1372. [Accessed 26 March 2026].

Stirling, D. R. et al. (2021). CellProfiler 4: improvements in speed, utility and usability. *BMC bioinformatics*, 22 (1), p.433. [Accessed 24 November 2025].

Stock, K. et al. (2016). Capturing tumor complexity in vitro: Comparative analysis of 2D and 3D tumor models for drug discovery. *Scientific reports*, 6 (1), p.28951. [Accessed 11 October 2025].

Stoner, M. et al. (2002). Hypoxia Induces Proteasome-Dependent Degradation of Estrogen Receptor α in ZR-75 Breast Cancer Cells. *Molecular Endocrinology*, 16 (10), pp.2231–2242. [Accessed 24 November 2025].

Stortz, M. et al. (2017). Mapping the dynamics of the glucocorticoid receptor within the nuclear landscape. *Scientific reports*, 7 (1), p.6219. [Accessed 25 November 2025].

Stravodimou, A. and Voutsadakis, I. A. (2025). The level of estrogen receptor (ER) expression and the length of adjuvant hormonal therapy in ER positive breast cancer. *Gland Surgery*, 14 (2), pp.246–251. [Accessed 13 March 2026].

Strillacci, A. et al. (2022). ERα-LBD, an isoform of estrogen receptor alpha, promotes breast cancer proliferation and endocrine resistance. *NPJ breast cancer*, 8 (1), p.96. [Accessed 21 October 2025].

Sun, J., Nawaz, Z. and Slingerland, J. M. (2007). Long-range activation of GREB1 by estrogen receptor via three distal consensus estrogen-responsive elements in breast cancer cells. *Molecular endocrinology (Baltimore, Md.)*, 21 (11), pp.2651–2662. [Accessed 26 August 2025].

Sun, J.-M. et al. (2005). Estrogen regulation of trefoil factor 1 expression by estrogen receptor alpha and Sp proteins. *Experimental cell research*, 302 (1), pp.96–107. [Accessed 17 September 2025].

Sundaram, A. Y. M. et al. (2016). A comparative study of ChIP-seq sequencing library preparation methods. *BMC genomics*, 17 (1), p.816. [Accessed 24 November 2025].

Sung, H. et al. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: a cancer journal for clinicians*, 71 (3), pp.209–249.

Sutherland, R. L., Hall, R. E. and Taylor, I. W. (1983). Cell proliferation kinetics of MCF-7 human mammary carcinoma cells in culture and effects of tamoxifen on exponentially growing and plateau-phase cells. *Cancer research*, 43 (9), pp.3998–4006. [Accessed 15 September 2025].

Svensson, V., Vento-Tormo, R. and Teichmann, S. A. (2018). Exponential scaling of single-cell RNA-seq in the past decade. *Nature Protocols*, 13 (4), pp.599–604. [Accessed 24 November 2025].

Swaby, R. F., Sharma, C. G. N. and Jordan, V. C. (2007). SERMs for the treatment and prevention of breast cancer. *Reviews in endocrine & metabolic disorders*, 8 (3), pp.229–239. [Accessed 21 October 2025].

Swartz, J. E. et al. (2022). Correlation and colocalization of HIF-1α and pimonidazole staining for hypoxia in laryngeal squamous cell carcinomas: A digital, single-cell-based analysis. *Oral oncology*, 128 (105862), p.105862. [Accessed 8 October 2025].

Swinstead, E. E. et al. (2016). Steroid receptors reprogram FoxA1 occupancy through dynamic chromatin transitions. *Cell*, 165 (3), pp.593–605. [Accessed 3 November 2025].

Szostakowska, M. et al. (2019). Resistance to endocrine therapy in breast cancer: molecular mechanisms and future goals. *Breast cancer research and treatment*, 173 (3), pp.489–497.

Takaku, M. et al. (2020). Cancer-specific mutation of GATA3 disrupts the transcriptional regulatory network governed by Estrogen Receptor alpha, FOXA1 and GATA3. *Nucleic Acids Research*, 48 (9), pp.4756–4768. [Accessed 21 March 2026].

Tangen, I. L. et al. (2014). Switch in FOXA1 status associates with endometrial cancer progression. *PloS one*, 9 (5), p.e98069. [Accessed 31 October 2025].

Taskindoust, M. et al. (2021). Survival outcomes among patients with metastatic breast cancer: Review of 47,000 patients. *Annals of surgical oncology*, 28 (12), pp.7441–7449. [Accessed 15 November 2025].

Teichgraeber, D. C., Guirguis, M. S. and Whitman, G. J. (2021). Breast cancer staging: Updates in the AJCC Cancer Staging Manual, 8th edition, and current challenges for radiologists, from the AJR special series on cancer staging. *AJR. American journal of roentgenology*, 217 (2), pp.278–290. [Accessed 25 November 2025].

Teiwes, J. and Toto, R. D. (2007). Epithelial sodium channel inhibition in cardiovascular disease. A potential role for amiloride. *American journal of hypertension*, 20 (1), pp.109–117. [Accessed 10 July 2025].

Telarovic, I., Wenger, R. H. and Pruschy, M. (2021). Interfering with tumor hypoxia for radiotherapy optimization. *Journal of experimental & clinical cancer research: CR*, 40 (1), p.197. [Accessed 23 October 2025].

Tevis, K. M. et al. (2017). Mimicking the tumor microenvironment to regulate macrophage phenotype and assessing chemotherapeutic efficacy in embedded cancer cell/macrophage spheroid models. *Acta biomaterialia*, 50, pp.271–279. [Accessed 8 October 2025].

Tevlek, A. et al. (2023). Spheroid engineering in microfluidic devices. *ACS omega*, 8 (4), pp.3630–3649. [Accessed 24 October 2025].

Teytelman, L. et al. (2013). Highly expressed loci are vulnerable to misleading ChIP localization of multiple unrelated proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 110 (46), pp.18602–18607. [Accessed 10 November 2025].

Theodorou, V. et al. (2013). GATA3 acts upstream of FOXA1 in mediating ESR1 binding by shaping enhancer accessibility. *Genome research*, 23 (1), pp.12–22. [Accessed 21 October 2025].

Theodorou, V. and Carroll, J. S. (2010). Estrogen receptor action in three dimensions - looping the loop. *Breast cancer research: BCR*, 12 (1), p.303. [Accessed 26 August 2025].

Thienpont, B. et al. (2016). Tumour hypoxia causes DNA hypermethylation by reducing TET activity. *Nature*, 537 (7618), pp.63–68. [Accessed 23 November 2025].

Thomas, P. et al. (2005). Identity of an estrogen membrane receptor coupled to a G protein in human breast cancer cells. *Endocrinology*, 146 (2), pp.624–632. [Accessed 21 October 2025].

Thomas, R. et al. (2017). Features that define the best ChIP-seq peak calling algorithms. *Briefings in bioinformatics*, 18 (3), pp.441–450. [Accessed 15 October 2025].

Tirpe, A. A. et al. (2019). Hypoxia: Overview on hypoxia-mediated mechanisms with a focus on the role of HIF genes. *International journal of molecular sciences*, 20 (24), p.6140. [Accessed 23 October 2025].

Tolaney, S. M. et al. (2023). AMEERA-3: Randomized phase II study of amcenestrant (oral selective estrogen receptor degrader) versus standard endocrine monotherapy in estrogen receptor-positive, human epidermal growth factor receptor 2-negative advanced breast cancer. *Journal of clinical oncology: official journal of the American Society of Clinical Oncology*, 41 (24), pp.4014–4024. [Accessed 25 November 2025].

Tremblay, A. et al. (1999). Ligand-independent recruitment of SRC-1 to estrogen receptor beta through phosphorylation of activation function AF-1. *Molecular cell*, 3 (4), pp.513–519. [Accessed 21 October 2025].

Tremont, A., Lu, J. and Cole, J. T. (2017). Endocrine therapy for early breast cancer: Updated review. *The Ochsner journal*, 17 (4), pp.405–411. [Accessed 22 October 2025].

Trukhacheva, E. et al. (2009). Estrogen receptor (ER) beta regulates ERalpha expression in stromal cells derived from ovarian endometriosis. *The journal of clinical endocrinology and metabolism*, 94 (2), pp.615–622. [Accessed 21 October 2025].

Tukey, J. W. (1949). Comparing individual means in the analysis of variance. *Biometrics*, 5 (2), pp.99–114. [Accessed 24 November 2025].

Tullius, T. D. and Dombroski, B. A. (1986). Hydroxyl radical 'footprinting': high-resolution information about DNA-protein contacts and application to lambda repressor and Cro protein. *Proceedings of the National Academy of Sciences of the United States of America*, 83 (15), pp.5469–5473. [Accessed 25 October 2025].

Turashvili, G. and Brogi, E. (2017). Tumor heterogeneity in breast cancer. *Frontiers in medicine*, 4, p.227. [Accessed 20 October 2025].

Turova, P. et al. (2025). The Breast Cancer Classifier refines molecular breast cancer classification to delineate the HER2-low subtype. *NPJ breast cancer*, 11 (1), p.19. [Accessed 20 October 2025].

Vakhshiteh, F. et al. (2023). Heterotypic tumor spheroids: a platform for nanomedicine evaluation. *Journal of nanobiotechnology*, 21 (1), p.249. [Accessed 3 November 2025].

Valášek, L. S. (2013). Kozak consensus sequence. In: *Encyclopedia of Systems Biology*. New York, NY: Springer New York. pp.1087–1087. [Accessed 26 August 2025].

Valencia, G. A. et al. (2022). Immunotherapy in triple-negative breast cancer: A literature review and new advances. *World journal of clinical oncology*, 13 (3), pp.219–236. [Accessed 26 November 2025].

Vale-Silva, L. A., Markowitz, T. E. and Hochwagen, A. (2019). SNP-ChIP: a versatile and tag-free method to quantify changes in protein binding across the genome. *BMC Genomics*, 20 (1), p.54.

Varia, M. A. et al. (1998). Pimonidazole: a novel hypoxia marker for complementary study of tumor hypoxia and cell proliferation in cervical carcinoma. *Gynecologic oncology*, 71 (2), pp.270–277. [Accessed 8 October 2025].

Vaupel, P. (2004). The role of hypoxia-induced factors in tumor progression. *The oncologist*, 9 Suppl 5 (S5), pp.10–17. [Accessed 24 October 2025].

Vinci, M. et al. (2012). Advances in establishment and analysis of three-dimensional tumor spheroid-based functional assays for target validation and drug evaluation. *BMC biology*, 10 (1), p.29. [Accessed 3 November 2025].

Visniauskas, B. et al. (2023). Estrogen-mediated mechanisms in hypertension and other cardiovascular diseases. *Journal of human hypertension*, 37 (8), pp.609–618. [Accessed 31 October 2025].

Walens, A. et al. (2020). Adaptation and selection shape clonal evolution of tumors during residual disease and recurrence. *Nature communications*, 11 (1), p.5017. [Accessed 15 November 2025].

Walsh, C. A. et al. (2012). The function of steroid receptor coactivator-1 in normal tissues and cancer. *International journal of biological sciences*, 8 (4), pp.470–485. [Accessed 23 October 2025].

Walter, P. et al. (1985). Cloning of the human estrogen receptor cDNA. *Proceedings of the National Academy of Sciences of the United States of America*, 82 (23), pp.7889–7893. [Accessed 20 October 2025].

Wang, G. L. et al. (1995). Hypoxia-inducible factor 1 is a basic-helix-loop-helix-PAS heterodimer regulated by cellular O2 tension. *Proceedings of the National Academy of Sciences*, 92 (12), pp.5510–5514.

Wang, G. L. and Semenza, G. L. (1995). Purification and characterization of hypoxia-inducible factor 1. *The journal of biological chemistry*, 270 (3), pp.1230–1237. [Accessed 23 October 2025].

Wang, H. et al. (2011). Calling Cards enable multiplexed identification of the genomic targets of DNA-binding proteins. *Genome research*, 21 (5), pp.748–755.

Wang, H. et al. (2012a). 'Calling cards' for DNA-binding proteins in mammalian cells. *Genetics*, 190 (3), pp.941–949.

Wang, H., Johnston, M. and Mitra, R. D. (2007). Calling cards for DNA-binding proteins. *Genome research*, 17 (8), pp.1202–1209.

Wang, J. et al. (2012b). The 2-nitroimidazole EF5 is a biomarker for oxidoreductases that activate the bioreductive prodrug CEN-209 under hypoxia. *Clinical cancer research: an official journal of the American Association for Cancer Research*, 18 (6), pp.1684–1695. [Accessed 8 October 2025].

Wang, J. et al. (2013). Genome-wide analysis uncovers high frequency, strong differential chromosomal interactions and their associated epigenetic patterns in E2-mediated gene regulation. *BMC genomics*, 14 (1), p.70. [Accessed 10 November 2025].

Wang, Q. et al. (2019a). CoBATCH for High-Throughput Single-Cell Epigenomic Profiling. *Molecular cell*, 76 (1), pp.206-216.e7.

Wang, R. et al. (2014a). Comparison of mammosphere formation from breast cancer cell lines and primary breast tumors. *Journal of thoracic disease*, 6 (6), pp.829–837. [Accessed 11 October 2025].

Wang, R. et al. (2019b). The Clinicopathological features and survival outcomes of patients with different metastatic sites in stage IV breast cancer. *BMC cancer*, 19 (1), p.1091. [Accessed 15 November 2025].

Wang, S. et al. (2023). The evolution of single-cell RNA sequencing technology and application: Progress and perspectives. *International journal of molecular sciences*, 24 (3), p.2943. [Accessed 24 October 2025].

Wang, X. et al. (2024). Spatial transcriptomics reveals substantial heterogeneity in triple-negative breast cancer with potential clinical implications. *Nature communications*, 15 (1), p.10232. [Accessed 20 October 2025].

Wang, Y. et al. (2014b). Clonal evolution in breast cancer revealed by single nucleus genome sequencing. *Nature*, 512 (7513), pp.155–160. [Accessed 15 November 2025].

Wang, Y. et al. (2018). 3D hydrogel breast cancer models for studying the effects of hypoxia on epithelial to mesenchymal transition. *Oncotarget*, 9 (63), pp.32191–32203. [Accessed 24 October 2025].

Wang, Y. and Tang, S.-C. (2022). The race to develop oral SERDs and other novel estrogen receptor inhibitors: recent clinical trial results and impact on treatment options. *Cancer metastasis reviews*, 41 (4), pp.975–990. [Accessed 22 October 2025].

Wanigasekara, J. et al. (2023). Three-Dimensional (3D) in vitro cell culture protocols to enhance glioblastoma research. *PloS one*, 18 (2), p.e0276248. [Accessed 3 November 2025].

Wardle, F. C. and Tan, H. (2015). A ChIP on the shoulder? Chromatin immunoprecipitation and validation strategies for ChIP antibodies. *F1000Research*, 4, p.235. [Accessed 25 October 2025].

Ware, A. W. et al. (2021). The epithelial sodium channel has a role in breast cancer cell proliferation. *Breast cancer research and treatment*, 187 (1), pp.31–43. [Accessed 9 July 2025].

Webster, T. H. et al. (2019). Identifying, understanding, and correcting technical artifacts on the sex chromosomes in next-generation sequencing data. *GigaScience*, 8 (7), p.giz074. [Accessed 10 November 2025].

Weikum, E. R., Liu, X. and Ortlund, E. A. (2018). The nuclear receptor superfamily: A structural perspective: The Nuclear Receptor Superfamily. *Protein science: a publication of the Protein Society*, 27 (11), pp.1876–1892. [Accessed 21 October 2025].

Welboren, W.-J. et al. (2009). ChIP-Seq of ERalpha and RNA polymerase II defines genes differentially responding to ligands. *The EMBO journal*, 28 (10), pp.1418–1428. [Accessed 25 October 2025].

Wen, L. and Tang, F. (2025). Single-cell omics sequencing technologies: the long-read generation. *Trends in genetics: TIG*. [Online]. Available at: doi:10.1016/j.tig.2025.07.012 [Accessed 24 October 2025].

Wen, X. et al. (1999). Kinase activation of the non-receptor tyrosine kinase Etk/BMX alone is sufficient to transactivate STAT-mediated gene expression in salivary and lung epithelial cells. *The journal of biological chemistry*, 274 (53), pp.38204–38210. [Accessed 22 August 2025].

Wenger, S. L. et al. (2004). Comparison of established cell lines at different passages by karyotype and comparative genomic hybridization. *Bioscience reports*, 24 (6), pp.631–639. [Accessed 15 September 2025].

Whitington, T. et al. (2011). Inferring transcription factor complexes from ChIP-seq data. *Nucleic acids research*, 39 (15), p.e98. [Accessed 26 October 2025].

Whitman, N. A. et al. (2019). Hypoxia differentially regulates estrogen receptor alpha in 2D and 3D culture formats. *Archives of biochemistry and biophysics*, 671, pp.8–17. [Accessed 8 July 2025].

Williams, C. G. et al. (2022). An introduction to spatial transcriptomics for biomedical research. *Genome medicine*, 14 (1), p.68. [Accessed 23 November 2025].

Wilson, B. J. and Giguère, V. (2008). Meta-analysis of human cancer microarrays reveals GATA3 is integral to the estrogen receptor alpha pathway. *Molecular cancer*, 7 (1), p.49. [Accessed 7 November 2025].

Wilson, C. A. and Dering, J. (2004). Recent translational research: microarray expression profiling of breast cancer--beyond classification and prognostic markers? *Breast cancer research: BCR*, 6 (5), pp.192–200.

Wolff, A. C. et al. (2023). Human epidermal growth factor receptor 2 testing in breast cancer: ASCO-College of American Pathologists guideline update. *Journal of clinical oncology: official journal of the American Society of Clinical Oncology*, 41 (22), pp.3867–3872. [Accessed 26 November 2025].

Wolff, A. C. and Davidson, N. E. (2001). Use of SERMs for the adjuvant therapy of early-stage breast cancer. *Annals of the New York Academy of Sciences*, 949 (1), pp.80–88. [Accessed 22 October 2025].

Woodard, L. E. and Wilson, M. H. (2015). piggyBac-ing models and new therapeutic strategies. *Trends in biotechnology*, 33 (9), pp.525–533. [Accessed 26 October 2025].

Wortham, N. C. et al. (2009). The DEAD-box protein p72 regulates ERalpha-/oestrogen-dependent transcription and cell growth, and is associated with improved survival in ERalpha-positive breast cancer. *Oncogene*, 28 (46), pp.4053–4064. [Accessed 17 September 2025].

Wu, C. H. et al. (2006). The Universal Protein Resource (UniProt): an expanding universe of protein information. *Nucleic acids research*, 34 (Database issue), pp.D187-91. [Accessed 18 November 2025].

Wu, D. et al. (2024). Bioprinting of perfusable vascularized organ models for drug development via sacrificial-free direct ink writing. *Advanced functional materials*, 34 (30), p.2314171. [Accessed 24 October 2025].

Wu, D., Wang, L. and Huang, H. (2021). Protocol to apply spike-in ChIP-seq to capture massive histone acetylation in human cells. *STAR Protocols*, 2 (3), p.100681.

Wu, S. J. et al. (2021). Single-cell CUT&Tag analysis of chromatin modifications in differentiation and tumor progression. *Nature biotechnology*, 39 (7), pp.819–824. [Accessed 26 October 2025].

Xie, G. et al. (2012). Mammosphere cells from high-passage MCF7 cell line show variable loss of tumorigenicity and radioresistance. *Cancer letters*, 316 (1), pp.53–61. [Accessed 15 September 2025].

Xie, J. et al. (2016). Hypoxia regulates stemness of breast cancer MDA-MB-231 cells. *Medical oncology* , 33 (5). [Online]. Available at: doi:10.1007/S12032-016-0755-7.

Xu, F. et al. (2023a). The role of G protein-coupled estrogen receptor (GPER) in vascular pathology and physiology. *Biomolecules*, 13 (9), p.1410. [Accessed 25 November 2025].

Xu, J. et al. (2021). To mock or not: a comprehensive comparison of mock IP and DNA input for ChIP-seq. *Nucleic acids research*, 49 (3), p.e17. [Accessed 25 October 2025].

Xu, L. et al. (2024). A comprehensive single-cell breast tumor atlas defines epithelial and immune heterogeneity and interactions predicting anti-PD-1 therapy response. *Cell reports. Medicine*, 5 (5), p.101511. [Accessed 15 November 2025].

Xu, Y. et al. (2023b). Global trends and forecasts of breast cancer incidence and deaths. *Scientific data*, 10 (1), p.334. [Accessed 20 October 2025].

Yakavets, I. et al. (2020). Advanced co-culture 3D breast cancer model for investigation of fibrosis induced by external stimuli: optimization study. *Scientific reports*, 10 (1), p.21273.

Yan, Y. et al. (2018). HIF-2α promotes conversion to a stem cell phenotype and induces chemoresistance in breast cancer cells by activating Wnt and Notch pathways. *Journal of experimental & clinical cancer research: CR*, 37 (1). [Online]. Available at: doi:10.1186/S13046-018-0925-X.

Yang, J. et al. (2015). Estrogen receptor-α directly regulates the hypoxia-inducible factor 1 pathway associated with antiestrogen response in breast cancer. *Proceedings of the National Academy of Sciences of the United States of America*, 112 (49), pp.15172–15177. [Accessed 8 July 2025].

Yang, J. et al. (2025). Dynamic culture system advances the applications of breast cancer organoids for precision medicine. *Scientific reports*, 15 (1), p.8852. [Accessed 24 October 2025].

Yang, J., Harris, A. L. and Davidoff, A. M. (2018). Hypoxia and hormone-mediated pathways converge at the histone demethylase KDM4B in cancer. *International journal of molecular sciences*, 19 (1), p.240. [Accessed 8 July 2025].

Yang, W. et al. (2023). Single-cell RNA reveals a tumorigenic microenvironment in the interface zone of human breast tumors. *Breast cancer research: BCR*, 25 (1), p.100. [Accessed 23 November 2025].

Yang, Z., He, N. and Zhou, Q. (2008). Brd4 recruits P-TEFb to chromosomes at late mitosis to promote G_1 gene expression and cell cycle progression. *Molecular and cellular biology*, 28 (3), pp.967–976. [Accessed 5 November 2025].

Yaşar, P. et al. (2017). Molecular mechanism of estrogen–estrogen receptor signaling. *Reproductive medicine and biology*, 16 (1), pp.4–20.

Yasir, M. et al. (2022). Long-read sequencing for identification of insertion sites in large transposon mutant libraries. *Scientific reports*, 12 (1), p.3546. [Accessed 24 November 2025].

Yates, M. E. et al. (2024). ESR1 fusions invoke breast cancer subtype-dependent enrichment of ligand-independent oncogenic signatures and phenotypes. *Endocrinology*, 165 (10), p.bqae111. [Accessed 23 October 2025].

Yen, A. et al. (2023). Calling Cards: a customizable platform to longitudinally record protein-DNA interactions over time in cells and tissues. *bioRxiv*, p.2023.06.07.544098. [Online]. Available at: doi:10.1101/2023.06.07.544098 [Accessed 2 May 2024].

Yersal, O. and Barutca, S. (2014). Biological subtypes of breast cancer: Prognostic and therapeutic implications. *World journal of clinical oncology*, 5 (3), pp.412–424. [Accessed 21 October 2025].

Yi, J. et al. (2020). Trefoil factor 1 (TFF1) is a potential prognostic biomarker with functional significance in breast cancers. *Biomedecine & pharmacotherapie [Biomedicine & pharmacotherapy]*, 124 (109827), p.109827.

Yi, P. et al. (2002). Differences in the abilities of estrogen receptors to integrate activation functions are critical for subtype-specific transcriptional responses. *Molecular endocrinology (Baltimore, Md.)*, 16 (8), pp.1810–1827. [Accessed 21 October 2025].

Yi, P. et al. (2013). SRC-3 coactivator regulates cell resistance to cytotoxic stress via TRAF4-mediated p53 destabilization. *Genes & development*, 27 (3), pp.274–287. [Accessed 23 October 2025].

Yin, L. et al. (2020). Triple-negative breast cancer molecular subtyping and treatment progress. *Breast cancer research: BCR*, 22 (1), p.61. [Accessed 26 November 2025].

Yoshida, J. et al. (2017). Chromatin states shape insertion profiles of the piggyBac, Tol2 and Sleeping Beauty transposons and murine leukemia virus. *Scientific Reports 2017 7:1*, 7 (1), pp.1–18.

Yu, D. et al. (2012). Regional differences in rat conjunctival ion transport activities. *American journal of physiology. Cell physiology*, 303 (7), pp.C767-80. [Accessed 21 November 2025].

Yu, F. et al. (2001). HIF-1alpha binding to VHL is regulated by stimulus-sensitive proline hydroxylation. *Proceedings of the National Academy of Sciences of the United States of America*, 98 (17), pp.9630–9635. [Accessed 23 October 2025].

Yun, Z. and Lin, Q. (2014). Hypoxia and regulation of cancer cell stemness. *Advances in experimental medicine and biology*, 772, pp.41–53. [Accessed 24 October 2025].

Yusa, K. et al. (2011). A hyperactive piggyBac transposase for mammalian applications. *Proceedings of the National Academy of Sciences of the United States of America*, 108 (4), pp.1531–1536. [Accessed 15 October 2025].

Zaborowski, A. M. et al. (2023). A systematic review of oncological outcomes after nipple-sparing mastectomy for breast cancer. *Journal of surgical oncology*, 127 (3), pp.361–368. [Accessed 22 October 2025].

Zagami, P. and Carey, L. A. (2022). Triple negative breast cancer: Pitfalls and progress. *NPJ breast cancer*, 8 (1), p.95. [Accessed 26 November 2025].

Zanoni, M. et al. (2020). Modeling neoplastic disease with spheroids and organoids. *Journal of hematology & oncology*, 13 (1), p.97. [Accessed 24 November 2025].

Zappia, L., Phipson, B. and Oshlack, A. (2018). Exploring the single-cell RNA-seq analysis landscape with the scRNA-tools database. *PLoS computational biology*, 14 (6), p.e1006245. [Accessed 25 October 2025].

Zarnegar, M. A. et al. (2017). Targeted chromatin ligation, a robust epigenetic profiling technique for small cell numbers. *Nucleic acids research*, 45 (17), pp.e153–e153.

Zhang, B. et al. (2021a). Tracking histone modifications in embryos and low-input samples using ultrasensitive STAR ChIP-seq. *Methods in molecular biology (Clifton, N.J.)*, 2214, pp.241–252. [Accessed 25 October 2025].

Zhang, C. et al. (2016). Hypoxia induces the breast cancer stem cell phenotype by HIF-dependent and ALKBH5-mediated m^6A-demethylation of NANOG mRNA. *Proceedings of the National Academy of Sciences of the United States of America*, 113 (14), pp.E2047-56. [Accessed 25 November 2025].

Zhang, C. et al. (2020). 3D culture technologies of cancer stem cells: promising ex vivo tumor models. *Journal of tissue engineering*, 11, p.2041731420933407. [Accessed 24 October 2025].

Zhang, J. et al. (2021b). Neoadjuvant chemotherapy or endocrine therapy for invasive ductal carcinoma of the breast with high hormone receptor positivity and human epidermal growth factor receptor 2 negativity. *JAMA network open*, 4 (3), p.e211785. [Accessed 22 October 2025].

Zhang, X. et al. (2013a). A renewable tissue resource of phenotypically stable, biologically and ethnically diverse, patient-derived human breast cancer xenograft models. *Cancer research*, 73 (15), pp.4885–4897. [Accessed 24 October 2025].

Zhang, X. H.-F. et al. (2013b). Metastasis dormancy in estrogen receptor-positive breast cancer. *Clinical cancer research: an official journal of the American Association for Cancer Research*, 19 (23), pp.6389–6397. [Accessed 25 November 2025].

Zhang, Y. et al. (2021c). HIF-1α is necessary for activation and tumour-promotion effect of cancer-associated fibroblasts in lung cancer. *Journal of cellular and molecular medicine*, 25 (12), pp.5457–5469. [Accessed 3 November 2025].

Zhang, Z. et al. (2015). FOXA2 attenuates the epithelial to mesenchymal transition by regulating the transcription of E-cadherin and ZEB2 in human breast cancer. *Cancer letters*, 361 (2), pp.240–250. [Accessed 24 November 2025].

Zhao, S. et al. (2016). PiggyBac transposon vectors: the tools of the human gene encoding. *Translational lung cancer research*, 5 (1), pp.120–125. [Accessed 26 October 2025].

Zhao, W. et al. (2024). ZMIZ1 enhances ERα-dependent expression of E2F2 in breast cancer. *Journal of molecular endocrinology*, 73 (1). [Online]. Available at: doi:10.1530/JME-23-0133 [Accessed 13 November 2025].

Zheng, G. X. Y. et al. (2017). Massively parallel digital transcriptional profiling of single cells. *Nature communications*, 8 (1), p.14049. [Accessed 24 November 2025].

Zheng, P. et al. (2023). Nanopore sequencing technology and its applications. *MedComm*, 4 (4), p.e316. [Accessed 24 November 2025].

Zheng, Z.-Z. et al. (2022). Super-enhancer-controlled positive feedback loop BRD4/ERα-RET-ERα promotes ERα-positive breast cancer. *Nucleic acids research*, 50 (18), pp.10230–10248. [Accessed 22 August 2025].

Zhou, D. et al. (2019). Chemotherapy modulates endocrine therapy-related resistance mutations in metastatic breast cancer. *Translational oncology*, 12 (5), pp.764–774. [Accessed 23 October 2025].

Zhou, J. et al. (2020). SRC promotes tamoxifen resistance in breast cancer via up-regulating SIRT1. *OncoTargets and therapy*, 13, pp.4635–4647. [Accessed 23 October 2025].

Zhou, W. et al. (2023). Submersion and hypoxia inhibit alveolar epithelial Na+ transport through ERK/NF-κB signaling pathway. *Respiratory research*, 24 (1), p.117. [Accessed 21 November 2025].

Zhou, Y. et al. (2014). Quantitation of DNA double-strand break resection intermediates in human cells. *Nucleic acids research*, 42 (3), p.e19. [Accessed 22 August 2025].

Zhou, Y. and Liu, X. (2020). The role of estrogen receptor beta in breast cancer. *Biomarker research*, 8, p.39. [Accessed 25 November 2025].

Zhu, B. et al. (2019). MOWChIP-seq for low-input and multiplexed profiling of genome-wide histone modifications. *Nature protocols*, 14 (12), pp.3366–3394. [Accessed 25 October 2025].

Zhu, Z., Jiang, L. and Ding, X. (2023). Advancing breast cancer heterogeneity analysis: Insights from genomics, transcriptomics and proteomics at bulk and single-cell levels. *Cancers*, 15 (16), p.4164. [Accessed 15 November 2025].

Zinn, D. A., Mehner, C. and Patel, T. (2023). Protocol for generation of multicellular spheroids through reduced gravity. *STAR protocols*, 4 (2), p.102264. [Accessed 24 October 2025].

Zwart, W. et al. (2010). The hinge region of the human estrogen receptor determines functional synergy between AF-1 and AF-2 in the quantitative response to estradiol and tamoxifen. *Journal of cell science*, 123 (Pt 8), pp.1253–1261. [Accessed 21 October 2025].

The clonal evolution of tumor cell populations: Acquired genetic lability permits stepwise selection of variant sublines and underlies tumor progres.