Chromatin Dynamics During *in vitro* Differentiation of Human Urothelium

Carl Thomas Fishwick

PhD Thesis

University of York

Department of Biology

September 2013

Abstract

Ligand-activation of the constitutively expressed nuclear receptor PPARγ in normal human urothelial (NHU) cells *in vitro* entrains a programme of transcriptional changes resulting in expression of gene and protein markers associated with *in vivo* differentiated urothelium.

It was hypothesised that, after induction of differentiation, PPARγ would translocate to the nucleus and facilitate targeted expression of differentiation-associated genes through altering chromatin constitution. To address this hypothesis, differential solubility of chromatin and nuclear-matrix bound proteins were exploited to observe changes in PPARγ localisation. In addition, label-free mass spectrometric analysis of NHU extracts was undertaken to discern if differential relative abundance of chromatin-associated proteins could be detected, and next-generation sequencing technologies were employed to assess the changes induced in the chromatin environment by sequencing RNA transcripts (RNA-seq), performing high-throughput chromosome conformation (HiC), assessing transcription factor binding via formaldehyde-assisted isolation of regulatory elements (FAIRE), and histone epigenetic markers of transcription (ChIP-seq).

Potentially novel PPARγ isoforms were observed by western blot, with little localisation alterations between differentiated and control NHU cells. Differentiation markers were downregulated after treatment with siRNA specifically targeting PPARγ2, without significant reduction in PPARγ abundance. Label-free mass spectrometry detected peptides from chromatin-associated proteins involved as having differential abundance between extracts from differentiated and control NHU cells. RNA transcriptomics revealed upregulation of novel transcription factors not previously associated with urothelial differentiation. Preliminary FAIRE results revealed the presence of regulatory elements unique to terminally differentiated cells.

This study extends the understanding of the behaviour of PPARγ in NHU differentiation, identified chromatin constituents with potential roles in differentiation and provided a rich transcriptomics resource which will be a valuable tool in assessment of the impact of transcription factor binding on local chromatin organisation during urothelial differentiation.

Table of Contents

List of Figures

NHU cells.

List of Tables

List of Tables

Supplementary Materials on Disc in Back Cover

1. Supplementary Materials for Chapter 3

1.1. siRNA Densitometry Calculations

2. Supplementary Materials for Chapter 4

2.1. CSK-NaCl

2.1.1 Protein Identifications Dish1-3 0.1-2 M NaCl CSK DDA

2.1.2 emPAI quantitation Dish1 Dish 30.1-2 M NaCl CSK

2.1.3 Intensity Based LF-MS Replicate 0.1-2 M NaCl CSK

2.1.4 Mascot Scores of Peptides from Proteins with 1 Peptide for Intensity Quant.

2.1.5. GO analysis CSK 0.1-2 M NaCl

2.2. CSK-DNase

2.2.2.1. Mascot Identifications – Peptide Data

2.2.2. DDA Identifications

2.2.1.1.2. NHUc DNase Extract

2.2.1.1.1. NHUc Control Extract

2.2.1.1.3. NHUd Control Extract

2.2.1.1.4. NHUd DNase Extract

2.2.1.2. Protein Scape Exports - Protein Data

2.2.1.2.1. CSK-DNase All Extracts Merged DDA

2.2.1.2.2. NHUc Control Extract

2.2.1.2.3. NHUc DNase Extract

2.2.1.2.4. NHUc DNase Pellet

2.2.1.2.5. NHUd Control Extract

2.2.1.2.6. NHUd DNase Extract

2.2.1.2.7. NHUd DNase Pellet

2.2.1.2.8. NHUc NHUd Matched Peptides DDA

2.2.2.Label-Free Mass Spectrometry

2.2.2.1. Manually Calculated p-values for Bruker Features

2.2.2.2. Intensity LF-MS CSK-DNase Using DDA

2.2.2.3. CSK-DNase Protein emPAI Values

2.2.2.4. NHUd SPL Bruker 2+ 3+ ions Protein IDs 1

2.2.2.5. NHUd SPL Bruker 2+ 3+ ions Protein IDs 2

3. Supplementary Materials for Chapter 5

3.1 RNA-seq

Acknowledgements

Special thanks are owed to PhD supervisors Professor Jennifer Southgate, Professor Jane Thomas-Oates and Dr Andrew Nicholls for all their help and advice during the development and execution of work carried out in this thesis.

Additional thanks are owed to the members of the Jack Birch Unit of Molecular Carcinogenesis, and the Centre of Excellence for Mass Spectrometry for their help and support with lab work and equipment maintenance, as well as staff at TGAC for processing transcriptomics samples and data.

Author's Declaration

The candidate confirms that the work submitted in this thesis is his own work, and that it has only been submitted to this institution. Appropriate credit has been given where reference is made to the work of others, or for the efforts of collaborators. Part of figure 7.1.1.1 was reproduced from a copyrighted publication under licence number 3333011300599 from Elsevier publications.

1 **Introduction**

1.1 **The Urothelium**

It was the aim of the work contained in this thesis to understand the changes in the chromatin environment which occur when differentiation is established in proliferating human urothelial cells *in vitro* through treatment with agonists targeting nuclear receptors and inhibition of the cell cycle. This brief introductory chapter outlines the advances in the culture of urothelial cells which have enabled this study to take place, and a rationale for the approaches used in studying urothelial differentiation.

In situ, urothelium is a transitional epithelium which functions as the protective urinary barrier of the bladder and the ureter (1). Superficial urothelial cells (Figure 1.1.1) express a range of cell-cell contact proteins which aid the formation of the barrier to fluid transfer (2), as well as some apical plaque proteins which are only known to be expressed by urothelium (3, 4).
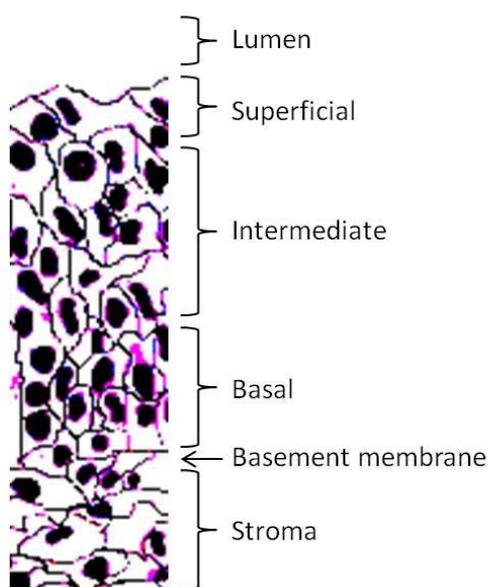
Figure 1.1.1. Illustration of cell layers in human urothelium. Superficial cells cover several intermediate cells, earning them the moniker "umbrella cells". They express a range of plaque proteins which maintain a protective barrier in the lumen of the bladder. Superficial and intermediate cells have a range of cell-cell contact proteins which help to maintain tight cell junctions, thought to aid the formation of a strong trans-epithelial barrier. Basal cells are connected to the basement membrane and are more cylindrical than the other layers.

1.2 Tissue Engineering and Cellular Differentiation

A major goal of tissue engineering is to repair or replace tissue or organs. Various approaches have been employed in pursuit of this goal, including use of synthetic or biological-derived matrix to encourage *in vivo* repair of native tissues, or *in vitro* culture of cells on a matrix under conditions designed to facilitate the cells to form functional replacement tissues or organs.

Use of matrices with and without pre-seeded cells has been successfully employed *in vivo* to encourage repair of tissues such as skin and cornea (5, 6). Implantation of tissues constructed *in vitro* has been achieved in some cases, with one example in which decellularised donor trachea was used as a scaffold for the expansion in culture of the patients' existing tracheal cells prior to successful implantation (7). For more complex organs, successes have been modest. Muscle and epithelial cells from the bladders of human myelomeningocele patients have been cultured *in vitro* on a shaped acellular matrix to produce an organ with gross tissue architecture and morphology resembling that of normal bladder. However, limited follow up after transplantation showed the engineered bladders had at best equivocal function to the diseased native bladder (8). These examples have demonstrated proof of principle that construction and transplantation of engineered tissues and organs are possible, but questions remain about whether more effort should be committed to improving and investigating the functionality of the engineered organs before transplantation into patients.

In order to understand how to culture a fully functional tissue *in vitro,* knowledge is required of the molecular cues which drive the differentiation of cells into the desired phenotype. Much research in recent years has been directed towards controlled differentiation of stem cells, which are a preferable source of cells for tissue engineering in cases where primary cells from the target organ or tissue are unsuitable for expansion. "Stem cells" is a term which describes a population of undifferentiated cells which retain the potential to differentiate into one or more cell types. The range of tissue types which the stem cell can differentiate into depends on the source. Freshly fertilised embryos are totipotent as the constituent cells can differentiated into any tissue, inner cells from the blastocyst are pluripotent as they can differentiate into many different tissue types, and progenitor populations present in tissues may be committed to a specific lineage and thus are considered unipotent. In the case of human urothelium, a resident

stem cell population is yet to be identified, although such a population is presumed to exist as candidate slow-cycling potential progenitor cells have been identified in mouse and rat (9, 10). These stem cells in bladder may not inherently be "stem" cells, as the possibility remains that the urothelial cells are being kept in a progenitor state by the niche in which they reside (11). In the absence of lineage-specific stem cells from the target tissue, stem cells from other sources can potentially be "transdifferentiated" into the required cell lineage. These can include stem cells from other tissues, embryonic stem cells cloned from the host, or induced pluripotent stem cells (iPS) (12, 13).

The most complete transformation of pluripotent progenitor cells directed to differentiate into urothelium has been reported in mouse embryonic stem cells (ESCs), which were induced to differentiate solely into urothelium by xenografting cultures of ESCs and fetal bladder mesenchyme into adult rats (14, 15). A similar study from some of the same group of authors also achieved urothelial differentiation of rat bone-marrow derived mesenchymal stem cells co-cultured with fetal bladder mesenchyme xenografted into nude mice (16). Although the instructive nature of the fetal mesenchyme is vital to direct the urothelial differentiation of the ESCs, in both studies the specific signalling mechanisms which enabled this specification remained unstudied. Some potential clues as to the molecular signals directing urothelial differentiation have been uncovered in studies where mouse ESCs were induced to express urothelial markers *in vitro* by addition of all-trans retinoic acid (17), but this method has not been successfully used to generate organised or functional tissue.

Across various tissue types, some progress has been made in directed differentiation of stem cells *in vitro* (18-20). Transdifferentiation of tissue-specific stem cells has been attempted with the use of intestinal stem cells implanted into mouse bladders, but with limited expression of urothelial markers (21). However, safety concerns about the therapeutic use of iPS and transdifferentiated cells remain due to potential existence of epigenetic differences with the native tissue progenitor cells which could affect the capacity to form functionally differentiated tissue of the desired phenotype (22-24), as well as the ability of iPS cells to form teratomas when injected into nude mice (13).

In the case of the urothelium, primary tissue samples retain significant proliferative capacity *in vitro,* making it possible to study differentiation of the tissue (25-28). Function differentiation of primary cell cultures *in vitro* has its

own inherent issues, chiefly heterogeneity between donors due to differing genetic backgrounds and often unknown environmental influences. This is exemplified by a study which showed *in vitro* differentiation of pre-adipocytes into adipocytes varies between donors, with some having a poor expression of differentiation markers (29). Although problems caused by heterogeneity of primary cell cultures in *in vitro* research into differentiation can be partly overcome by using immortalised or tumour-derived cell lines, immortalised urothelial cells have been reported to have limited capacity for differentiation, and are ultimately not suitable for transplant (30, 31).

Therefore, when investigating *in vitro* differentiation of normal primary cells, the use of replicate donor lines to test reproducibility of observed experimental outcomes is required. When interpreting the results from such primary cultures it is important to keep in mind that in addition to variation between donors, within the culture there may be a heterogeneous response to any stimuli. This type of variable response to induction of gene expression is exemplified by studies showing that even within homogeneous cell populations induced to express particular genes, at any one time a significant proportion of the population is not actively transcribing the gene (32, 33). In primary cell cultures, this heterogeneous response to stimuli likely reflects the natural phenotypic plasticity of the cells, wherein a relatively uniform progenitor population can differentiate into each layer of the tissue. Thus any induction of differentiation *in vitro* will likely not result in the complete ablation of the expression of markers which are associated with a less differentiated phenotype. Therefore any novel differentiation-associated proteins discovered through *in vitro* work with primary cultures would benefit from validation such as reproduction of *in vitro* results with multiple donors, or identification of candidate protein localisation within *in situ* differentiated cells *in vivo*.

Current models of directed *in vitro* differentiation of human urothelium are based on pharmacological activation of peroxisome-proliferator-activated receptor gamma (PPARγ), which in turn upregulates several other transcription factors, including forkhead box A1 (FOXA1), which have been shown to contribute to expression of urothelial-specific marker genes (26, 34). Other studies have associated expression of several transcription factors with the development or maintenance of the urothelial phenotype in the mouse or human embryo, including Foxa1, Foxa2 (14), grainyhead like 3 (Grhl3) (35), Kruppel-like factor 5

(KLF5) (36), PPARγ (37). This seemingly complex web of transcription factors contributing to cell identity mirrors findings in current genomics research, which has highlighted that cell-type specific gene expression is achieved by hierarchies of interacting transcription factors promoting and repressing transcription at sites along the genome, with the outcome dependent on the mix of transcription factors interacting with each locus (38-40). It is therefore desirable to further understand which urothelial differentiation-associated genes are influenced by which DNA-binding proteins in order to better understand which transcription factors markers could be used as targets to improve, or as markers to better assess, *in vitro* differentiation.

1.3 Proliferation and Differentiation in Urothelium

Cells in all layers of i*n situ* urothelium from many distinct mammalian species are mostly quiescent (41, 42), and a similar status is observed in explant cultures of human bladder (43, 44). However, cells from all layers of the urothelium have been observed to have markers of active proliferation (41, 45), and when disaggregated in culture the vast majority of human urothelial cells survive and go on to express markers of the cell cycle (25). Thus, even the superficial cells which express unique terminal differentiation markers are able to re-enter the cell cycle and proliferate, the ability to do which is thought to enable the urothelium to regenerate a functional barrier rapidly after injury. The capability to progress from a quiescent differentiated state to a proliferative phenotype and then revert back again makes the urothelium unlike other stratified epithelia, in which terminally differentiated cells are generally replenished solely by proliferation and subsequent differentiation of underlying cell layers. Understanding the plasticity of the urothelial phenotype is complex, but beginning to be elucidated.

Evidence suggests that the signalling from the stroma which drives proliferation of candidate stem cells in mouse urothelium after injury is mediated through the sonic hedgehog-wnt pathway (9). Bladder urothelium and prostate are derived from the same embryonic progenitor population, and a study has shown that the balance of expression of the nuclear receptors peroxisome proliferator activated receptor gamma (PPARγ) and androgen receptor (AR) in human prostate-derived cells was able to alter the expression of urothelial and prostate differentiation markers (46). The same study showed that when these cell cultured were explanted into mice, further phenotypic changes occurred when both PPARγ and

AR were reduced in abundance in response to knockdown of PPARγ, whereupon the cells expressed differentiation markers of neither prostate or urothelium but did express markers of a disease state known as keratinising squamous metaplasia (KSM).

The study of such animal-derived and closely related systems such as prostate can give clues to the signalling which is potentially active in urothelium, but does not necessarily translate to understanding how human urothelial tissue functions. Early experiments of *in vitro* cultures of human tissue explants showed that urothelial cells survive in culture, but that serum was necessary for the superficial cells to retain their differentiated morphology and generate a functional barrier (43). Research using *in vitro* explant models has advantages over isolated culture of cells as stromal-epithelial signalling can be maintained, but such studies of human tissue are hampered by the need for relatively large amounts of normal tissue and the difficulties associated with obtaining it. By contrast, the ability of urothelial cells to proliferate and re-differentiate has been exploited to enable development of *in vitro* culture models which circumvent the problems of studying undifferentiated cells in culture.

Normal human urothelial (NHU) cells have been shown to be able to proliferate as finite cell lines in culture (25, 47), with proliferation stimulated by an autocrine epidermal growth factor (EGF) feedback loop wherein the cells produce the EGF ligand and stimulate their own proliferation (48). Culturing NHU cells in the absence of nuclear receptor ligands by using serum-free medium prevents the pro-differentiation signalling which allows this proliferation to take place. This results in the expression of cytokeratin 14 (CK14) and a reduction in cytokeratin 13 (CK13). CK13 is expressed by basal and intermediate urothelial cells *in vivo,* and CK14 is usually only observed in urothelium in disease states such as KSM which are characterised by a switch in differentiation status (49, 50). Addition of serum to culture medium has been shown by several different research groups to be able to generate phenotypically normal urothelium with functional barriers (27, 47, 51). Although it is likely that this is achieved by restoration of nuclear receptor signalling, the complexity of factors in serum means that these studies were not able to answer the question of how the differentiation was induced.

A more reductionist approach to identifying differentiation mechanisms has enabled identification of the likely nuclear receptor signalling pathways which facilitate expression of differentiation markers in cultured normal human

urothelial (NHU) cells. Squamous metaplasia is a condition where the normal urothelium is replaced by stratified squamous epithelium, it occurs in a wide range of epithelia and is often associated with vitamin A deficiency (52, 53). Vitamin A metabolites are ligands required for function of nuclear receptors such as retinoid X receptors (RXR)(54) and retinoid A receptors (RAR) (55). Varley *et al* (56) showed that NHU cells cultured under serum-free conditions treated with the RAR ligand 13-cis-RA showed a reduction in CK14 and increase in CK13, suggesting the generation of a phenotype more akin to the *in vivo* situation. However, as normal *in vivo* superficial human urothelial cells express cytokeratin 20 (CK20), and no expression of this marker could be detected after treatment with 13-cis-RA, the authors determined that the cells had returned from an undifferentiated state to a more basal-intermediate urothelial cell type, but that terminal differentiation had not been initiated.

Further work by Varley *et al* (56) indicated that *in vivo* squamous metaplasia of urothelium was associated with loss of nuclear localisation of the nuclear hormone receptor PPARγ, which is a known RXRα heterodimerisation partner (57). Treatment of cultured NHU cells with the PPARγ agonist troglitazone (TZ) instigated the same CK14 to CK13 switch as 13-*cis*-RA, but still without the CK20 expression (26). Blocking the EGF receptor with PD153035 (PD) in combination with TZ resulted in a more reproducible initiation of the CK14-CK13 switch and the upregulation of CK20 (Table 1.1.1). Addition of epidermal growth factor receptor (EGFR) inhibitors alone could not induce this change in expression, and inhibition of PPARγ activity using antagonists prevented this change in gene expression, establishing the probability that PPARγ activation was required for *in vitro* differentiation in urothelium.

Treatment→ Cytokeratin Marker ↓	Untreated	+13-cis-RA	+TZ	+TZ +PD153035
CK20 -Terminal Differentiation	-	-	-	+
CK13 – Intermediate Differentiation	-	+	+	+
CK14 - Undifferentiated	+	-	-	-

Table 1.1.1. Cytokeratin (CK) expression as markers of differentiation in human urothelial cells. Treatment of NHU cells *in vitro* with compounds targeting RAR (13-cis-RA), PPARγ (Troglitazone) and EGFR signalling pathway (PD153035) results in upregulation (+) or downregulation of specific cytokeratins.

It was later established that although EGFR or phosphoinositide 3 kinase (PI3K) inhibition in combination with TZ treatment could effectively induce differentiation, specific inhibition of the downstream kinase Extracellular signal-regulated kinase 1 (ERK1) and ERK2 with U0126 also improved differentiation in combination with TZ. ERK1/2 are known inhibitors of PPARγ activity through modulation of a phosphoserine at position 84 of PPARγ (58-60), and it was demonstrated that EGFR inhibition reduced the relative levels of activated phosphorylated ERK1/2 and concurrently reduced the relative levels of PPARγ phosphorylated at serine 84. As the EGFR-ERK signalling pathway is central to proliferation of *in vitro* urothelial cells, these results suggested that ERK1/2 simultaneously drives proliferation and inhibits differentiation in NHU cells.

Uroplakin 2 (UPK2) is an integral component of the barrier plaques formed at the apical surface of the urothelium, and is only known to be expressed in superficial urothelial cells (3, 61). In the *in vitro* differentiation of NHU cells, there is a delay of 2-3 days before significant upregulation of UPK2. However, the promoter sequence 2000 bp upstream from the transcription start site of UPK2 does not contain any PPARγ sequence-specific binding motifs. This observation led to investigations into the possibility that other transcription factors upregulated by PPARγ were involved in the regulation of differentiation-associated genes. The resulting study established that RNA from several transcription factors, including forkhead box A1 (FOXA1) and interferon-regulatory factor (IRF1), were upregulated within 12 h of induction of differentiation. FOXA1 and IRF1 were shown to have a role to play in the expression of differentiation-associated genes

when siRNA targeting each gene reduced upregulation of CK13 after induction of differentiation (34).

These studies have reasonably well established that PPARγ is the master regulator of differentiation in urothelium. However, much remains unanswered about how the interactions of PPARγ and other transcription factors co-ordinate the change in urothelial phenotype after PPARγ is activated. New genomic and post-genomic techniques have made it possible to attempt to identify multiple chromatin-binding factors and to understand how multiple factors bind to the DNA and influence the activity of the chromatin across large genomic regions. These genome-wide studies of transcription factor binding undertaken by the encyclopaedia of DNA elements (ENCODE) consortium have shown that the interaction of myriad factors is needed to control gene expression (38). Utilising such approaches to assess changes in chromatin composition during urothelial differentiation could allow insight into the complexities of differentiation in a more comprehensive, hypothesis-free manner.

1.4 Thesis Aims

It was the aim of the work in this thesis to establish whether PPARγ expression and localisation altered during *in vitro* differentiation, and to utilise post-genomic technologies to identify novel chromatin-binding factors which alter their association with the chromatin during differentiation.

The three results chapters in this thesis are split into sections along the basis of the approaches that were used to attempt to investigate the chromatin landscape and transcription factors which control human urothelial differentiation:

- Investigations of PPARγ isoform presence and distribution during differentiation using molecular biology and indirect immunofluorescence approaches.
- Label-free mass spectrometry based investigation of changes in chromatin-associated protein abundance during differentiation.
- Utilisation of next-generation sequencing technologies to investigate epigenetic and chromatin changes occurring during of human urothelial differentiation.

2 Materials and Methods

2.1 Practical Work and Collaborations

Practical work was carried out in the Jack Birch Unit in the Department of Biology at the University of York, the Centre of Excellence in Mass Spectrometry (CoEMS) in the Department of Biology at the University of York, or at The Genome Analysis Centre (TGAC) in Norwich.

Mass spectrometric equipment in CoEMS was maintained by Dr David Ashford. TGAC staff was responsible for the final steps in the preparation of samples for next generation sequencing and the execution of bioinformatics workflows.

2.2 H$_2$O and Buffers

All references to H$_2$O in this thesis refer to double distilled H$_2$O which was further purified by reverse-osmosis in an ELGA purewater system (ELGA process water, UK) to a resistivity of 18.2 Ω.

To reduce contamination from nucleases in molecular biology experiments involving manipulation of DNA, ELGA purified water was treated with 1 mL diethylpyrocarbonate (DEPC) per 1 L water and autoclaved before adding to relevant buffers. Where certified nuclease-free water (Promega, Cat # P1193) was used this is noted.

Buffer recipes are in the appendix, section 7.4.

2.3 Tissue Culture

2.3.1 Overview

Tissue culture was carried out using aseptic technique in class II laminar air flow hoods. Surfaces were cleaned with 70% (v/v) ethanol before and after use. Keratinocyte serum-free medium (Life Technologies, Cat #10744-019) was replenished on cell monolayers every 2 to 3 days or as stated. KSFM was supplemented with provided 5 ng/mL recombinant human epidermal growth factor (EGF), 50 µg/mL bovine pituitary extract (BPE) and additional 30 ng/mL

cholera toxin (Sigma Aldrich, Cat # C8052). This "complete" KSFM (KSFMc) was used for all NHU cell cultures.

Cultures were maintained in Hera*Cell* 240 incubators (Thermo Scientific) at 37°C in a humidified atmosphere of 5% CO_2 in air. All other cell culture reagents were tissue culture grade and purchased from Sigma, unless otherwise stated. Waste cells and medium were aspirated by vacuum into a Buchner flask containing 10% (w/v) Virkon sterilising agent (SLS, Cat # CLE-1554) for decontamination. All tissue culture centrifugation steps were performed in a Sigma benchtop swing-angle centrifuge at 250 *g* for 5 min. Waste cells and medium were decontaminated in 10% (w/v) Virkon.

2.3.2 Urothelial Samples, Cell Culture and Storage

Human urological specimens of urinary bladder, ureter and renal pelvis were obtained with informed consent from patients with no previous history of urothelial neoplasia with relevant Research Ethics Committee approval. On arrival, samples were allocated a laboratory record number (Y-number, e.g. Y967).

Primary urothelial cell cultures were established from urological specimens as previously described (25). Samples were stripped of fat and connective tissue in sterile Petri dishes using scissors and forceps. The remaining sample was incubated in medium containing 0.1% (w/v) ethylenediaminetetraacetic acid (EDTA) for 4 h at 37°c to aid dissociation of the urothelium from the basement membrane. After incubation, sheets of urothelial cells were gently detached from the stroma using forceps, collected by centrifugation, resuspended in 2 mL (200 units per mL) collagenase IV (Sigma Aldrich, Cat # C5138) in Hank's balanced salt solution (HBSS) with Ca^{2+} Mg^{2+} (Life Sciences, Cat # 24020-091) and 10mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) (pH 7.6) to disaggregate cells, and incubated at 37°C for 20 min. Cells were collected by centrifugation and seeded at a density of at least 4 x10^4 cells / cm² in KSFMc medium. All NHU cells were cultured on Primaria™ 10 cm dishes, 25 cm² flasks or 75 cm² flasks (SLS, Cat # 353803, 353808 and 353810 respectively).

NHU cells were subcultured when they reached near confluence by incubating cell monolayers in PBS with 0.1% (w/v) EDTA for 5 min at 37 °C, or until cells visibly

dissociated from each other. Cells were then incubated in 0.3 - 1 mL, depending on flask size, of Hank's balanced salt solution (Life Technologies, Cat # 14170-070) containing 0.25% (w/v) trypsin (Sigma Aldrich, Cat # T4549) and 0.02% (w/v) EDTA for 1 min at 37°C. Cells were harvested into 5 mL KSFMc containing 1.5 mg/mL trypsin inhibitor (Sigma Aldrich, Cat # T6522). Cells were centrifuged, supernatant aspirated and resuspended in KSFMc before being split into fresh flasks or dishes. Cells were passaged using split ratios of 1:3 to 1:5. All experiments were performed on cells of passage 3-5.

Cells were cryopreserved in a liquid nitrogen-containing storage Dewar at -196 °C after harvesting as for passaging, and collecting by centrifugation. Cells from one T75 cm^2 flask were resuspended in 6 mL ice-cold KSFMc containing 10% (v/v) fetal bovine serum (FBS) and 10% (v/v) dimethyl sulfoxide (DMSO). 1 mL of cell suspension was aliquoted into 1 mL polypropylene cryovials and transferred to an isopropanol-filled freezing container (Sigma Aldrich, Cat # C1562) to buffer temperature change to approximately -1°C per min. Cells were kept at -80 °C overnight (not more than 24 hours), and transferred to liquid nitrogen. Cells were recovered by thawing in a 37°C water bath. Cells were then immediately diluted with 5 mL KSFMc, centrifuged and medium aspirated. Cells were resuspended in KSFMc and plated at the required density. Medium was changed 24 h after seeding to remove unattached cells.

2.3.3 Cell Counting

When cells needed to be counted, counts of single-cell suspensions were performed using an "Improved Neubauer" haemocytometer (VWR International). Cells were counted in four of the 4x4 grid and average number of cells per grid calculated. This average is then multiplied by 1×10^4 to obtain a cell count per mL of medium.

2.3.4 *In vitro* Differentiation of Normal Human Urothelial Cells

Troglitazone (TZ) (Sigma Aldrich, Cat # T2573) was solubilised in sterile filtered DMSO (SLS, Cat # D2650), to a final molarity of 100 mM. PD153035 (PD) (Calbiochem, Cat # 234490) was solubilised in DMSO, to a final molarity of 10 mM.

NHU cells were cultured to 70-80% confluence in KSFMc. For differentiation, medium was replaced with KSFMc containing TZ (1 μM) and PD (5 μM) diluted in DMSO totalling 0.1% of the medium volume. For control experiments performed without the presence of TZ and PD, KSFMc with 0.1% DMSO (v/v) was added as a vehicle control. After 24 h, medium was replenished with KSFMc with PD (5 μM) for differentiation-induced cells, or with 0.1% (v/v) DMSO for control cells. For later time-points, medium was changed as at 24 h after another 48 h (72 h total), and again after another 48 h (120 h total). All differentiation experiments were performed after four or five routine passages of 1:2 split of cells after an initial split of 1:6 from freshly isolated cells.

2.3.5 Fixation of Cultured Cells in Formaldehyde

Cells were cultured in 10 cm Primaria™ dishes (SLS, Cat # 353803) and treated with 7 mL medium containing required compounds. Dishes were placed on an orbital shaker and 189 μL of 37% (w/v) formaldehyde added drop-wise directly to the medium to a final concentration of 1%. After 10 min, cross-linking was quenched by addition of 798 μL 125 mM glycine to a final concentration of 12.5 mM, followed by incubation for 5 min. Cells were rinsed and scrape harvested in PBS and then transferred to 15 mL centrifuge tubes. Cells were pelleted by centrifugation at 800 *g* for 5 min, and the supernatants thoroughly aspirated and cell pellets stored at -80 ℃ until use.

2.3.6 Transfection of NHU cells with siRNA

NHU cells were cultured 25 cm² Primaria™ flasks to 70-80% confluence. For each dish, 30 μL of 20 μM relevant siRNA oligomers (Materials and Methods 1.7.3) were mixed with 3 mL KSFM medium without supplements and incubated for 10 minutes. 9 μL oligofectamine (Life Technologies, Cat # 12252-011) was mixed with 36 μL KSFM without supplements, then mixed with the siRNA from the previous step, and then incubated for 20 minutes. Oligofectamine forms stable complexes with oligonucleotides which the cell imports, allowing their interaction with the RNAi machinery and subsequent degradation of target mRNA.

The siRNA was diluted to 3 mL with KSFM without supplements to give a final siRNA concentration of 200 nm. Medium was aspirated from NHU cells and they

were then washed with 3 mL KSFM without supplements, before adding the siRNA. Cells were placed in the 37^oC incubator (5% O_2) for 7 h.

After incubation cells were mixed with 1.3 mL medium containing supplements at three times the normal dilution, and either TZ (3 μM) and PD (3 μM) or 0.3% DMSO for 24 h. For cells cultured for 72 h, medium was replaced with 3.3 mL KSFM containing the normal dilution of supplements and 1 μM PD in the case of TZ and PD treated cells, or 0.1% DMSO in the case of DMSO treated cells.

2.4 Western Blotting

2.4.1 Protein Harvesting

Culture medium was aspirated and cells washed twice briefly with 3-5 mL cold PBS before aspirating PBS. Cells were scrape-harvested in 50 μL 2% (w/v) SDS western blot buffer per $\sim 2 \times 10^6$ cells and transferred to a 1.5 mL micro-centrifuge tube.

Cell material was sonicated on ice for 2 X 10 s, with a 10 s rest between bursts, using a Branson Sonifier set to 25 W, 40% amplitude. Whole cell lysate was left to chill for 30 min on ice before centrifuging at 18,000 g for 30 min in a centrifuge chilled to 4^oC. Supernatant was removed to fresh micro-centrifuge tube(s) and stored at -80^oC until use.

2.4.2 Coomassie Assay

The protein content of sodium dodecyl sulfate (SDS) free samples was determined using the Coomassie protein assay reagent kit (Pierce, Cat # 23236), which is a colorimetric variant of the Bradford assay. Coomassie reagent donates an electron to proteins in the solution being measured, and then binds non-covalently the hydrophobic pockets exposed by the disruption of tertiary structure caused by the electron donation. Under acidic conditions the reagents' absorbance spectrum is shifted by this binding, allowing detection of a colorimetric change in the solution. To achieve this, protein samples were diluted 1:12.5 in H_2O, and 10 μL of each sample aliquoted in duplicate into a 96-well plate with 200 μL Coomassie reagent.

Absorbance of sample at test (570 nm) and reference (630 nm) wavelengths was measured using a MRX II 96-well plate spectrophotometer (Dynex). Samples

were measured alongside bovine serum albumin (BSA) (Pierce, Cat # 23210B) diluted to concentrations ranging from 0 to 1000 µg/mL in H_2O. The Revelation software package (Dynex) was used to plot BSA standard curve and calculate the R^2 value for the concentration range. The average of duplicate absorbance readings from each sample was used to estimate the protein concentration by comparison with the BSA standard curve.

2.4.3 BCA Assay

Where protein samples containing SDS, but without DTT, needed to be quantified, the bicinchonic acid (BCA) assay was used. The BCA assay uses the reduction of Cu^{2+} to Cu^{1+} by protein in an alkaline medium (the biuret reaction) to generate a reaction with bicinchonic acid and the Cu^+ ion, which changes its absorbance maxima after the reaction. BCA reagents were mixed with samples following the instructions in the BCA protein assay kit (Thermo Scientific, Cat # PN23227). Samples were then prepared and absorbance measured using the same approach as the Coomassie assay with the exception that the absorbance maxima measured was at 562 nm.

2.4.4 SDS-PAGE and Western Blot

Protein samples (20 µg as measured by Coomassie assay unless otherwise stated) were mixed with 4X lithium dodecyl sulfate (LDS) (Life Technologies, Cat # NP0007) and 10X reducing agent (Life Technologies, Cat # NP0004). Reagents were diluted to a final concentration of 1X with H_2O into volumes suitable to fit into wells, and then heated to 70 °C for 10 min.

Proteins were resolved by electrophoresis through 1 mm thick 10 or 15 well 4-12% Bis-Tris NuPAGE pre-cast polyacrylamide gels as required. 5 µL of All-Blue pre-stained marker (BioRad, Cat # 161-0373) was run as a protein size marker. 4-12% Bis-Tris gels were run in 1X MOPS buffer (Life Technologies, Cat # NP0001-02) which gives maximum separation of proteins around 50 kDa. 500 µL antioxidant (Life Technologies, Cat # NP0005) was added to the inner chamber of the tank during electrophoresis, which was performed at 200 V for 50 minutes or until the running buffer reached the foot of the gel.

Gels were removed from casts, the gel from around the wells and foot removed, and then rinsed in H_2O before equilibration in transfer buffer for 10 min. Polyvinyldifluoride (PVDF) sheets (Millipore, IPVH20200) were cut to size and dipped in methanol to wet, rinsed in H_2O and equilibrated in transfer buffer for 10 min. Protein was transferred to PVDF by semi-dry transfer between fibre blotting pads (Life Sciences, E-PAGE™ Blotting Pads (Reusable), 8.6 cm x 13.5 cm, Cat # LC2101) and filter paper (Whatman Grade 1, 150 mM) at 30 V for 2 h, keeping the tank on ice. PVDF membrane was rinsed in Tris buffered saline (TBS) (pH 7.4).

2.4.5 Antibody Labelling of Membranes and Scanning

Before application of antibodies, all PVDF membranes with transferred protein were blocked for 1 h in 1:1 mixture of TBS (pH 7.4) and Odyssey blocking buffer (Li-CoR, Cat # 927-40000) which is optimised to reduce autofluorescence of the membranes at detection wavelengths of the Li-CoR scanner system. For specificity control, 680 or 800 nM Alexa tagged secondary antibody (Table 2.4.1) raised against the host of the primary antibody were diluted in 1:1 mix of TBS with added 0.1% (w/v) Tween-20 (Sigma Aldrich, Cat # P9416) (TBST) and blocking buffer, and incubated with membrane for 1 h in the dark. Membranes were then washed four times for 5 min in TBST.

Antigen	Supplier	Antibody Clone or Catalogue #	Host	Production	Dilution for Labelling
PPARγ	Santa Cruz	Sc-7273 (E8)	Mouse	Monoclonal	1:500
PPARγ2	Sigma Aldrich	P0744	Rabbit	Polyclonal	1:500
FOXA1	Santa Cruz	Sc-6553 (C20)	Goat	Polyclonal	1:500
Claudin 4	Zymed	32-9400	Mouse	Monoclonal	1:1000
Cytokeratin 13	AbNova	mab 1864	Mouse	Monoclonal	1:2000
Beta Actin	Sigma Aldrich	A5441	Mouse	Monoclonal	1:10,000
Histone H3	Cell Signalling	4620	Rabbit	Monoclonal	1:2000
LAMP1	R&D Systems	AF4800	Sheep	Polyclonal	1:1000
Alexa 680 conjugated anti-Rabbit IgG	Rockland	611-131-122	Goat	Polyclonal	1:10,000
Alexa 700 conjugated anti-mouse IgG	Molecular Probes	A10038	Donkey	Polyclonal	1:10,000
Alexa 700 conjugated anti-goat IgG	Molecular Probes	A21084	Donkey	Polyclonal	1:10,000

Table 2.4.1 Antibodies used for western blotting. All antibodies were diluted in a 1:1 mixture of TBS (pH 7.4) and blocking buffer (Odyssey).

After incubation with secondary antibodies, membranes were then scanned on the Li-CoR Odyssey scanner (Odyssey CLx Scanner, LiCoR). The LiCoR uses laser excitation at 685 and 785 nm to stimulate emission from the Alexa-conjugated secondary antibodies and a scan-head with 700 and 800 nm filters to detect their respective emissions. Scans were analysed using Odyssey v1.1 software (Li-CoR).
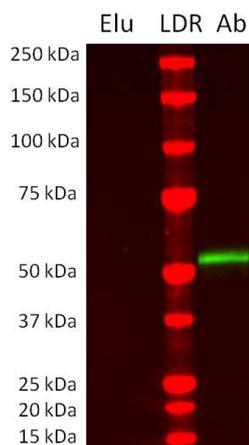
Figure 2.4.1 Example secondary only control western blot. Eluates (Elu) from a Pierce Catch and release immunoprecipitation and 1 µL of anti-PPARγ antibody (81B8 clone) (Ab) were subjected to western blot and labelled with anti-mouse (700 nm, red) and anti-rabbit secondary (680 nm, green) antibodies as described in Materials and Methods 2.4. The eluates show no reactivity to the secondary antibodies, whereas the PPARγ antibody binds to the anti-rabbit antibody (green). LDR = protein size ladder.

If no labelling was visible after secondary only control incubations, blots were used for labelling with primary antibodies, whereby membranes were mixed with primary antibody diluted in 1:1 mixture of TBST and incubated overnight at 4°C. Membranes were washed four times for 5 min each in TBST.

Band intensities were measured using the Odyssey software by drawing boxes around the protein band to calculate densitometry following background subtraction using the median intensity of pixels on the left and right of the box out to three pixels each side. Densitometry of beta-actin protein was measured on the same blot for use as a loading control to normalise protein loading in each lane.

2.4.6 Recycling western blot membranes

If membranes required removal of bound antibodies before re-use with other antibodies, these were stripped by incubating for 30 min in high pH western blot recycling kit (Source Bioscience, Cat # 90100). After stripping, secondary antibody was applied and membrane scanned as above to confirm removal of primary antibodies.

2.5 Cell Extractions and Immunoprecipitation

2.5.1 Nuclear Complex Co-IP Extracts

The Nuclear Complex Co-IP kit (Active Motif, Cat # 54001) contains hypotonic lysis buffers, chromatin fragmentation and low-salt nuclear extraction buffers specifically developed for the extraction of protein-DNA complexes. The methods in the manufacturer's manual were followed, and using the reagents supplied in the kit. Samples and centrifugation steps were maintained at 4°C or on ice unless otherwise stated.

For nuclear extract preparation, buffers and inhibitors supplied in concentrated form were diluted to 1X with H_2O fresh for each experiment. Cells were scrape-harvested into the required volume of phosphate buffered saline (PBS) (Sigma Aldrich, Cat # p4417) containing the phosphatase inhibitors provided. The cell suspension was centrifuged at 300 g for 5 min and the cell pellet resuspended in hypertonic buffer (500 µL per 8.8 X 10^6 cells) for 15 min. After incubation, 25 µL of the provided detergent was added and the suspension centrifuged at 18,000 g for 30 s. The supernatant was retained as the cytoplasmic fraction. The pellet containing nuclei was suspended in 100 µL 'Complete Digestion buffer' with added 0.5 µL 'Enzymatic Shearing Cocktail', and incubated for 90 min. 2 µL 0.5 M EDTA was added to stop the enzymatic reaction and samples centrifuged for 10 min at 18,000 g. The supernatant containing the nuclear fraction was transferred to pre-chilled micro-centrifuge tubes and the protein concentration calculated using the Coomassie protein assay (Materials and Methods 2.4.2).

2.5.2 Millipore Catch and Release Immunoprecipitation

The Millipore Catch and Release IP kit (Millipore Cat # 17-500) is supplied with a proprietary resin which reversibly binds a ligand with affinity for immunoglobulin domains of antibodies. This resin-ligand complex was mixed with antibody and 200 µg (as measured by Coomassie assay) of protein-DNA complexes which had been extracted using the Nuclear Complex Co-IP kit diluted to 500 µL in Millipore "Wash Buffer". Samples were kept at 4 °C or on ice unless otherwise stated and all centrifugation was carried out at 2000 g.

The mixture was incubated in the provided spin-filter column for 90 min with end-over-end rotation. After incubation, the non-bound protein fraction was removed by centrifugation for 30 s. Resin-antibody-protein complexes were

washed by centrifuging with 6 x 400 µL of "Wash Buffer", with 0.1 % (w/v) BSA in the first three washes to aid removal of non-specifically bound proteins. Antibody-protein complexes were eluted by addition of 75 µL of "Denaturing Elution Buffer" with freshly added 5 % (v/v) 2-mercaptoethanol before centrifugation. 25 µL of eluates and non-bound fraction was used for western blotting as described in Materials and Methods 2.4.

2.5.3 Pierce Direct IP

The Pierce Direct IP kit (Pierce, Cat # 26148) utilises sodium cyanoborohydride to form a Schiff's base between primary or secondary amines in the antibody and functionalised aldehyde compounds on the proprietary agarose resin. The agarose then retains the antibody after elution performed at or below 70°C. All wash steps were carried out by centrifugation in spin-filter columns at 1,000 *g* for 1 min, with centrifugation carried out at 4°C after immunoprecipitation.

20 µL resin slurry was added to spin-filter columns and washed twice by centrifuging through 200 µL 'Coupling Buffer' before adding the desired amount of antibody in 200 µL 'Coupling Buffer'. 5 M sodium cyanoborohydride was mixed with resin-antibody mixture and incubated for 2 hours at ambient temperature with end-over-end rotation. The reaction was quenched by mixing in 'Quenching Buffer'. The agarose was then washed by centrifuging with 6 x 200 µL 'Wash Solution' and prepared for immunoprecipitation by washing with 3 x 200 µL 'IP Wash Buffer'.

Agarose-antibody complex was mixed with 200 µg of chromatin extract obtained with Nuclear Complex Co-IP kit diluted to 500 µL in 'IP wash' buffer. The mixture was incubated for 90 min at 4°C with end-over-end rotation. Unbound proteins were removed by centrifugation and retained. Agarose-antibody-protein complexes were washed by centrifuging with 5 x 400 µL ice-cold 'IP wash buffer', and then 1 x with pH neutral 'Conditioning Buffer' before elution of the bound protein at ambient temperature in 25 µL low-pH 'Elution Buffer'. A further 50 µL 'Elution Buffer' was added to the agarose for 10 min at ambient temperature, and then the buffer collected by centrifugation. 25 µL of eluates and non-bound fraction was used for western blotting as described in Materials and Methods 2.4.

2.5.4 Antibodies used for Immunoprecipitation from Co-IP Nuclear Extracts

When using the Millipore Catch and Release system, specified amounts or dilutions of antibodies were incubated directly in 200 μg of protein-DNA extracts diluted to 500 μL. When using the Pierce Direct IP kit, the amount of antibody bound to the agarose was based on the final immunoprecipitation volume concentration. Antibodies used are detailed in Table 2.3.4.1.

As a control for immunoprecipitation of proteins which bind to immunoglobulin G (IgG) proteins, parallel precipitations were performed using anti-IgG antibodies and the eluates analysed alongside those of target antigens.

Antigen	Supplier	Antibody Clone or Catalogue #	Amount per IP	Host	Production
PPARγ	Cell Signaling	81B8	1:50	Rabbit	Monoclonal
Normal Rabbit IgG	Santa Cruz	Sc-2027	2 μg	Rabbit	Polyclonal

Table 2.5.1. Antibodies used for immunoprecipitation. Cell Signaling provide antibodies in solution without giving the concentration of the antibody, instead giving a recommended dilution at which to use for particular. As such, when using the Pierce Direct IP kit, if the immunoprecipitation was to be performed in 500 μL volume then 10 μL of antibody solution was bound to the agarose.

2.5.5 Western Blots of Immunoprecipitations

Western blots were always performed using antibodies which were not raised in the same host as the antibody used for immunoprecipitation to prevent bands from anti-host secondary antibodies appearing on membranes and confounding results (Figure 2.4.1).

2.6 Cytoskeletal (CSK) Extractions

2.6.1 Sequential CSK-NaCl extraction

Nuclear receptor proteins have been shown to have differential resistance to extraction by salt concentrations depending on their ligand-bound or transcriptional status (62). Using buffers containing detergent to perforate the cell membrane and salt of increasing concentrations to draw out the proteins across an osmotic gradient, isoforms of proteins have been shown to be differentially extracted, implying occupation of alternative functional compartments within the cell such as the nuclear matrix or chromatin (63). These buffers are known as cytoskeletal (CSK) buffers, as they were originally designed to extract all cell contents but the cytoskeleton.

CSK buffers were made to recipes described in Ainscough *et* al (64). Sodium chloride (NaCl) concentrations were variable and are described for each step. CSK buffer contained 0.1% v/v Triton-X100 (Sigma Aldrich, Cat # 93443) unless stated. For extractions, cells were cultured in 10 cm dishes were rinsed twice with ice-cold PBS and scrape harvested in 1 mL CSK buffer with 0.1 M NaCl, no detergent and freshly added 0.1 mM ATP with 1:500 protease inhibitors (Sigma Aldrich, Cat # P8340), 1 mM DTT and 1 mM PMSF. Cells were transferred to 15 mL centrifuge tubes along with remaining cell material harvested from dishes in another 3 mL of buffer. Cells were centrifuged at 4°C for 5 min at 300 *g*. Supernatant was discarded and cells re-suspended in 100 μL CSK with 0.1 M NaCl and 0.1% (v/v) Triton-X100, then incubated on ice for 5 min, then gently pipetted up and down five times using a cut-end 20-200 μL pipette tip. Cells were centrifuged at 8,000 *g* for 2 min, and supernatants retained. Incubations were then repeated with CSK buffers containing 0.1% Triton-X100 and sequentially increasing NaCl concentrations of 0.2 M, 0.4 M, 0.5 M, 1 M and 2 M. After the final incubation, the cell pellet was washed briefly without disruption in CSK with 0.1 M NaCl and no detergent. Pellet was then re-suspended in either 100 μL 2% (w/v) SDS western blot buffer or 100 μL CSK buffer with no added NaCl or detergent.

Cell pellets solubilised in SDS were then sonicated as for whole cell extracts taken for western blotting (Materials and Methods 2.4.1). Cell material in CSK buffer was pre-mixed with 4X LDS buffer, 10X reducing agent solution and H$_2$O to a final volume of 160 μL with buffers diluted to 1X. Cell pellets in CSK were sonicated as for SDS samples, but with centrifugation at ambient temperature for

10 min as chilled centrifugation was not required to remove SDS. For western blotting, protein from pellets taken in SDS were prepared as whole cell extracts with the exception that 10 μg protein rather than 20 μg protein was loaded as higher inputs appeared to affect the quality of electrophoresis, possibly due to the high salt content of the pellets. For cell material taken in CSK buffer, 10 μL of the final 160 μL volume after dilution in LDS and reducing agents was heated to 70°C for 10 minutes and loaded onto western blots.

2.6.2 CSK-DNase Extractions

Although CSK extraction using high salt can reveal which proteins which have high affinity for the nuclear compartment, this does not discern if the protein is bound primarily to the DNA or is in complex with the structural proteins of the nucleus. To shed light on whether the proteins with high resistance to extraction (>0.5 M NaCl) were bound primarily to the DNA or the structural elements of the nucleus, cells were pre-extracted with CSK buffer with detergent and 0.5 M NaCl, and then incubated with the nuclease DNaseI to digest DNA before re-extraction of solubilised DNA and protein with CSK buffer with detergent and 0.5 M NaCl. Therefore if the protein is part of a complex bound only to the DNA and not the nuclear matrix then it should be released when the DNA is digested.

CSK buffer was prepared as for sequential extractions, except that aliquots with NaCl and detergent had 0.1% Triton-X100 and 0.1 M or 0.5 M NaCl, or 0.2% Triton-X100 and 1 M NaCl.

Three dishes were prepared per time-point, one for NaCl extraction and one each for DNaseI and control extractions to be performed with and without enzyme respectively. Cells were cultured as required, then scrape-harvested and the first incubation performed in CSK with 0.1% Triton-X100 and 0.1 M NaCl, all as CSK-NaCl extractions.

After first extraction, the second incubation was performed with CSK with 0.1% Triton-X100 and 0.5 M NaCl. After harvesting supernatants, two pellets were then re-suspended in 50 μL 1X DNaseI buffer (Cambio, Cat # D9902K) with or without a 1:30 dilution of DNaseI enzyme (Cambio, Cat # D9902K) and incubated at 25°C for 30 min. Meanwhile, the remaining pellet was re-suspended in 100 μL CSK buffer with no added NaCl or detergent and prepared for western blotting as CSK-NaCl pellets. After incubation, the two other pellets were mixed 1:1 with CSK with 0.2% Triton-X100 and 1 M NaCl, and then incubated on ice for 5 min. The cell

suspension was then gently pipetted up and down 5 times with a cut-end 20-200 µL pipette tip before centrifugation and collection of supernatants. Pellets were then prepared for western blotting in CSK buffer as previous pellets. Supernatants and pellets were prepared for western blotting as for CSK-NaCl extractions.

2.6.3 Indirect Immunofluorescence Confocal Microscopy

Cells were seeded onto 12-well glass slides (CA Hendley, Cat # PH-057) in 50 µL droplets at 1×10^5 cells per mL Cells were left for 4 h to attach, and chambers of Heraeus boxes (Greiner, Cat # 96077308) flooded with 7 mL of appropriate medium. Cell were then cultured until 70-80% confluent and treated as required. To maintain structural integrity of cells after fixation to allow 3D confocal microscopy to be performed, cells were fixed in 1:1 mixture of methanol: acetone for 30 s before sequentially adding 1:1 methanol: acetone diluted with 25, 50 and 75 % PBS for 30 s each. Cells were then rinsed twice with PBS and transferred into fresh Heraeus boxes, and kept in PBS.

Immunolabelling was performed immediately after fixation. Appropriate primary antibodies (Table 2.6.1) were prepared by diluting in TBS with 0.1% w/v bovine serum albumin (BSA). Wells on slides were surrounded by liquid repellent grease (Dako Pen, Dako, Cat # S2002) to contain liquid within individual wells. Slides were tapped to remove PBS and 20 µL spots of antibody solution or TBS added to wells. In experiments where cellular differentiation had been induced, positive control antibodies targeting markers of differentiation were included to test for presence or localisation of differentiation markers.

Antibodies were incubated overnight at 4°C, and then slides were rinsed in PBS before 3 X 5 min washes in PBS on an orbital shaker. Slides were tapped to remove excess PBS and relevant secondary antibodies (Table 2.6.1) at pre-titrated dilutions were added in 25 µL drops to each well, and then incubated for 1 h at ambient temperature in the dark to protect fluorophores from UV light. Slides were rinsed and washed in PBS as before, and then incubated for 5 min on an orbital shaker with slides immersed in PBS with 0.1 µg/mL Hoechst 33258 to stain nuclear DNA. Slides were washed 1 X 5 min in PBS on an orbital shaker and then rinsed in ddH$_2$O. Slides were covered in antifade solution and mounted with glass coverslips.

Antigen	Supplier	Antibody Clone or Catalogue #	Host	Production	Dilution for Labelling
PPARγ	GlaxoSmithKline	P&A53.25	Mouse	Monoclonal	1:400
AQP3	Santa Cruz	9885	Goat	Polyclonal	1:100
Alexa 594 anti-Mouse IgG	Molecular Probes	A-11005	Donkey	Polyclonal	1:500
Alexa 488 anti-Goat IgG	Molecular Probes	A11055	Donkey	Polyclonal	1:500

Table 2.6.1 Antibodies used for confocal microscopy.

Microscopy was carried out on an LSM 710 instrument (Zeiss), fitted with appropriate lasers (488, 555 and 594 nm) for antibody fluorophore excitation and the x60 objective lens used. Although secondary antibodies were chosen to avoid cross-over of excitation and emission spectra (Alexa 488, 555 and 647 fluorophore conjugated secondary antibodies), further care was taken by optimising the instrument to avoid acquisition of emissions from wavelengths with potential for bleed-through. Each fluorophore was excited separately and individual images taken and merged as a single image. Secondary antibody only (no primary) controls were included with each experiment to ascertain if significant non-specific adsorption of secondary antibodies was occurring (Figure 2.6.1).
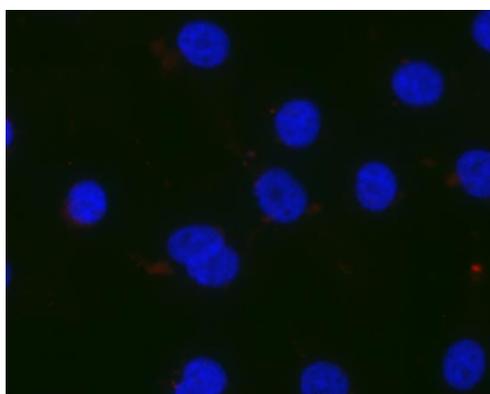
Figure 2.6.1. NHU cells extracted with CSK buffer containing 0.5 M NaCl prepared for immunofluorescence. Cell prepared as described above with only secondary antibodies (anti-mouse 594 nm and anti-goat 488 nm) applied. Cells show weak background at 594 nm, and none at 488 nm.

2.7 Genomics

2.7.1 RNA purification

Cells were cultured in 25 cm^2 flasks, and at the relevant time points were rinsed with phosphate buffered saline (PBS). 3 mL of Trizol™ solution was added to flasks and left to incubate on a rocking platform for 5 min before scrape-harvesting and transferring to a diethylpyrocarbonate (DEPC) treated 12 mL centrifuge tube. RNA was isolated in the aqueous phase following partition induced by the addition of 0.6 mL chloroform. The upper, aqueous phase was withdrawn after centrifugation at 12,000 x*g* for 30 min at 4°C. 1.5 mL isopropanol was added to the aqueous phase and the tube was incubated at ambient temperature for 10 min. RNA was precipitated by centrifugation at 12,000 x*g* for 20 min at 4°C. Isopropanol was aspirated and the RNA pellet washed by vortexing in 3 mL 75% ethanol before centrifugation at 7,500 *g* for 5 min at 4°C. Ethanol was aspirated and the pellets air-dried. Pellets were re-suspended in 900 μL nuclease-free H$_2$O, then mixed with 100 μL 3 M sodium acetate solution in 1.5 mL microcentrifuge tubes. Samples were incubated at -80°C for 1 h and centrifuged at 18,000 *g* for 30 min at 4°C. Supernatant was removed and pellets washed with 500 μL 75% ethanol. Pellets were air-dried and re-suspended in 30 μL nuclease-free H$_2$O. DNA was digested by mixing with 3.3 μL 10X DNaseI buffer and 1 μL (2 U/μL) DNaseI enzyme from Ambion DNA-free kit (Life Technologies, Cat # AM1906). Samples were incubated for 30 minutes at 37°C. 3.3 μL DNase inactivation reagent was mixed into samples before incubation at ambient temperature for 2 min. Samples were centrifuged at 8,000 *g* for 90 s to pellet the DNase inactivation reagent and the supernatant removed to a fresh tube and retained as the purified RNA fraction. 1 μL (40 U/μL) of RNAseOUT (Life Technologies, Cat # 10777-019) was mixed into samples to prevent degradation of RNA by RNases.

2.7.2 cDNA Synthesis with Random Hexamers

Complementary DNA (cDNA) was synthesised from purified mRNA by second strand synthesis by the use of reverse transcriptase and random 6-mer primers to the RNA template from the SuperScript® II First Strand Synthesis Kit (Life Technologies, Cat # 18064-014). 1 μg of DNase treated RNA was mixed with 1 μL (50 ng/μL) random hexamers and nuclease-free H$_2$O in a final volume of 10 μL. Samples were incubated at 65°C to anneal primers to RNA. To each sample the

following volumes of reagents from the kit were added: 2 μL 10X RT buffer, 4 μL 25 mM $MgCl_2$, 2 μL 0.1 M DTT and 1 μL dNTP mix. Samples were incubated at 25°C for 2 minutes. As a control for the presence of contaminating DNA, two of each sample was prepared and at this stage one sample mixed with 1 μL (50 U/μL) of Superscript II reverse transcriptase, and the other with 1 μL nuclease-free H_2O. Samples were then incubated at 25°C for 10 min followed by 50 min at 42°C to allow the reverse transcription process to occur, and then heated to 70°C for 15 min to inactivate the enzyme. Where required, 1 μL RNase H (2 U/μL) was added to samples to improve PCR efficiency over long templates by degrading the RNA: cDNA complexes. Samples were then used neat for RT-PCR reactions.

2.7.3 Polymerase Chain Reaction

PCR was performed using a T100 thermal cycler (BioRad) and the GoTaq reagent kit (Promega, Cat # M5005). PCR reactions were made to total volumes of 20 μL to the manufacturers recommended recipe, with 2 μM of sense and antisense primers, 2.5 mM $MgCl_2$, and nuclease-free H_2O. The amount of DNA included depended on the experiment being performed. DNA was denatured by heating to 95°C for 5 minutes, followed by the desired number of cycles of amplification with set to denature for 30 s at 95°C, anneal at optimal temperature

DNA was denatured for 2 minutes at 94°C followed by 35 cycles of denaturation for 30 seconds at 94°C, annealing for 30 seconds at 50-65°C (depending on optimal temperature for primer) and 1 minute/kb DNA at 72°C for extension. A final elongation phase of 10 minutes at 72°C was followed by incubation at 4°C.

2.7.4 Gel Electrophoresis

Isolated DNA was visualised under UV light after separation by gel electrophoresis. Electrophoresis grade agarose was boiled in 1x Tris-Borate-EDTA (TBE) buffer and cooled to 50°C before adding 1/10,000 (v/v) GelRed (Cambridge Bioscience, Cat # BT41003), which fluoresces under UV light when it intercalates with double stranded DNA. Gels were cast and allowed to set PCR products, diluted 1:5 (v/v) in Blue/Orange loading dye (Promega, Cat # G1881), and Hyperladder I or IV (Bioline, Cat # BIO-33029 and BIO-33025 respectively) were electrophoresed on the gel submerged in 1x TBE at 5V/cm for required amount of time to resolve bands. Gel images were captured digitally using a Gene Genius Gel Imaging System (Syngene) with GeneSnap software.

2.7.5 Primer Design and Optimisation

Primers were designed against target regions using the National Center for Biotechnology Information (NCBI) primer-BLAST tool with default settings applied and target T_m of $60^{\circ}C$ (http://www.ncbi.nlm.nih.gov/tools/primer-blast/). Primers were subjected to gradient PCR in the presence of template human genomic DNA (Roche, Cat # 11691112001) for ChIP based experiments (Figure 2.7.1), or 3C / HiC libraries for chromosome conformation capture (Figure 2.7.2).
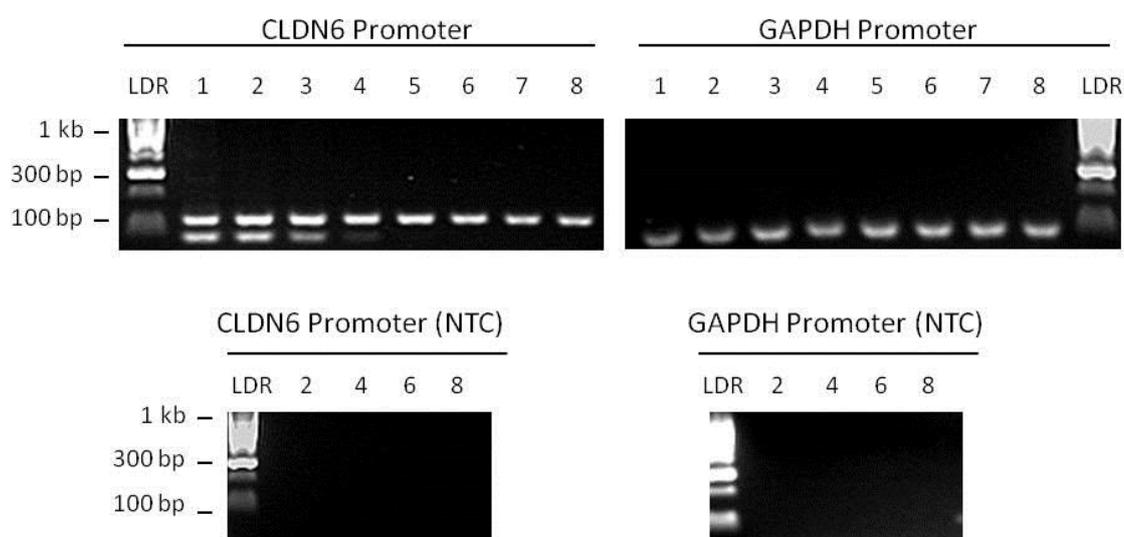
Figure 2.7.1. PCR gradients for ChIP primers. Primers targeting regions 400-600 bp upstream from the transcription start sites of CLDN6 and GAPDH were subjected to 28 cycles of PCR in the presence of 200 ng template human genomic DNA. Wells labelled 1-8 were subjected to annealing temperatures of $55^{\circ}C$, $56.4^{\circ}C$, $57.8^{\circ}C$, $59.2^{\circ}C$, $60.7^{\circ}C$, $62.2^{\circ}C$, $63.5^{\circ}C$ and $65^{\circ}C$ respectively. No template control reactions were run in parallel as negative controls. $60^{\circ}C$ was chosen as the annealing temperature for both sets of primers (sequences in Materials and Methods 2.7.5.3).
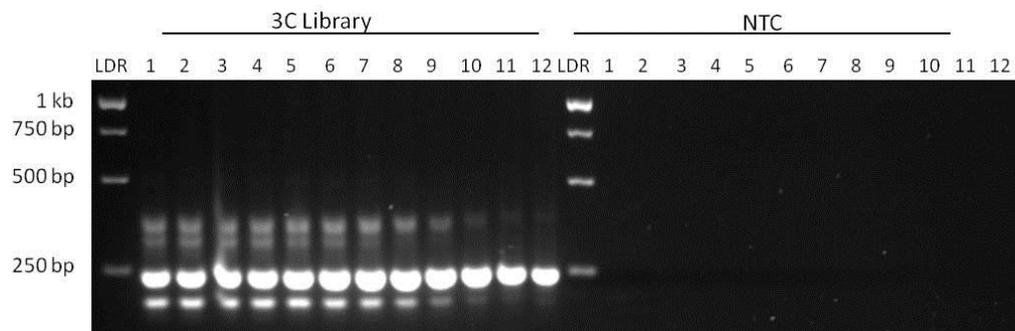
Figure 2.7.2. PCR gradient for HiC primers. Chromosome conformation capture was performed on NHU cells and 500 ng of DNA subjected to 34 cycles of PCR. Wells 1-12 were subjected to annealing temperatures of 55.1°C, 55.5°C, 56.3°C, 57.7°C, 59.4°C, 61.4°C, 63.3°C, 65.3°C, 67.6°C, 69.0°C, 69.7°C and 70.2°C respectively. 69°C was chosen as the optimal temperature for this set of primers targeting two potential interacting regions around the GAPDH locus. Primer 1: 5'-CAAGCATTCCTGGGGTGGCA-3', Primer 2: 5'-TGCAGCATCTCCTTACCCCCAGGA-3'.

2.7.6 siRNA oligomers

siRNA was purchased from MWG Eurofins. PPARγ2 siRNA sequence: 5'-AACUCUGGGGAGAUUCUCC-3', PPARγ1/2 siRNA sequence: 5'-GAAGACAUUCCAUUCACAA-3', control luciferase sequence: 5'-CGUACGCGGAAUACUUCGA-3'. siRNA sequences were checked against the human genome assembly hg19 using ensembl (www.ensemble.org). After substituting uracil for thymine, both sequences were observed to match their intended targets in exon 1 and 2 for PPARγ1 and PPARγ1/2 respectively (Figure 1.7.3). Firefly luciferase sequence is used as a control for the oligonucleotide transfection process as it has little homology to any part of the human genome, and thus should not interfere significantly with the expression of known genes (65). Although nonspecific effects have been observed in some circumstances (66), inclusion of a nonspecific control as such controls for cell systems prone to reactions to the presence of double stranded RNA. Luciferase sequence used matches that used by Sakurai *et al* (67).

Figure 2.7.3. siRNA targets exons of PPARγ. Transcript name PPARG-002, with siRNA targets underlined in red. Image from www.ensemble.org retrieved using build hg19 of the human genome.

2.7.7 RNA-seq

Whole RNA from differentiated and control NHU cells at 24 and 144 h in three donor lines was isolated for massively-parallel sequencing. RNA was quantitated and a 1 µL aliquot of 1 µg / µL RNA was electrophoresed on a Bioanalyzer (Agilent), using the RNA integrity number (RIN) interpretation of electrophoretic trace features as a guide to RNA quality (68).

2.7.8 High-throughput Chromosome Conformation Capture (HiC)

Chromosome conformation capture (3C), first described by Dekker *et al* (69) aims to probe the DNA interactions by ligating together regions of DNA which were interacting as part of the same protein-DNA complex at the time of fixation. This allows the long-range interactions of the genome to be analysed, furthering understanding of the compartmentalisation of the genome and how regions of genes are transcribed in a co-ordinated manner.

High-throughput 3C (HiC) aims to describe these interactions across the whole genome by utilising next-generation sequencing. To achieve ligation of separate but interacting DNA strands, DNA is digested with an enzyme which cuts every few kb and leaves complementary overhangs as for standard 3C. In HiC, these overhangs are filled in with nucleotides including a biotinylated cytosine and then randomly re-ligated. This is done because sequencing a standard 3C library would result in very low numbers of reads which originated from ligated DNA, as they are relatively rare within the overall DNA sequence. Incorporation of biotin enables affinity purification of the ligated DNA, thus vastly increasing the information yield per number of sequenced bases.

Protocols used here were developed based on the supplementary methods of the HiC publications by Belton *et al* (70) and Lieberman-Aiden *et al* (71).

2.7.8.1 *Cell Lysis and Chromatin Digestion with HindIII*

To construct HiC libraries, DNA was first extracted by lysing cells and then digested with the nuclease HindIII. Pellets of approximately 24×10^6 scrape-harvested, formaldehyde-fixed cells (combined pellets of 6 x 10 cm diameter dishes at 4×10^6 per 10 cm dish) from each experimental time-point stored at -80°C were thawed and re-suspended in 1 mL lysis buffer with 1:500 protease inhibitors (Sigma Aldrich, P8340).

To separate the nuclei from the cell debris, the cell lysate was incubated on ice for 15 min and then passed through a 21-gauge needle 10 times before transfer to a 1.5 mL micro-centrifuge tube and centrifuged at 2,000 *g* for 5 min. The supernatant was discarded and the pelleted cell material was washed in 500 µL ice-cold NEBuffer 2. The suspension was then pelleted for 5 min at 2,000 *g* at ambient temperature, then the supernatant discarded and the previous wash repeated. The pellet was re-suspended in NEBuffer 2 so that the final volume of pellet plus buffer was 260 µL and then split into 5 x 50 µL aliquots in 1.5 mL micro-centrifuge tubes.

To remove proteins not cross-linked to DNA, 312 µL NEBuffer 2 and 38 µL 1 % SDS was added to each tube, and then mixed carefully to avoid foaming. Samples were incubated for 10 min at 65°C for 10 min and then immediately placed on ice. 44 µL of 10% Triton-X100 was added to each tube to allow SDS and Triton-X100 to form mixed micelles and remove SDS from proteins. 400 units HindIII enzyme (NEB, Cat # R0104T) were added to each tube and samples incubated at 37°C overnight on a rocking platform.

2.7.8.2 *Fill-in of DNA Overhangs and Ligation*

To the four tubes destined for HiC, the components in Table 2.7.1 were added to catalyse the incorporation of nucleotides into the 5' overhangs left by the HindIII digestion. To enable downstream assessment of sequences present at ligation junctions which had or had not been filled in, 60 µL H_2O was added to one tube in place of other reagents so that no filling in of complementary overhangs would

occur. This procedure is the same as that used in the standard 3C protocol where the nuclease-digested DNA ends are ligated together without the incorporation of nucleotides.

End-Fill Master Mix	Supplier	Cat #	1 x Reaction
Nuclease-free H₂O	Promega	P1193	2.0 µL
10x NEBuffer 2	NEB	B7002S	6.0 µL
10 mM dATP	Life Technologies	10297-018	1.5 µL
10 mM dGTP	Life Technologies	10297-018	1.5 µL
10 mM dTTP	Life Technologies	10297-018	1.5 µL
0.4 mM biotin-14-dCTP	Life Technologies	19518-018	37.5 µL
5 U/ µL Klenow Fragment	NEB	M0212L	10.0 µL

Table 2.7.1 Components of reaction for HiC fill-in of overhangs after HindIII digestion.

Samples were incubated at 37°C for 75 min with mixing by inversion every 15 min, and then placed on ice. 96 µL 10% SDS was added to all tubes and mixed carefully to avoid foaming. Samples were incubated at 65°C for 30 min and placed on ice.

To facilitate ligation of DNA ends which were part of the same complex at the time of fixation, and not simply proximal to one another by chance in solution, samples were transferred to 15 mL centrifuge tubes and mixed with 7.58 mL ligation buffer. 50 µL of 1 U/µL T4 DNA ligase (NEB, Cat # M0203L) solution was added to HiC samples, and 10 µL added to 3C samples. More ligase was added to HiC samples as the blunt-end ligation is less efficient than that between complementary overhangs. Samples were incubated at ambient temperature overnight.

2.7.8.3 *Reversal of Cross-Links and Purification of DNA*

To remove protein and recover DNA from fixed complexes, 50 µL of 10 mg/mL proteinase K solution was added before incubation at 65°C overnight. Another 50 µL of proteinase K solution was then added to each tube and incubation continued for a further 2 h. Tubes were cooled to ambient temperature and the contents transferred to 50 mL conical tubes. DNA was extracted by twice adding two volumes of a 1:1 mixture of saturated phenol (pH 8.0):chloroform to each tube,

vortexing for 30 s and spinning at 3,000 x g for 10 min. Aqueous phases from matched HiC or 3C time-points were retained and mixed in fresh 50 mL tubes. Volumes were brought up to 40 mL with TE buffer for HiC samples and 10 mL for 3C samples. 1/10 volume of 3 M sodium acetate (pH 5.2) was added and mixed. 2.5 x volume of ice-cold ethanol was added to samples, before inverting several times to mix well. Samples were split over several 13 mL centrifuge tubes designed to withstand centrifugation >10,000 x g (Sarstedt, Cat # 55.518) and incubated at -80°C for 1 h. Tubes were spun at 12,000 x g and throughout for 20 min at 4°C. Supernatants were carefully discarded so as not to dislodge pellets. HiC pellets were-recombined into the same 2 mL of TE buffer, or 1 mL for 3C sample, and placed in a fresh 15 mL centrifuge tube. Samples were mixed 1:1 with phenol (pH 8.0): chloroform and then vortexed for 1 min and centrifuged at 2,500 g for 5 min at ambient temperature. The aqueous phase was transferred to a fresh 15 mL centrifuge tube and mixed with 1/10 volume 3 M sodium acetate (pH 5.2). 2.5x volume of ice-cold ethanol was added, and sample mixed before and transferred to 1.5 mL micro-centrifuge tubes. Samples were incubated at -80°C for 30 min, and then spun at 18,000 g for 30 min at 4°C. Supernatants were discarded and pellets air-dried. Pellets from the same time-point were combined by re-suspending in the same 500 µL TE. Each sample was loaded onto a 30 kDa spin-filter (Millipore, Cat # UFC503096) and volume reduced by centrifugation at 18,000 g for 10 min. Flow-through was discarded and 450 µL TE centrifuged through as before another three times. DNA was recovered by inverting the filter in a fresh collection tube and spinning for 2 min at 18,000 g. Volumes of HiC samples were adjusted to 100 µL, and those of 3C samples to 25 µL, using TE buffer. 2 µL RNAse A (1 mg/mL) was added to each sample, before incubation at 37°C for 30 min. 5 µL aliquots were taken and stored at -20°C. Remaining DNA was stored at -80°C.

2.7.8.4 *Library Quality Control – Library Size Distribution and PCR Digest*

DNA was quantitated by absorbance at 260 nm on a Nanodrop N-8000 spectrophotometer. 500 ng of each library was electrophoretically separated on a 0.75 % agarose gel. Libraries were expected to run as bands around 10 kb, with as little smearing towards smaller fragment sizes as possible. Smearing of DNA was compared to that in the supplementary methods of Lieberman-Aiden *et al* (71) as a benchmark.

To assess success of biotin incorporation at ligated junctions, digestion of a PCR product was performed. After HindIII digestion of the DNA, the cut-ends which are closer together are more likely to be part of the same complex and re-ligate to one another, but these may still be many kilobases apart on the genome. As ligation in HiC is done as a blunt-end ligation after filling in overhangs, the orientation of the strands can be reversed. This reversal of strand orientation at re-ligation sites allows primers to be designed against the reference genome in the sense orientation directly upstream of two HindIII recognition sequences which will only result in a PCR product in the case of a successful ligation. Successfully filled-in HiC samples will have altered sequence at the ligation junction, which should cleave in the presence of NheI enzyme as opposed to HindIII.

Primers were designed around two HindIII sites near the GAPDH locus, which is likely to have open chromatin amenable to digestion and ligation because GAPDH is constitutively expressed in all human cell types tested (72). Primers were designed using PrimerBlast with the sequence downstream from the HindIII recognition site reversed to allow primers to be designed against the intended target sequence which would be produced upon ligation of the target HindIII junctions.

500 ng DNA was added to each PCR reaction and to ensure sufficient yield of PCR product for digestion, eight reactions were performed for each time-point. 34 cycles of PCR were performed using the Go-Taq hot start polymerase (Promega, Cat # M5001) PCR kit, with the reagents detailed in Table 2.5.2.

Component	Supplier	Catalogue #	Volume
500 ng HiC / 3C DNA	N/A	N/A	As required
5X GoTaq Buffer (Green)	Promega	M5001	4 µL
10 mM dNTP mix	Life Technologies	10297-018	0.4 µL
25 mM MgCl$_2$	Promega	M5001	2 µL
Forward primer (100 mM)	MWG Eurofins	Custom	2 µL
Reverse primer (100 mM)	MWG Eurofins	Custom	2 µL
GoTaq polymerase	Promega	M5001	0.1 µL
Nuclease-free H$_2$O	Promega	P1193	Sufficient to bring final volume to 20 µL

Table 2.7.2. GoTaq PCR components.

PCR was performed on a thermal cycler (Bio-Rad T100) under the conditions outlined in Table 2.7.3.

Step	Temperature	Time
1	95°C	5 min
	go to 2	
2	95°C	30 s
3	60°C	30 s
4	72°C	45 s
	go to step 2	34 cycles
	go to step 5	
5	72°C	5 min
	go to step 6	
6	4°C	Hold

Table 2.7.3. HiC PCR protocol.

The eight PCR products from the each time point were combined and purified using a Qiagen PCR purification kit to remove reagents, then resuspended in H_2O. PCR products were incubated with HindIII, NheI, no enzyme or both enzymes to test digestion (Table 2.7.4).

	Supplier	Catalogue #	No Enzyme (µL)	HindIII (µL)	NheI (µL)	Both Enzymes (µL)
PCR product	N/A	N/A	15.0	15.0	15.0	15.0
10X NEBuffer 2	NEB	B7002	1.9	1.9	1.9	1.9
10 mg/mL BSA	NEB	B9001	0.19	0.19	0.19	0.19
HindIII (1 U / µL)	NEB	R0104	---	0.95	---	0.95
NheI (1 U / µL)	NEB	R0131	---	---	0.95	0.95
Nuclease-free H_2O	Promega	P1193	1.9	0.95	0.95	---

Table 2.7.4. Components of nuclease digestion reaction for HiC PCR products.

The digestion mixture was incubated for 2 hours and mixed with 6x loading buffer, and then electrophoresed on a 2% agarose gel until the digested and undigested bands could be resolved from one another. Images of the gel were

taken using a GeneSnap camera with a long-wave UV source lamp (Syngene). Intensity of cut and uncut bands in each digest lane were measured in the ImageJ (http://rsbweb.nih.gov/ij/) software (version 1.47) by taking the sum pixel intensity within a box of equal size drawn around all bands. Percentage efficiency of digestion was estimated by comparing the ratio of intensity from cut to uncut band. Typical efficiency reported in the literature for digestion of HiC bands by NheI is 20-30% (71), which was used as a minimum baseline for accepting libraries as successful.

2.7.8.5 *Removal of Biotin from Unligated Ends*

To remove unligated HindIII cut sites where biotin had been incorporated, exonuclease activity of Klenow polymerase was used to recede exposed DNA ends and thus remove biotin from unligated cut sites. 25 µg of HiC DNA was split into five tubes and each incubated with T4 DNA polymerase at 20°C for 4 hours with the reagents detailed in Table 2.7.5.

Reagent	Supplier	Catalogue #	Amount
Hi-C DNA sample	N/A	N/A	5 µg
10 mg/mL BSA	NEB	B9001	0.5 µL
10X NEBuffer 2	NEB	B7002	5 µL
2.5 mM dATP	Life Technologies	10297-018	0.5 µL
2.5 mM dGTP	Life Technologies	10297-018	0.5 µL
3,000 U/mL T4 DNA polymerase	NEB	M0203	5 µL
Nuclease free water	Promega	P1193	Sufficient to bring final volume to 50 µL

Table 2.7.5 Reaction components for removal of biotin from unligated ends.

After completion of the reaction, 2 µL 0.5 M EDTA solution was added to stop the reaction. Samples were pooled into matched time-points and subjected to a

phenol: chloroform purification and clean-up on 30 kDa spin-filter columns as in Materials and Methods 1.7.4, with the exception that washes on filters were performed with nuclease-free water. Samples were recovered from spin filters and volumes brought to 105 µL.

2.7.8.6 *HiC Library Sonication*

Current generation high-throughput sequencing platforms are designed to process many relatively short read lengths (25-100 bp) in parallel. To demonstrate DNA interactions in HiC data, a sequence read must encounter the ligation junction. To make the 10 kb HiC libraries amenable to next-generation sequencing, samples were sheared to 150 bp on a Covaris S2 (Covaris) instrument with the parameters detailed in Table 2.7.6. After three cycles of sonication, the size-distribution of libraries was checked by electrophoretically separating 2 µL of the sample on a 2% agarose gel.

Duty Cycle	10%
Intensity	5
Cycles per Burst	200
Set Mode	Frequency sweeping
Continuous degassing	Yes
Process time:	60 s
Number of cycles	$3 - 4$

Table 2.7.6 Covaris S2 settings for HiC library sonication.

2.7.8.7 *Size Fractionation of Libraries*

Ideal fragment length for sequencing HiC samples is 100-200 bp. Ampure XP beads (Beckman Coulter, Cat # A63880) size-selectively precipitate DNA depending on the ratio of the liquid phase to beads in the mixture. Fragments >300 bp were selectively removed by mixing a low ratio of beads into the fragmented library and recovering the beads. Remaining DNA not bound to the beads was precipitated from the supernatant by the addition of a higher ratio of beads.

Sonicated libraries were diluted to 500 µL with buffer EB (Qiagen, Cat # 19086), mixed with 450 µL of Ampure XP bead mixture and incubated at ambient temperature for 10 min. Beads were recovered on a magnetic particle separator,

and the supernatant with the <300 bp fragments was removed to a new micro-centrifuge tube. The beads with the >300 bp fragments were retained.

In a fresh tube, 500 µL of Ampure XP beads were collected on a magnetic particle separator, the supernatant was removed and the beads re-suspended in a fresh aliquot of 100 µL Ampure XP bead mixture. This concentrated bead mixture was mixed with the supernatant from the initial HiC >300 bp fragment removal, thus creating a high ratio of beads to aid recovery of small DNA fragments >100 bp. This mixture was incubated for 10 min before recovering the beads on a magnetic particle separator and removing the supernatant.

Both sets of beads with >300 bp and 100-300 bp fragments were washed twice with 1 mL of 70% ethanol, recovering beads on the magnet between each wash. Beads were then air-dried until ethanol was completely evaporated. 150 µL buffer EB was added to beads to elute the DNA, beads recovered and supernatant removed to a fresh micro-centrifuge tube. Supernatants were concentrated by centrifuging for 2 min at 18,000 x *g* in tubes with built-in 30 kDa filter membranes. Membrane tubes with DNA concentrate were inverted into fresh collection tubes and spun as before to recover DNA. The volume of samples was brought to 52 µL with nuclease-free water.

The size of the DNA fragments was checked on a Bioanalyzer (Agilent) system. Firstly, samples were quantitated on the QuBit Fluorometer (Life Technologies) and then diluted to 1 ng/µL. Size distribution of DNA in samples was analysed using electrophoresis separation on the Bioanalyzer, using a high sensitivity kit (Agilent, Cat # 5067-4626). Fragmented, size selected libraries were expected to have peak DNA content centred on 200 bp. Where libraries had such a distribution, processing was continued.

2.7.8.8 ***End Repair A-tailing***

To repair the DNA ends damaged by sonication, end-repair was performed by recessing DNA ends to blunt ends with 5´-phosphates and 3´-hydroxyl groups to enable downstream annealing of sequencing primers. The end repair reaction components used are detailed in Table 2.7.7.

Component	Supplier	Catalogue #	Volume
HiC Library (100-300 bp)	N/A	N/A	50 µL
10X repair buffer	NEB	E6050	6 µL
End repair mix (NEB)	NEB	E6050	3.5 µL
Nuclease-free water	NEB	B1052	0.5 µL

Table 2.7.7. A-tailing reaction.

The reaction was incubated for 30 min at 20°C in a thermal cycler. Samples were purified using a MinElute PCR purification kit (Qiagen, Cat # 28004), using multiple columns if libraries had more than the 5 µg DNA capacity of min-elute columns. To improve recovery from columns, samples were eluted with two double-step elutions. The first elution was performed using 20 µL TLE (10 mM Tris (pH 8.0), 0.1 mM EDTA) at 6,000 *g* for 1 min, then the eluate re-loaded and the column centrifuged at 18,000 *g* for 1 min. The second elution was a repeat of the first, with fresh 15 µL TLE. Eluates were then combined.

To enable ligation of sequencing primers and prevent concatenation of fragments during ligation, dAMP was incorporated onto the 3' of the blunt end of DNA, known as "A-tailing". A-tailing was carried out using the reaction detailed in Table 2.7.8.

Component	Supplier	Catalogue #	Volume
HiC Library	N/A	N/A	35 µL
10X NEBnext A-tailing	NEB	E6053	5 µL
Nuclease free water	NEB	B1052	6 µL
Klenow fragment (5U/ µL)	NEB	M0212	4 µL

Table 2.7.8. A-tailing reagents

The reaction was incubated in a thermal cycler for 30 min at 37°C then at 65°C for 20 min to inactivate Klenow fragment. Samples were then combined and diluted to 400 µL with TLE.

2.7.8.9 ***Biotin Pull Down***

To further increase the proportion of reads which contain target ligation junctions, streptavidin purification of biotinylated DNA was performed. 2 µL of magnetic streptavidin C1 beads (Life Technologies, Cat # 65001) per µg of DNA were aliquoted into a 1.5 mL Lo-Bind microcentrifuge tube (Eppendorf, Cat # 0030108116). 400 µL Tween wash buffer (TWB) was added to the beads, then the beads reclaimed on a magnet. All subsequent library preparation was carried out in Lo-Bind microfuge tubes to reduce adsorption of DNA onto tubes.

Supernatant was removed and the TWB wash repeated. Supernatant was removed and beads suspended in 400 µL 2X binding buffer (BB) (10 mM Tris-HCl (pH 8.0), 1 mM EDTA and 2 M NaCl). Beads were then mixed with the HiC library. The sample was incubated for 15 min with rotation at ambient temperature. Beads were collected on a magnet, and supernatant discarded. Beads were resuspended in 400 µL 1X BB and transferred to a fresh tube. Beads were collected against a magnet, and the supernatant retained. Beads were resuspended in 100 µL 1X T4 DNA ligation buffer (Life Technologies, Cat # 15224017) and again transferred to a new tube. Beads were reclaimed against the magnet, and supernatant removed before moving immediately to adapter ligation.

2.7.8.10 ***Illumina adapter ligation and paired end PCR***

To amplify the captured library fragments, the sequencing primers were ligated prior to amplification by PCR. Streptavidin beads with HiC library beads were resuspended in T4 DNA ligase mix and sequencing adapters as detailed in Table 2.7.9. Samples were to be duplexed in the sequencer, so barcoded adapters were used to allow downstream identification of libraries.

Component	Supplier	Catalogue #	Volume
HiC library beads	N/A	N/A	N/A
5X T4 ligation buffer	Life Technologies	15224017	40 µL
T4 ligase 1 U / µL	Life Technologies	15224017	20 µL
TruSeq adapter	Illumina	FC-121-2001	1 µL
Nuclease free water	NEB	B1052	139 µL

Table 2.7.9. Illumina paired-end sequencing adapters.

Libraries were incubated overnight with rotation at ambient temperature. Beads were then washed twice with 400 µL TWB and incubation for 5 min, reclaiming the beads on the magnet after each wash. Beads were resuspended in 200 µL 1X BB and transferred to a new tube. Beads were reclaimed and washed twice in 200 µL NEBuffer 2, transferring to a new tube between washes. Beads were reclaimed on a magnet and resuspended in 20 µL NEBuffer2.

Two PCR reactions were set up for each library with the components in Table 2.7.10.

Component	Supplier	Catalogue #	Volume
TruSeq master mix	Illumina	FC-121-2001	10 μL
Primer cocktail	Illumina	FC-121-2001	2 μL
Resuspension buffer	Beckman Coulter	A63880	6.5 μL
Bead-bound Hi-C DNA	N/A	N/A	1.5 μL

Table 2.7.10. HiC PCR components

Insufficient amplification results in too little material for sequencing, and excessive amplification can result in formation of PCR artefacts and introduce biases into the content of the library. To optimise the number of amplification cycles, one PCR was run for 12 cycles and one for 18 cycles on a thermal cycler using the programme outlined in Table 2.7.11.

Step	Temperature	Time
1	98°C	30 s
2	98°C	10 s
3	65°C	30 s
4	72°C	30 s
5	Go to 2	12 or 18 cycles
6	72°C	7 min

Table 2.7.11 HiC library amplification PCR protocol

Amplified libraries were purified using a Qiagen PCR purification kit. DNA quantity and size-distribution was assessed by QuBit and Bioanalyzer as in Materials and Methods 1.7.8. If libraries had a total of >20 ng of DNA after 12 cycles they were accepted for sequencing. If DNA was insufficient at 12 or 18 cycles, PCR was repeated with extra template in the mix (up to 4 μL template solution). If total DNA was <10 ng after using increased template volume in PCR, libraries were rejected.

2.7.8.11 *Quality Control – Library Digestion*

Biotinylated ligation junctions should contain NheI digestion sites, as demonstrated in the PCR product digest. Although libraries may contain some NheI sites by chance, amplified libraries should contain significant amounts of NheI sites. Belton *et al* (70) demonstrated that libraries which degraded significantly in the presence of NheI had high proportions of ligation junctions and gave good results in sequencing. 10 ng of library solution was incubated with NheI enzyme as outlined in Table 2.7.12.

Component	Supplier	Catalogue #	Volume
10X NEBuffer 2	NEB	B7002	1.5 µL
Nuclease Free Water	NEB	B1052	to 15 µL
NheI (1 U / µL)	NEB	R0131	1 µL
BSA 10X	NEB	B9001	1.5 µL
Hi-C Library	N/A	N/A	10 ng

Table 2.7.12. HiC library digestion.

Samples were incubated overnight on a thermal cycler at 37°C, with the lid heated to 105°C to prevent condensation of the sample. Samples were purified using a Qiagen MinElute kit, and eluted in 10 µL TE. Size-distribution of 1 µL aliquots of library solutions was assessed on the Bioanalyzer system. If the peak at 350 bp had significantly degraded and the distribution shifted toward smaller fragments, as compared to the assessment done after amplification, this was taken as demonstration that sequencing was likely to produce satisfactory results.

2.7.9 ChIP-PCR, ChIP-Seq and FAIRE

2.7.9.1 *Fragmentation of DNA by Sonication*

Each 1 x 10 cm dish of formaldehyde-fixed cells were resuspended in 1 mL of cold "swelling buffer with freshly added 0.5% NP-40 and 1:500 protease inhibitors. Samples were incubated on ice for 20 min with gentle shaking on an orbital shaker. Samples were then centrifuged at 800 *g* for 5 min at 4°C. Pellet was then resuspended in modified Tris-EDTA (TE) buffer.

Samples were suspended in a Bioruptor pre-chilled to 4°C (Diagenode). For initial optimization, samples were sonicated for 15 cycles of 5 min sonication at full

power, with 30 second on / off cycles. After each 5 min cycle, ice was replenished in the sonicator bath, a 10 μL aliquot was taken from the sample, and buffer volume topped up with modified TSE.

Aliquots were topped up to 250 μL with modified TE buffer, and then incubated at 65°C overnight to reverse cross-links. Samples were cooled to ambient temperature and 240 μL modified TE buffer and 10 μL of 10 mg/mL proteinase K was added to each aliquot before incubation for 2 hours at 37°C.

Fragmented DNA was purified by phenol: chloroform phase separation. To each sample, 500 μL of a 1:1 mixture of phenol: chloroform was added before vortexing for 30 s. Samples were then centrifuged for 15 min at 14,000 *g* for 15 min at ambient temperature. The transparent upper phase was removed to a fresh 1.5 mL centrifuge tube. To this tube, 1 mL of ice-cold ethanol was added before incubating the sample at 80 °C for 30 min. Samples were then centrifuged at 18,000 *g* for 30 min at 4°C. The supernatant was removed and the pellet washed with 1 mL 70% ethanol. Samples were then centrifuged at 18,000 *g* for 15 min at 4°C. The ethanol was removed and the pellets air-dried. Pellets were resuspended in 25 μL modified TE buffer. To each sample, 1 μL of 1 U / μL RNase A solution was added, and samples incubated at 37°C for 15 min.

DNA in each sample was quantitated by measurement of fluorescence at 260 nm using a Nanodrop (Thermo Scientific) spectrophotometer. 500 ng of DNA was electrophoretically separated on a 0.75% agarose gel. The number of cycles where the DNA was fragmented to 100-300 bp was chosen as the optimum number of cycles for sonication (9 cycles).

For all sonication performed, aliquots were always taken after the final sonication cycle and subjected to purification, and fragmentation of DNA assessed by electrophoresis through agarose gels as above.

2.7.9.2 *ChIP*

Magnetic protein-G conjugated Dynabeads (Life Sciences, Cat # 100-04D) were used for immunoprecipitations. Before beads were used for immunoprecipitations (IP), beads were pre-blocked to reduce non-specific binding. To achieve this, every 100 μL of bead suspension was mixed with 1 mL of buffer to be used for IP. Beads were recovered on a magnet and mixed with another 1 mL of IP buffer a total of three times. Beads were then mixed with 1 mL blocking buffer (0.2 mg/mL

glycogen, 0.2 mg/mL BSA and 0.2 mg/mL yeast tRNA in IP buffer) and incubated with rotation overnight at 4°C. Beads were recovered and washed twice as before with IP buffer before diluting to starting volume in IP buffer.

Sonicated chromatin from 1 x 10 cm dish was buffer-exchanged into radioimmunoprecipitation assay (RIPA) buffer for immunoprecipitation by loading onto an Amicon 30 kDa filter unit (Millipore, Cat # UFC503096). Sample was concentrated by centrifugation at 18,000 *g* for 10 min. 200 μL RIPA buffer was added to the unit, and centrifugation repeated. This process was repeated twice more and samples recovered into a new tube and diluted to 200 μL with RIPA buffer. A 10 μL aliquot of chromatin was kept as an "input" control for PCR reactions.

Chromatin solution was diluted to 2 mL with radioimmunoprecipitation assay buffer (RIPA) buffer and 40 μL of blocked Dynabeads added to pre-clear chromatin which binds non-specifically to Dynabeads. Chromatin was incubated for 90 min with rotation at 4°C before recovery of beads. Chromatin supernatant was aspirated from the beads and then split into 500 μL aliquots before addition of appropriate antibodies against target or control IgG (Table 2.7.13). The chromatin-antibody mixture was incubated overnight at 4°C with rotation. 20 μL blocked Dynabead suspension were added to the chromatin-antibody mixture, and incubated at 4°C with rotation for a further 90 min.

Antigen	Supplier	Antibody Clone or Catalogue #	Host	Production	Dilution
Histone H3	Cell Signalling	4620	Rabbit	Monoclonal	1:2000
Histone H3K4me3	Cell Signalling	9733	Rabbit	Monoclonal	1:2000
Histone H3K27me3	Cell Signalling	9751	Rabbit	Monoclonal	1:2000
Rabbit IgG	Rockland	611-131-122	Goat	Polyclonal	1:10,000

Table 2.7.13 Antibodies used for ChIP.

Beads were recovered and the supernatant retained as the non-bound fraction. Beads were then washed three times by resuspending in 500 μL RIPA buffer and recovery on the magnet, and transferred to a fresh tube. Beads were then washed

another three times in RIPA buffer, with a final wash in TE buffer with 50 mM NaCl. Supernatant was removed and beads resuspended in elution buffer (1 % w/v SDS, 10 mM EDTA, 50 mM Tris-HCl pH 8). Samples were then heated to 65°C for 30 min. Beads were retained and the supernatant removed to a fresh tube. 1 µL of 1 U / µL RNase A solution was added to samples, which were then mixed and incubated at 37°C for 15 min. DNA was purified from samples using a Qiagen PCR purification kit according to the manufacturer's instructions, with elution in 20 µL TE buffer. Samples were stored at -80°C until use.

2.7.9.3 *ChIP-PCR*

Chromatin immunoprecipitation (ChIP) success was ascertained by ChIP-PCR. Targets for ChIP were histone H3 trimethylated at lysine 4 (H3K4me3) or lysine 27 (H3K27me3), which are respectively markers for transcriptionally active and repressed genomic regions across multiple human cell types. Although the specific genomic sites with these histone markers varies depending on transcriptional activity of the surrounding genes, where they are present the abundance is highest upstream of transcription start sites (TSS) with a peak around 500 bp upstream. Two candidate genes were chosen based on expected transcription levels: GAPDH as a target for H3K4me3, as it ubiquitously expressed in most human tissues (72); claudin 6 for H3K27me3, as it has been demonstrated not to be expressed in NHU cells under the differentiated and control conditions used in this study (2).

The Ensemble genome browser (www.ensemble.org) was used to identify the transcription start sites for CLDN6 (16:3064713:-1) and GAPDH (12:6643093:1), and export 1 kb of 5' flanking nucleotide sequences. This sequence was input into primer design software (http://www.ncbi.nlm.nih.gov/tools/primer-blast/). The software was instructed to design a PCR product of 100-200 bp within 250 bp of the 500th base in the sequence (Table 2.7.14). The search target was the human reference genome (hg19) and all other settings were left as default. The highest ranked primer pair with no other hits in the genome was chosen.

CLDN6

AGATGGCACCACTGCACTCTAGCCTGGGCCACAGAGCGAGACTCCTCGTCTCAAACAAA
CAAACAAAACAAATGGACAAGTTTTGAGCCTTAAACTCAGGGCTAGAATTGAACCATTC
TACACGGAAACTGCTTGGAAATCGCACCGTCCTGTCATCCTGCTT**TTTGCCGCCTTGC
AGCCTGA**TCACTGGCTCTGGCCTTGGACTGTCATTGTATAGGTTTCATCCTGTTATCTT
TCATTTGCATTCAGG**ATAGCCAGGGCAGGGCTGCT**TGCACAGTCACTGTCATCAGCC
AGTGCCTGTCACATAACTAGAAAGAATAATAAATGCTAAAGCATTAAGTGTCCATTTTGC
CCACGAGAAGTGAGGCTCAGGGAGCATGTCTTGACTGAGTTCCAGGATGTGATGACCAC
TTTCCAGGGTGCATGCCCCTAGAGTCCTGTCCACATGTGGCCTGAGGTCTGGGCCTGCC
AGGCCTCACCTTGTTGTGCTTCTGTCCCAAACACAGTGCAGCTCCTTCAACCTCGCC***ATG***
GCCTCTG

GAPDH

AGGAGGGACTTAGAG**AAGGGGTGGGCTTGCCCTGT**CCAGTTAATTTCTGACCTTTAC
TCCTGCCCTTTGAGTTTGATGATGCTGAGTGTACAAGCGTTTTCTCCCTAAAGGGTGCA
GCTGAGCTAGGCAGCA**GCAAGCATTCCTGGGGTGGCA**TAGTGGGGTGGTGAATACC
ATGTACAAAGCTTGTGCCCAGACTGTGGGTGGCAGTGCCCCACATGGCCGCTTCTCCTG
GAAGGGCTTCGTATGACTGGGGGTGTTGGGCAGCCCTGGAGCCTTCAGTTGCAGCCAT
GCCTTAAGCCAGGCCAGCCTGGCAGGGAAGCTCAAGGGAGATAAAATTCAACCTCTTGG
GCCCTCCTGGGGGTAAGGAGATGCTGCATTCGCCCTCTTAATGGGGAGGTGGCCTAGG
GCTGCTCACATATTCTGGAGGAGCCTCCCCTCCTCATGCCTTCTTGCCTCTTGTCTCTTA
GATTTGGTCGTATTGGGCGCCTGGTCACCAGGGCTGCTTTTAACTCTGGTAAAGTGGAT
ATTGTTGCCATCAATGACCCCTTCATTGACCTCAACTAC***ATG***

Target Gene Promoter	Sense Primer 5'→ 3'	Antisense Primer 5'→ 3'
CLDN6	TTTGCCGCCTTGCAGCCTGA	AGCAGCCCTGCCCTGGCTAT
GAPDH	AAGGGGTGGGCTTGCCCTGT	TGCCACCCCAGGAATGCTTGC

Table 2.7.14 ChIP-PCR Primer Design. Partial sequence of the CLDN6 and GAPDH promoter regions shown for illustrative purposes. Primer targets are underlined in bold text, and ATG TSS codon of the gene is in bold italicised text. CLDN6 primers target a region 277-364 bp upstream from the TSS, and the GAPDH primers targeted a regions 412-549 bp upstream from the TSS.

1 µL of purified ChIP-DNA solution and chromatin IP "input" were prepared for PCR using the GoTaq (Promega) PCR kit with the reagents outlined in Table 2.7.15.

Reagent	Volume (µL)
GoTaq 5X Flexi Buffer	4
MgCl$_2$ 25 mM	2
Pre-mixed nucleotides (10 mM each)	0.4
Sense primer (10 mM)	2
Antisense primer (10 mM)	2
H$_2$O	8.6
DNA	1
Total Volume	**20 µL**

Table 2.7.15. PCR reagents for ChIP-PCR

PCR was then performed using the conditions outlined in Table 2.7.16.

Step	Temperature	Time
1	98 °C	30 s
2	98°C	10 s
3	60°C	30 s
4	72°C	30 s
5	Go to 2	12 or 18 cycles
6	72°C	7 min

Table 2.7.16. ChIP-PCR amplification PCR settings.

10 µL of PCR reactions were separated by electrophoresis on a 2 % agarose gel and gels photographed under UV light to visualise DNA bands.

2.7.10   Next Generation Sequencing

Once samples were generated in a format ready for sequencing, they were submitted to TGAC (Norwich, UK).

RNA and FAIRE samples were submitted as unprocessed purified samples, then fragmented and sequencing primers adapted by TGAC. ChIP-seq libraries were submitted after primers were adapted and final amplification tested using the same approach that was used for the HiC libraries.

All sample libraries were tagged with Illumina barcoded sequencing primers to allow multiplexing where necessary. Samples were all quantitated by QuBit and average fragment lengths calculated using the Bioanalyzer data. Samples were then diluted to the same molarity and small aliquots compared for amplification characteristics by qPCR. Samples which amplified successfully were considered as having passed the quality control criteria.

2.7.10.1 *RNA-Seq Library Construction*

To prepare RNA samples for sequencing, the submitted RNA was adapted to sequencing primers using the Illumina TruSeq RNA kit (Illumina, Cat # RS-122-2201), following the manufacturer's instructions. In brief, 1 µg of total RNA was purified and enriched for mRNA using provided poly-T oligonucleotide attached magnetic beads. Two rounds of purification were used. During the second elution of poly-A RNA, the RNA was fragmented and primed for cDNA synthesis. cDNA synthesis was carried out using SuperScript II Reverse Transcriptase (Invitrogen, Cat # 18064022) and random primers. Second strand cDNA synthesis was carried out and the DNA was subjected to end repair, "A" tailing and ligation. cDNA templates were enriched by 15 cycles of PCR as per manufacturer's instructions. The amplified library was quantified using a Bioanalyzer DNA 100 Chip. The library was normalised to 10 nM for generation of sequence clusters on a sequencing flow-cell on the Illumina c-Bot instrument.

Sequencing library cluster generation was carried out on a flow cell on the Illumina cBot according to the manufacturer's instructions. Following the clustering procedure 100 cycles of single end sequencing was performed with TruSeq Rapid SBS sequencing chemistry, HCS 2.0.10 software and RTA 1.17 on the HiSeq platform. FASTQ files were generated and demultiplexed according to library-specific indices by CASAVA (Illumina).

Libraries were multiplexed 6 per lane, yielding a minimum of 17 million pairs of reads per sample.

2.7.10.2 *ChIP-Seq Library Construction*

To prepare DNA captured by ChIP for sequencing, the TruSeq ChIP sample preparation kit (Illumina, Cat # IP-202-1012) was used by TGAC following the

manufacturer's recommended protocol. In brief, 5–10 ng ChIP DNA was blunt-ended and phosphorylated, and a single 'A' nucleotide added to the 3' ends of the fragments in preparation for ligation to an adapter with a single-base 'T' overhang. The ligation products were purified and accurately size-selected by agarose gel electrophoresis. Size-selected DNA was then purified and PCR-amplified to enrich for fragments with adapters on both ends. The final purified product was then quantitated using a combination of Bioanalyzer DNA HS Chip (Agilent) on a 2100 Bioanalyzer and Qubit 2.0 (Invitrogen). The libraries were normalised to 10nM, pooled and q-PCR was performed prior to cluster generation.

Samples were sequenced on an Illumina HiSeq 2500 (Rapid-Run mode) clustering and sequencing using 50 bp single-end reads. The 8 libraries were normalised and equimolar pooled to 10 nM. The library pool was then diluted to 2 nM with NaOH and 4.5µL transferred into 995.5 µL HT1 to give a final concentration of 9pM. 135 µL of the diluted library pool was then transferred into a 200 µL strip tube and placed on ice before loading onto the Illumina cBot utilising the Rapid Duo cBot sample loading kit (Illumina, Cat # CT-402-4001) using the RR_TemplateHyb_FirstExt_vR recipe to allow the pool to be run on a single lane of a HiSeq 2500 Rapid flow cell. The rapid flow cell was loaded onto the Illumina HiSeq2500 instrument following the manufacturer's instructions and onboard clustering was performed using the TruSeq Rapid Paired-end cluster generation kit. Following the clustering procedure 50 cycles of single end sequencing was performed with TruSeq Rapid SBS sequencing chemistry, HCS 2.0.10 and RTA 1.17 softwares.

Libraries were multiplexed 8 per lane, yielding a minimum of 13 million pairs of reads per sample.

2.7.10.3 *FAIRE Library Construction*

DNA obtained using the FAIRE protocol was prepared for sequencing as for ChIP-seq. Samples were multiplexed four per 100 bp paired-end HiSeq lane, yielding a minimum of 55 million pairs of reads per sample.

2.7.11 Bioinformatics

2.7.11.1 *RNA-seq Mapping*

To allow sequenced RNA to be mapped to the genome, reads were aligned using tophat2 version 2.0.8 (http://tophat.cbcb.umd.edu/) using default parameters (73). Alignment was performed to (74)both the Ensembl GRCh37 and UCSC hg19 annotation. The genome sequence indexes were downloaded from the TopHat website (http://tophat.cbcb.umd.edu/igenomes.shtml). TopHat aligns RNA-Seq reads using the bowtie short read aligner (Bowtie2 version 2.0.6 http://bowtie-bio.sourceforge.net/index.shtml) and then analyzes the mapping results to identify splice junctions between exons.

2.7.11.2 *ChIP-seq Mapping*

Sequence reads from ChIP DNA were aligned to the genome using bowtie2 version 2.0.6 using default parameters against the UCSC hg19 genome downloaded from iGenomes (ftp://igenome:G3nom3s4u@ussd-ftp.illumina.com/Homo_sapiens/UCSC/hg19/Homo_sapiens_UCSC_hg19.tar.gz) (75).

Duplicated aligned reads were removed using samtools rmdup. This removes potential PCR duplicates: if multiple read pairs have identical external coordinates, only the pair with highest mapping quality is retained (76).

Peak were called using MACS version 1.4.2 using the following options --nolambda –nomodel. This skips the model building step and the local background estimation as recommended for histone modification (77).

2.7.11.3 *FAIRE-seq Mapping*

DNA sequence reads obtained from FAIRE samples were aligned to the genome using bowtie2 version 2.0.6 against the UCSC hg19 genome using the following options --no-discordant --no-mixed –maxins 400 (properly paired reads with a maximum fragment length of 400). Duplicated reads were removed from the alignment bam files using samtools rmdup. The deduplicated bam files were converted to bed files using bedtools bamtobed.

Peaks were called using zinba . The first step involved generating an alignability file (athresh=4). Minimum number of hits per read allowed during the mapping process extension =120   average fragment library length. The second step involve generating a basealigncount file extension – average fragment library length – 120 filetype='bed'. The third step involves calling the peaks. zinba(align='alignallchr/', numProc=8, seq='LIB4142_LDI3552.bed', basecountfile= 'LIB4142.basecount', filetype="bed", outfile="L4142Rzinba", twoBit="hg19.2bit", extension=120, printFullOut=1, refinepeaks=1, broad=F, input="none")

2.7.11.4 *Gene Ontology and Promoter Analysis*

Ensemble sequence tags (ENST) were converted to "associated gene names" using the Ensemble BioMart tool (http://www.ensembl.org/biomart/martview). Genes lists were submitted to GSEA (http://www.broadinstitute.org/gsea) the gene ontology service as separate lists, and the top ten gene sets returning a p-value of ≤0.05 under the headings "canonical pathways", "transcription factor targets" and "GO molecular function".

2.8 Mass Spectrometry

2.8.1 Filter Aided Sample Preparation (FASP)

Protein solutions of 0.1-2 M CSK extracts (not mixed with LDS which would interfere with Coomassie assay) were measured by Coomassie assay to contain 0.25-0.4 µg/µL protein when 1 x 10 cm dish of cells was extracted in volumes of 100 µL (data not shown). To obtain at least 4 µg of protein, 25 µL of CSK extracts pre-mixed with NuPAGE LDS and reducing buffers to 1X were loaded onto an Amicon 30 kDa filter unit (Millipore, Cat # UFC503096) and an adapted version of the filter aided sample preparation protocol from Wisniewski *et al* (78)was followed to buffer exchange and tryptically digest proteins in CSK extracts.

200 µL UA buffer (8 M urea with 0.1 M Tris-HCl, pH 8.5) was added to each unit. Samples were then centrifuged at 18,000 *g* for 15 min at 20°C. The annotation of the filter pore size is based on the average molecular weight globular protein which the pores will retain when proteins are in a native conformation. Under denaturing conditions, such as presence of LDS or urea, the tertiary protein structure unfolds and the protein will not pass through the pores in the filter.

Flow-through from filters was discarded, a further 200 μL UA buffer was added to the samples and centrifugation was repeated, then the process was repeated once more. Cysteines in protein samples were then alkylated by addition of 100 μL 0.05 M iodacetamide in UA buffer, and incubating for 15 min at ambient temperature in the dark. Samples were washed through once with 200 μL UA as before.

After washing in UA buffer, sample was buffer exchanged by centrifuging through four times with 200 μL 50 mM ammonium bicarbonate (ABC) pH 7.8 in H_2O. Samples were then proteolytically digested by addition of 0.04 μg trypsin in 30 μL 50 mM ABC (approximately 100:1 protein: trypsin ratio). Filter units were sealed with parafilm and incubated overnight at 37°C. Digested peptides are sufficiently small to pass through the filter pores, so were collected by centrifugation at 18,000 g for 10 min. A further 50 μL of 50 mM ABC was centrifuged through the filter units. Samples were analysed for peptide content using 260/280nm ratios on a Nanodrop spectrophotometer.

2.8.2 In-gel digestion

When CSK extract samples for proteomics were processed using in-gel digestion, 25 μL of samples were pre-mixed with NuPAGE LDS and reducing buffer to 1X dilution and electrophoresed into NuPAGE 4-12% BisTris gels at 200 V until the dye front reached 1 cm into the gel. Samples were spaced with an empty lane between each sample to avoid cross-contamination, and each 1 cm portion of the lane containing the protein was excised using a fresh scalpel. Gel pieces were cut into approximate 1 mm cubes and placed into Lo-bind microcentrifuge tubes. Gel pieces were washed by addition of 200 μL 100 mM ABC in a 1:1 (v:v) mixture of acetonitrile and H_2O for 20 min. The supernatant was removed and the wash step repeated. A further wash was performed with 100% acetonitrile, and gel pieces were dried in a Speedvac centrifugal concentrator on medium heat setting for 20 min.

Protein in dried gel pieces was reduced by addition of 200 μL 100 mM ABC with 10 mM dithiothreitol, followed by one hour incubation at 56°C and then removal of supernatant. Cysteines in proteins were alkylated by addition of 200 μL 100 mM ABC with 50 mM iodoacetamide and incubation at ambient temperature in the dark for 30 min.

The supernatant was removed and the gel pieces washed in 200 µL 100 mM ABC for 15 min. The supernatant was again removed and gel pieces washed with 25 mM ABC in a 1:1 mixture of acetonitrile and H_2O for 15 min. Supernatant was removed and gel pieces washed in 100% acetonitrile for 5 min. Supernatant was removed and gel pieces dried in a Speedvac set to medium heat for 20 min.

0.04 µg of trypsin dissolved in 20 µL 25 mM ABC solution was allowed to absorb into gel pieces for 10 min. 25 mM ABC was added to gel pieces until they were covered in liquid, the tubes sealed and incubated overnight at 37°C. Supernatant was recovered from gel pieces and transferred to a separated tube. Gel pieces were washed 3 x 15 minutes with 50 % ACN in H_2O and supernatant retained and added to previous tube after each wash. Solutions were lyophilised for 20-30 minutes in a Speedvac on medium drying setting. Samples were re-suspended in 12 µL 0.1 % trifluoroacetic acid in H_2O and peptide content estimated by measuring 280 nm emission on a Nanodrop spectrophotometer.

2.8.3 Liquid Chromatography and Mass Spectrometry (LC-MS)

Peptide samples were diluted to 100 ng/µL with 0.1% TFA in H_2O and loaded onto a nanoAcquity UPLC system (Waters) equipped with a nanoAcquity Symmetry C18, 5 µm trap (180 µm x 20 mm, Waters) and a nanoAcquity BEH130 1.7 µm C_{18} capillary column (75 µm x 250 mm, Waters). The trap was washed for 5 min with 0.1% (v/v) aqueous formic acid at a trapping flow rate of 10 µL/min, after which flow was switched to the capillary column. Separation was achieved using a gradient elution of two solvents. Solvent A was 0.1% (v/v) formic acid and solvent B was acetonitrile containing 0.1% (v/v) formic acid, and the gradient profile was as follows: initial conditions 5% solvent B (2 min), followed by a linear gradient to 35% solvent B over 120 min, then a linear gradient to 50% solvent B over 5 min, followed by a wash with 95% solvent B for 10 min. During the gradient, the flow rate for the capillary column was 300 nL/min and the column temperature was set to 60°C. After completion of the gradient, the column was returned to initial conditions and re-equilibrated for 30 min before subsequent injections.

The nanoLC system was interfaced with a maXis MS/MS System (Bruker Daltonics) with a nano electrospray source fitted with a steel emitter needle (180 µm O.D. x 30 µm I.D., Proxeon).

To ensure samples had sufficient peptides to yield identifications, initial runs used a single injection of 300 ng of peptide onto the LC-MS system, and positive ESI-MS and MS/MS spectra were acquired using the 'AutoMSMS' mode to enable data dependent acquisition (DDA) of product ion spectra. The instrument settings used were: ion spray voltage= 1,400 V, dry gas= 4 L/min, dry gas temperature =160 °C, ion acquisition range= m/z 50-2,200. AutoMSMS settings were: MS: 0.5 s (acquisition of survey spectrum), MS/MS (collision induced dissociation (CID) with N_2 as collision gas): ion acquisition range=m/z 300-1,500, 0.1 s acquisition time for precursor intensities above 100,000 counts, and for signals of lower intensities (down to 1,000 counts) acquisition time increased linearly to 1s. The collision energy and isolation width settings were automatically calculated using the AutoMSMS fragmentation table: 5 precursor ions, absolute threshold 1,000 counts, preferred charge states= 2 − 4, singly charged ions excluded. One MS/MS spectrum was acquired for each precursor ion and former target ions were excluded for 30 s.

When the data were collected for relative quantification, injection volumes were normalised based on the total ion intensity as observed when extracted ion chromatograms were visualised in the Data Analysis software (Bruker). 80 − 120 ng of peptides in solution, as measured by Nanodrop spectrophotometry were injected into the LC-MS system, and positive ESI- mass spectra were acquired using the same parameters as described above for MS/MS with the exception that AutoMSMS mode was switched off. Three injections where total ion intensity was relatively consistent within replicates were acquired per sample.

When data was collected using scheduled precursor lists (SPL), a single injection containing 1.2 μg of peptides in solution per sample was injected into the LC-MS system, and positive ESI mass spectra were acquired using the same parameters as described above for MS/MS with the exceptions that AutoMSMS mode was switched off and active exclusion of acquisition of MS/MS for ions not included in the SPL was switched on.

2.8.4 Generation of Compound Lists from MS-only Data

LC-MS data from MS-only analysis were imported into Data Analysis (Bruker) and processed using the Bruker "Find Molecular Features" (FMF) algorithm. The FMF script is proprietary, but is described as being based on the principle that the chromatographic peak maxima of a compounds' ions in each of the different charge states it is measured in will be highly correlated in elution time (Figure 2.8.1).
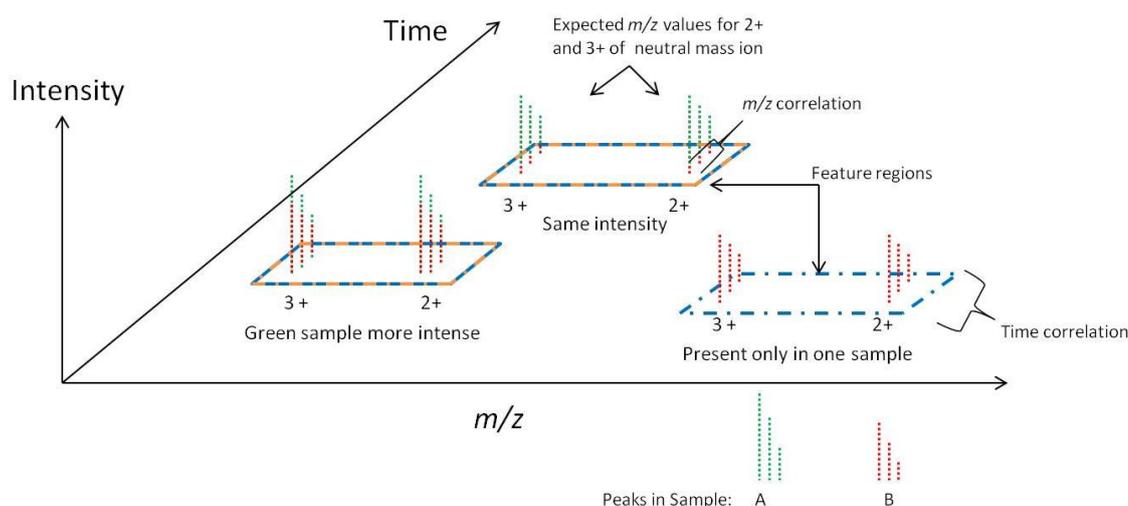
Figure 2.8.1. Feature picking by the FMF algorithm and alignment. Peaks are converted to line spectra and designated as a feature if they display characteristics such as tight time correlation, narrow *m/z* range and additional features such as 2+ and 3+ ions being the expected distance from one another in the *m/z* range. Features can then be aligned in the time dimension between several runs, such as triplicate MS-only data from two samples as was done here. The intensity of the peaks within feature which are reproducible between runs can then be compared across samples.

The algorithm converts profile spectra to line spectra and then defines features in sequential mass spectra which have peaks in a narrow mass range, calculated using the MS peak width. The criteria of time correlation and expected *m/z* differences between sequential charge states must be satisfied for the algorithm to designate the identification of a feature (compound).

Once a likely feature is identified, chromatographic traces are plotted and the peaks in each defined. Chromatographic peaks which have a difference in *m/z*

distance consistent with being the same compound in different charge states are assigned a charge status based on the algorithm's interpretation of these *m/z* distances. The algorithm then assigns a single calculated mass to each peak cluster based on the average mass of the peaks in the cluster, based on the interpretation of the charges. Compounds are then defined as peaks clusters which have the same neutral mass and highly correlated chromatographic peaks.

FMF was applied to MS-runs with a signal to noise ratio threshold of 3 (range 0.001 to 1000), which determines that the signal from a peak must be three times that of the background. The correlation coefficient was set to 0.7, which determines that clusters of peaks within a narrow *m/z* range with a calculated time correlation above 0.7 (range 0-1) will be subjected to evaluation as potential charge states of a compound. Minimum compound length was set to 20, which requires that each compound was observed in at least 20 consecutive mass spectra. The smoothing width was set to 10, which determines the number of compound peaks to be used for calculation of smoothed chromatographic peaks. Additional smoothing was applied, which smoothes individual peaks within compound prior to creation of chromatographic peaks. This prevents isolated peaks from non-related compounds interrupting true compounds in complex samples. The "Proteomics" option was also applied, which compares the intensities of compounds across the likely charge-states with the expected pattern of decreasing intensity with increasing charge, and omits the features if they do not follow the expected pattern. Retention times at the beginning (0-35 min) and end (120-165) of the elution gradient were excluded from feature generation. Additionally, only compounds with *m/z* 300-1600 were included, excluding a window of *m/z* 1221-1225 around the lockmass calibration compound.

2.8.5 Label Free Relative Quantifications in Bruker ProfileAnalysis

Each FMF processed trace was then imported into ProfileAnalysis (Version 2.0, Bruker). Traces were split into groups according to the sample from which they came for Student's t-tests calculation. The MS parameters for Student's t-test model generation included time-alignment of compounds, using an algorithm which takes into account non-linear shifts in retention times (79). This aligns features in all traces with similar elution characteristics, using an automatically determined master run as the anchor point for all elution times. Advanced bucket generation was used, which creates a new "bucket", or bin, for grouping of

compounds between runs with similar properties based on the properties of the compound calculated by the FMF algorithm. Data regions for bucket generation were given the same time and *m/z* exclusion parameters as the FMF files. Time and mass tolerances for bucket generation were calculated automatically from time alignment. Normalisation was set to quantile, which attempts to normalise the distribution of the intensity of features in LC-MS runs to correct for variations in injection volumes between runs. Student's t-test values can only be calculated by the ProfileAnalysis software when a group contains non-zero values. Therefore, with a comparison of three samples in each group, to increase stringency of compounds included in intensity calculations it was determined that only one missing value in the six would be tolerated. This was achieved by filtering buckets which contained values from <5 of the 6 runs. Student's t-tests models were generated for all pairwise comparisons between samples.

2.8.6 Scheduled Precursor List Generation

Retention times and *m/z* of compounds identified as having differential abundance ≥2-fold with a p≤0.05 in pairwise comparisons were exported into scheduled precursor lists. Lists were checked for density, so that less than 20 events occurred within every 1 min to allow the mass spectrometer to collect sufficient ions for MS/MS on each event. In cases where lists were too dense, events were split into separate lists until densities reduced to acceptable levels.

2.8.7 Mascot searches

MS/MS data from DDA and SPL analyses were either imported into ProteinScape (Bruker) and submitted to Mascot (Matrix Science) from within the Bruker software, or exported from Progenesis in .mgf format and submitted directly to the Mascot server. Mascot searches were performed with the following parameters: Database: IPI_Human, Taxonomy: All Entries, Enzyme: Trypsin, Modifications: Carbamidomethyl (C) Fixed, Oxidation (M): Variable, Peptide tol -/+ 10.0 ppm, #^{13}C: 0, 0.1 Da, Peptide Charge 2+ and 3+: monoisotopic, Instrument type CID: maXis.

2.8.8 Calculations of p-Values for Buckets Generated in ProfileAnalysis

When Student's t-tests were calculated manually for peptides which were absent in one of the samples, no filtering was applied in ProfileAnalysis and all compounds were exported. P-values were calculated using a 2-tailed t-test as performed in ProfileAnalysis (Bruker). Where all samples in one triplicate group were zero, it was required that intensity counts were present in all three of the comparator triplicate group for p-values to be accepted.

2.8.9 Nonlinear Dynamics Progenesis LC-MS

MS-only and MS/MS data were converted to mzXML using Compass (Bruker). Files were imported into the Progenesis software and the MS peaks aligned for all imported runs. Runs were then designated into groups for comparison, and MS/MS data from DDA analyses exported and converted to .mgf for submission to the Mascot search engine.

Peaks which had variance between groups sufficiently low to return a p-value of ≤0.05, and had a fold-change between the groups of ≥2 were included for comparison. Compounds were further filtered for those with retention times >0.25 s and with intensity counts of <1000. The SPL list was generated from the remaining MS peaks.

3 PPARγ in Differentiation of Human Urothelium *in vitro*

3.1 PPARγ

3.1.1 Nuclear Receptor PPARγ

Peroxisome-proliferator activated receptors (PPARα, PPARβ/δ, and PPARγ) are a sub-family of the nuclear receptor (NR) superfamily, so named from the ability of the first discovered of the sub-family, PPARα, to induce proliferation of the peroxisome sub-cellular organelles in response to a variety of compounds. PPARγ, although homologous with PPARα and PPARβ/δ, does not share this function. One or more isoforms of PPARγ are implicated in the control of cell identity in a variety of tissues, especially adipocytes. PPARγ has two known major protein variants, γ1 and γ2. There are four known promoters (γ1-4) for PPARγ. As the final transcripts for γ1 and γ3 only differ in 5' untranslated regions they both encode the same protein product (PPARγ1) (80-82). PPARγ2 coding sequence contains one more 5' exon than PPARγ1 that encodes extra 28 amino acid at the N-terminus, making it the largest of the canonical isoforms. PPARγ4 codes for a protein six amino acid longer than PPARγ1, whose gene expression was discovered first described in rhesus macrophages (83), and protein expression has been observed at low levels in human macrophages(84).

As a NR, PPARγ consists of distinct domains which are conserved across the superfamily and have well described function. It is the properties of these physical features that places PPARγ at the nexus of distinct cell signalling networks. A general outline of NRs is that the activation function 1 (AF1) and AF2 domains facilitate interactions with co-factors; the DNA binding domain binds sequence-specific DNA stands, and the hinge region connects between the DNA-binding domain and the ligand-binding domain (Figure 3.1.1). Ligands with different affinities vary in the extent to which their binding causes conformational changes in the AF2 domain, thus affecting the surface presented for interactions with co-factors (85), and hence the different effects on transcription that each compound exhibits.

PPARγ2

| Amino Acid: 1 | 28 | 104 | 160 | 207 | 273 | 424 | 505 |

Figure 3.1.1. Structural Outline of PPARγ2 domains. Numbered amino acids depict exon boundaries. Amino acid 1-28 are the additional exon unique to PPARγ, and 28-505 are the constituents of PPARγ1. The N-terminal AF1 region of PPARγ1 has weak ligand-independent transcriptional activation capability through interaction with proteins such as p300 (86). The DNA-binding domain provides sequence-specific DNA binding when dimerised with RXRα. The hinge region facilitates flexing of the DNA binding domain away from the ligand-binding domain (LBD). The LBD contains the AF2 domain which interacts with cofactors in a ligand-dependent fashion (86). The DNA binding, hinge region and LBD all contribute to heterodimerisation with RXR (87).

PPARγ2 was first identified in mammals alongside its heterodimerisation partner retinoid X receptor (RXRα), as part of the binding complex at an enhancer of the fat-specific AP2 gene (57, 88). Much of the initial work on PPARγ2 was undertaken to elucidate its role in adipose cell differentiation, but interest gathered pace once it was identified as the target for a class of anti-diabetic drugs known as thiazolidinediones (TZDs) which modulate insulin sensitivity (89, 90).

There is evidence that PPARγ isoforms are also involved in pathways which influence the differentiation status of other cell types, including, but not limited to, osteoblasts, macrophages, urothelium and prostate (46, 91-95). During the differentiation of osteoblasts and macrophages, PPARγ has been observed to associate with many different co-factors, some of which have proven to be essential for development of the differentiated phenotype (96-98). The co-factors involved in urothelial and prostate differentiation are less well understood. Across all cell types where it is implicated in differentiation, PPARγ has been found to be affected by several distinct cellular signalling pathways which directly post-translationally modify it or its interactors (60, 99-101).

3.1.2 PPARγ in Adipogenesis

In vitro differentiation of mouse adipocyte precursor 3T3-L1 cells can be initiated by inhibition of the cell cycle, after which Cyclin D1 drives upregulation of CCAAT enhancer-binding protein (C/EBP) δ and C/EBPβ, which in turn promote expression of PPARγ and C/EBPα (102). PPARγ then dimerises with RXRα, and the heterodimer binds DNA at sequence-specific motifs known as peroxisome proliferator response elements (PPREs) (103). During adipogenesis, PPARγ/RXRα binding occurs at PPREs in enhancers of adipogenic genes, including one for PPARγ itself, thus driving a positive feedback loop which results in a self-sustaining programme of changes in gene expression (39, 88).

This simple outline of adipogenesis can be misleading, as there are other subtle controls on PPARγ activity which must first be overcome before adipogenesis can begin. Activation of glucocorticoid receptor (GR) alongside the inhibition of the cell cycle enhances adipogenesis, as the activated GR remodels the epigenetic environment on the chromatin at sites around adipogenic genes, thus "preparing" the sites at which PPARγ will bind (104). PPARγ/RXRα heterodimers can then bind to DNA, but transcriptional activity remains inhibited as nuclear receptor co-repressor 1 (NCOR) and NCOR2 and their associated protein complexes (105) bind to the PPARγ/RXRα on the DNA. Further on in the differentiation process, NCOR1 or NCOR2 are replaced by transcriptional activators. In adipocytes, it is not clear what signals this cofactor exchange, but in other systems it is mediated by either protein kinase C (PKC), casein kinase (CK1), or glycogen synthase kinase (GSK) which phosphorylate NCOR complex members F-box-like/WD repeat-containing proteins TBL1 (TBL1X) or TBL1XR1 (106).

PPARγ and co-operating transcription factor proteins such as C/EBPα serve as foundations which bring transcriptional co-activator proteins into contact with the DNA (107). In adipogenesis, PPARγ has been shown to bind to general nuclear receptor co-activators, including thyroid hormone receptor associated protein 220 (TRAP220) (108), steroid receptor co-activator 1 (SRC-1) (109), and PPARγ coactivator proteins 1A and 1B (PGC-1α and PGC-1β) (110, 111). This profligate co-factor binding creates redundancy in the necessity for PPARγ co-factors, as demonstrated when genetic ablation of C/EBPα in mouse pre-adipocytes did not affect the adipogenic capacity of the cells (112).

In *in* vitro adipogenesis, differentiation of several different pre-adipocyte cell lines can be achieved through treatment with the same cocktail of compounds

(reviewed in (113)). Several transcription factors and co-factors have been shown to be necessary for adipogenesis (reviewed in (112)). However, only PPARγ (γ1 and γ2) has been shown to be able to trans-differentiate fibroblasts into adipocytes when over-expressed in the presence of PPARγ-activating ligands (114). It has been reported that ligand binding is necessary for PPARγ to facilitate binding to adipogenic co-factors, by virtue of the induced conformational change in the AF2 domain in the LBD to expose interaction domains (106).

However, it has also been reported that a PPARγ mutated in the ligand-binding domain (Q286P), which lacked increased transcriptional activity in response to exogenous agonists, was still capable of inducing adipogenesis when ectopically-expressed in fibroblasts lacking wild-type PPARγ (114). The authors proposed that although they could not rule out the possibility of endogenous ligands being present that bound the mutant PPARγ in preference to the introduced ligands, the relative amount of nuclear PPARγ in their expression system were similar to that of fully differentiated adipocytes. The expression observed in the PPARγ-transfected fibroblast cells was much higher than at the onset of adipogenesis and so a more abundant PPARγ with weaker transcriptional activity was proposed to have overcome the necessity for added PPARγ ligand to increase transcriptional activity during induction of differentiation. Interestingly, the same study showed that mutation of the AF2 domain (E499Q) prevented differentiation, implying that the C-terminus of PPARγ is necessary for adipogenesis. This was corroborated by a similar study, in which the truncation of five C-terminal amino acid of PPARγ2 produced a dominant negative isoform able to occupy the PPRE sites on the genome, but that was unable to induce adipogenesis(115). PPARγ splice variants lacking the ligand-binding domain have been reported to be expressed in cancer-derived cell lines, with the implication that cellular differentiation is deficient in these cancers due to the presence of such dominant negative isoforms of PPARγ which are unable to initiate transcription as well as the wild-type protein (116, 117).

The N-terminal region of PPARγ has also been shown to have significant effects on adipogenesis. PPARγ1 or PPARγ2 overexpression in mouse pre-adipocytes can induce adipogenesis, although PPARγ2 is more potent (112), implying that the additional domain in PPARγ2 influences adipogenesis, likely through recruitment of co-factors which do not bind PPARγ1. Replacing the N-terminus of PPARδ with that of PPARγ2 (residues 1-203) imparts significant adipogenic ability on the

fusion protein (118). However, deletion of the N-terminus of PPARγ2 results in greater adipogenic potential in cells expressing the truncated protein, an observation attributed to the loss of an inhibitory phosphoserine on residue 112 (88). Other domain deletion studies showed PPARγ2 with a truncated N-terminus (removal of residues 1-136) maintains the ability to upregulate 95% of PPARγ-induced adipogenic genes (86). This deletion did not affect the binding of the transcriptional regulator TRAP220 to PPARγ, but did prevent the attachment of Cyclic AMP-responsive element-binding protein (CREB) and p300.

3.1.3 PPARγ in Differentiation of Macrophages and Osteoblasts

Examples of the function of PPARγ in cell types other than adipocytes highlight its context-dependent mode of action. In macrophages and osteoblasts, the role of PPARγ is different from that in adipocytes. Signalling through PPARγ has been shown to influence mouse macrophage cells in their ability to become "alternately activated", anti-inflammatory macrophages. The way PPARγ affects gene expression in this context is very different to that in adipogenesis. In macrophages, monomeric ligand-bound PPARγ covalently bound to small ubiquitin-like modifier (SUMO) binds to NCOR1-nuclear factor κ B (NF-κB) complexes on the DNA, obstructing binding of the 19S proteasome (119). This prevents degradation of the inhibitory complexes, thereby repressing NF-κB target genes.

Cells harvested from bone mesenchyme can be differentiated into adipocytes, chondrocytes, myocytes or osteoblasts (reviewed in (91)).Thus, as PPARγ drives the adipogenic potential of these cells, it is an inhibitor of osteogenesis. Canonical Wnt-β-catenin and non-canonical Wnt signalling in osteoblasts suppress the activity of PPARγ via recruitment of histone deacetylases and methyltransferases to DNA-bound PPARγ, thus directly repressing PPARγ-influenced gene expression.

3.1.4 PPARγ in Differentiation of Prostate and Urothelium

PPARγ isoforms are expressed in both prostate epithelia and urothelium, both of which arise from the embryonic urogenital epithelium (120). The necessity for PPARγ activation in directing a programme of gene expression changes which culminates in late/terminal differentiation of urothelial cells is reasonably well established, although the cofactors that help drive PPARγ-based transcription are not well understood.

In normal human urothelial (NHU) cells, PPARγ is constitutively expressed. When NHU cells are cultured in serum free medium, spontaneous upregulation of the urothelial differentiation marker gene uroplakin 2 (UPK2) has been shown not to occur (25). Treatment of NHU cells with the PPARγ-specific ligand troglitazone (TZ) can induce upregulation of urothelial differentiation markers. However, concurrent inhibition of the EGFR with PD153035 (PD) or inhibition of the EGFR target proteins ERK1/2 or PI3K ensures a more robust expression of the differentiation marker UPK2, as well as cytokeratins and tight junction proteins associated with intermediate and superficial native urothelium (Table 3.1.1) (2, 26, 121). Two genes, aquaporin 3 (AQP3) and fatty acid binding protein 4 (FABP4), which have PPRE elements in their promoters are rapidly and highly upregulated after induction of differentiation in NHU cells *in vitro* (122). PPARγ-specific inhibitors that prevent TZD binding to PPARγ attenuate the differentiation-associated changes in gene and protein expression (2), as does whole-gene siRNA targeting of PPARγ (34).

Protein expression change after TZ&PD treatment *in vitro.* →	CLDN3	CLDN4	CLDN5	CLDN7	CK14	CK13
	↑↑	↑↑	↑↑	↑	↓	↑↑

Key (Protein expression change)	Upregulated ↑↑	Downregulated ↓	Slight upregulation ↑

Expression and localisation in normal urothelium *in vivo*		CLDN3	CLDN4	CLDN5	CLDN7	CK14	CK13
	Superficial	X	(Intercellular Junction)	(Intercellular Junction)			
	Upper Intermediate		(Diffuse Cytoplasmic)		(Intercellular Junction)		(Diffuse Cytoplasmic)
	Lower Intermediate		(Diffuse Cytoplasmic)		(Diffuse Cytoplasmic)		(Diffuse Cytoplasmic)
	Basal		(Diffuse Cytoplasmic)				(Diffuse Cytoplasmic)

Key (Protein localisation)	
Kissing Points of Cells	X
Intercellular Junction	(dark shading)
Diffuse Cytoplasmic	(grey shading)
Negative	

Table 3.1.1 Expression of selected proteins associated with urothelial differentiation *in vitro* and *in vivo*. Histological data from previous publications (2, 25) has demonstrated the distribution of keratins and claudins in native human urothelium, a selection of which are displayed here. Treatment of cells with TZ and PD153035 results in upregulation of the protein markers (2) apart from CK14. CK14 is a squamous marker that is not expressed in normal native urothelium (25) and is downregulated by *in vitro* differentiation.

Taken together, the above studies provide strong evidence that PPARγ activation is both necessary and sufficient for *in vitro* differentiation of NHU cells. However, although PPARγ is able to initiate differentiation, there is evidence that transcription factors upregulated by PPARγ have a role to play in maintaining expression of differentiation markers. The initial hypothesis for the existence of such intermediary factors came about because the urothelial differentiation marker UPK2 does not have a PPRE in the promoter region 2000 bp upstream from the transcription start site, and upregulation of UPK2 mRNA levels occurs 24-48 h after induction of differentiation with TZ and PD, later than other markers such as AQP3 and FABP4 which upregulate within 6 h (34). This circumstantial evidence suggested that PPARγ was driving expression gene expression after ligand treatment, but not directly regulating the terminal differentiation markers from their promoters. A subsequent investigation identified several transcription factors including forkhead box A1 (FOXA1) and interferon regulatory factor (IRF1) which had PPREs in their own promoters, and were upregulated within 12 h of induction of differentiation. Both factors were shown to bind to the promoter region of UPK2 (34). Their influence on terminal differentiation was further confirmed by siRNA knockdown, which inhibited the expression of the differentiation markers cytokeratin 13 (CK13) and claudin 3 (CLDN3). This established that PPARγ-driven upregulation of FOXA1 and IRF1 was able to drive upregulation of terminal differentiation markers in urothelium.

As well as transcription factors which upregulate urothelial differentiation, there is some knowledge of those which can inhibit it. Application of transforming growth factor beta-1 (TGFβ1) to NHU cells significantly reduces UPK2 upregulation after induction of differentiation with TZ and PD, without affecting ERK or RAC-alpha serine/threonine-protein kinase (AKT) signalling (123). TGFβ1 signalling affects transcriptional activity of mothers against decapentaplegic homolog (SMAD) proteins, of which SMAD3 was demonstrated to be upregulated in NHU cells after induction of differentiation.

Androgen receptor is considered the main driver of differentiation in prostate cells, but recent work in human prostate cells suggests that the relative abundance of PPARγ1 and PPARγ2 has a role to play in shaping prostatic cell identity (46). Cultured benign human prostate epithelial cells (BHPrEs) were shown to express a greater abundance of PPARγ2 as compared to PPARγ1. shRNA knockdown of PPARγ2 reportedly promoted transdifferentiation into urothelium, as measured

by expression of cytokeratin 20 which is normally restricted to superficial urothelial cells *in vivo*. However, the knockdown of PPARγ2 protein was far from complete, making the assertion that absence of PPARγ2 allows urogenital cells to differentiate into urothelium a questionable conclusion. In addition, although the PPARγ2 knockdown BHPrE cells expressed the urothelial differentiation marker CK20, they also expressed CK14. The authors suggest that CK14 is a marker of basal urothelium, when it is in fact only observed in human urothelium in squamous metaplasia, wherein cells are improperly differentiated (49). The conclusion that knockdown of PPARγ2 in prostate helping to achieve a transdifferentiation to urothelium is therefore based on keratin expression patterns which do not match with observations of native urothelium. When Strand *et al* cultured their PPARγ2 knockdown cells in an *in vivo* explant setting, structures resembling urothelium developed. However, the knockdown of PPARγ2 was not assessed in the *in vivo* explants and the conclusion that PPARγ2 knockdown aid urothelial differentiation also assumes that PPARγ2 is not expressed in differentiated urothelium, which is unresolved.

The same study by Strand *et al* (46) showed that concurrent shRNA knockdown of both PPARγ1 and PPARγ2 isoforms also reduced expression of androgen receptor, with cells adopting a squamous undifferentiated phenotype. This squamous phenotype in PPARγ1/2 knockdown BHPrE cells has a similar keratin expression profile to that seen in normal human urothelium cultured in serum and NR-ligand free conditions (25, 56), indicating that there may be some similarities in phenotype between proliferative undifferentiated urogenital-derived cells with non-ligand activated PPARγ.
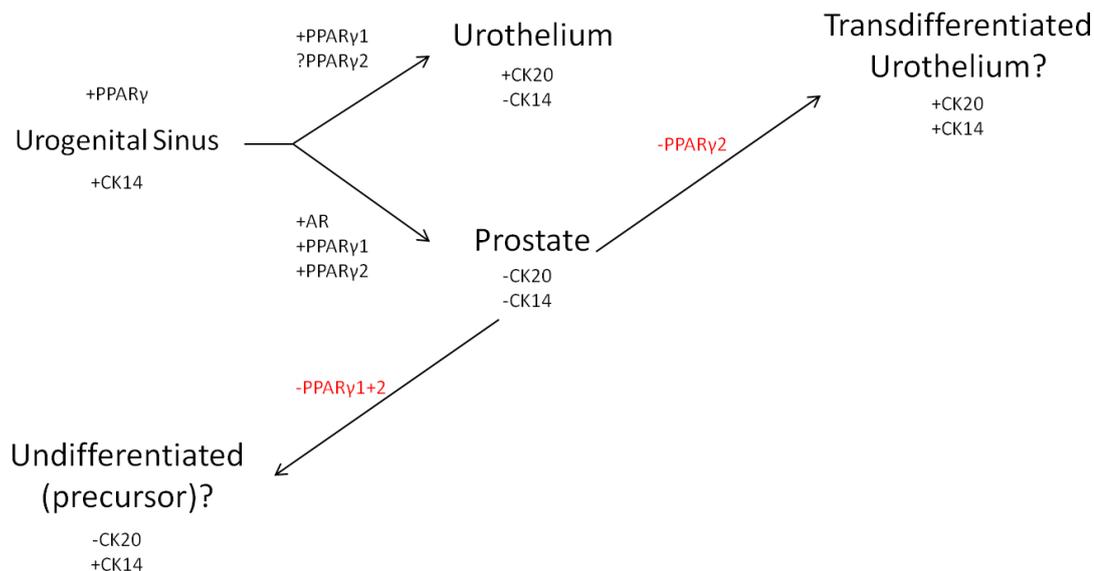
Figure 3.1.2. Cytokeratin and PPARγ expression in BHPrE cells as observed by Strand *et al* (46). Urothelium and prostate both derive from the same embryonic urogenital sinus, which expresses CK14 (124). Normal prostate cells show expression of PPARγ1, PPARγ2 and AR, and do not have expression of CK14 or CK20. Normal urothelium does not express AR or CK14, and expresses PPARγ1 and CK20, but expression of PPARγ2 is unknown. shRNA knockout of PPARγ2 by Strand *et al* in an *in vitro* setting resulted in a change in expression of CK isoforms in BHPrE cells from a prostate phenotype negative for CK14 and CK20, to a phenotype where it was positive for both. shRNA knockdown of PPARγ1/2 by the same authors resulted in a shift of expression to a CK20 negative and CK14 positive phenotype.

3.1.5 Isoforms of PPARγ in Urothelium

The work in prostate cells discussed above highlights that the isoforms of PPARγ that are present throughout the *in vitro* differentiation process are unknown. Varley *et al* (26) performed immunoprecipitation to PPARγ, and when immunoblotting against phosphorylated PPARγ observed two isoforms of PPARγ at 50 and 52 kDa in non-differentiated cells.

Georgopoulos *et al* (31) reported an increase an isoform of the presumed 52 kDa isoform of PPARγ 24 h after induction of differentiation with TZ and PD in NHU cells. This upregulation did not occur in telomerase (hTERT) immortalised human urothelial cells, which have compromised differentiation capacity (31). In addition, Georgopoulos *et al* also reported a PPARγ-reactive band above the PPARγ bands which resembled the 50 and 52 kDa isoforms reported by Varley *et*

al. Since Varley *et al* (26)have shown that the 50 and 52 kDa isoforms are both phosphorylated in non-differentiation induced cells, then these must be distinct isoforms and not a closely migrating phosphorylated and non-phosphorylated doublet which is often observed for PPARγ1 and PPARγ2 (125).

The calculated molecular weights of PPARγ1 and PPARγ2 based on the amino acid sequences in the Uniprot (www.uniprot.org) database are 55.4 and 58.6 kDa respectively. Tontonoz *et al* (88) cloned and expressed human PPARγ2 in an L[35S]-methionine *in vitro* translation system, which gave rise to two major bands at 56 and 52 kDa along with a faint band below 52 kDa, as detected by autoradiography. The band annotated as 52 kDa was equivalent to the separately expressed and purified PPARγ1 run in a parallel lane on their western blot (Figure 3.1.2). The identity of the smaller band below the PPARγ1 is not discussed in the above publication, despite the objective evidence of its presence.

This pattern of three PPARγ bands from *in vitro* translation resembles that obtained from immunoblotting of whole cell lysates of differentiated 3T3-L1 adipocytes, although the annotation of the which protein is PPARγ1 and which is PPARγ2 differs between publications (126, 127). Fleming (122) showed that extracts from nuclei of differentiated 3T3-L1 and NHU cells contained two PPARγ species each, as did whole cell lysates of undifferentiated NHU cells. As PPARγ2 is the larger isoform in 3T3-L1 cells, it was therefore assumed that the larger isoform of PPARγ in nuclear extracts from NHU cells was PPARγ2 and the PPARγ at 50 kDa was PPARγ1. However, the presence of only two PPARγ-reactive bands does not agree with the observations of Georgopoulos *et al,* who reported three PPARγ-reactive bands on western blots of whole cell lysates of NHU cells (Appendix Figure 7.1.1). Thus there remains an unanswered question about which isoforms of PPARγ are present in NHU cells, and what their contribution is to differentiation.

If the 50 and 52 kDa PPARγ in NHU cells are indeed PPARγ1 and PPARγ2 accompanied by a higher molecular weight isoform in NHU cells, there is evidence from other cell types as to what their origins may be. Publications which address the issue of high molecular weight PPARγ-reactive bands are most often studying the post-translational modification of PPARγ (119, 128-132). Gibbings *et al* (128) observed a close migration of two isoforms of PPARγ annotated as 50 and 55 kDa, with a higher molecular weight isoform annotated as 62 or 67 kDa depending on the cell type investigated. In addition to this, Gibbings *et al* saw a reduction in the

higher molecular weight isoforms of PPARγ annotated as 62 / 67 kDa of PPARγ when using siRNA targeting SUMO1. They thus concluded that these higher molecular weight isoforms are post-translationally modified by SUMOylation. This observation has been replicated elsewhere by Floyd *et al* (132), who were able to generate higher molecular weight PPARγ by mixing PPARγ and the proteins involved in SUMOylation together in an *in vitro* translation system. This *in vitro* translation system used PPARγ2 DNA as a template, yet gave rise to four bands detected by autoradiography. PPARγ2 transcript has two known transcription start sites, one of which is the start site for PPARγ1 (83). The two bands which were lower in molecular weight than PPARγ1 and PPARγ2, at 42 and 44 kDa were not discussed in the paper by Floyd *et al,* and have not been investigated elsewhere. It remains a possibility that these are variants produced by the presence of extra start codons; however Floyd *et al* did not detail whether they used plasmid-containing genomic DNA sequence of PPARγ2 or DNA amplified from translated PPARγ2 mRNA as a template for transcription. Genomic DNA would have included intronic DNA with potential alternative transcription start sites downstream of that of PPARγ1, but isolated PPARγ2 cDNA would not have.

This mixture of reports of sizes of PPARγ across several different cells types has makes it difficult to ascertain the identity of PPARγ proteins when interpreting the literature, and the identity of many potential transcript variants or post-translationally modified isoforms remains unresolved.

3.1.6 Localisation of Nuclear Receptors and Transcriptional Activity

PPARγ has been observed by immunofluorescence to change localisation to the nucleus in response to inhibition of EGFR using PD (26). This observation is in agreement with previous studies which showed PPARγ phosphorylation by ERK affects localisation (60). However, other studies of PPARγ in NHU cells have reported a mixed cytoplasmic and nuclear localisation (56) or a predominantly nuclear localisation in untreated near-confluent cells (56, 95, 121, 133).

One simple way to assess if an expressed transcription factor has the potential to initiate transcriptional activity is to observe its localisation in the cell. Nuclear localisation of NRs can be either ligand-dependent, as for glucocorticoid receptor, or ligand-independent as has been shown for the PPARs (134). However, although PPARs are generally nuclear, the picture is complicated by the addition of post-

translational modifications which can cause it to become localised in the cytoplasm (100, 135). Thus, for nuclear receptors the localisation pattern and binding affinity for the nuclear compartment is highly context dependent and indicative of the functional status of the protein.

Non-phosphorylated PPARs are reported to be localised to the nucleus regardless of their liganded status, with transcriptional control of PPARs being dependent on the change of binding partners from transcriptional repressors to activators upon binding of agonist ligands (100, 115, 136). Work using ectopic expression of PPARγ and RXRα demonstrated that the PPARγ/RXRα heterodimer is able to bind DNA in the absence of exogenous ligand (137), but that once ligand binds the diffusion rate of PPARγ is slowed as measured by fluorescence resonance energy transfer, suggesting the PPARγ/RXRα dimer is interacting with alternative binding partners (138). However, as the authors highlighted, these results are complicated by the observed cell-cell variation in localisation and abundance of PPARγ due to differing transfection efficiencies, and the inability to rule out the presence of endogenous ligands which might affect localisation.

This potential exchange of co-factors and movement of PPARγ underscores that the nucleus is a dynamic organelle, which contains many structures and complexes that influence the behaviour of proteins within it. The structural elements of the nucleus consist of what is loosely termed the "nuclear matrix" (NM), which is an overarching term for a group of structures including a filamentous skeleton in the nucleus and large RNA and ribonucleoprotein complexes implicated in the control of transcription (reviewed in (62)). Chromatin and its constituent proteins can interact with the NM to varying degrees. The binding affinity of an NR-containing chromatin-bound complex for the nuclear matrix or chromatin can be ascertained by the addition of buffers containing detergent and varying molarities of salt, a method dubbed cytoskeletal (CSK) extraction (139). Extraction with 2 M NaCl and 0.1% w/v Triton-X100 removes a large proportion of the histones from the cell, and the non-extracted nuclear proteins are thought to be bound to the NM. Pre-digesting DNA with nucleases before salt extraction can extract some proteins bound to DNA which are not extracted by salt alone. Studies with estrogen receptor alpha (ERα) exemplify the utility of this approach in identifying the development of NR-NM interactions after a change in liganded status of the NR. Within 30 minutes of addition of ER-specific ligands, nuclear foci of ERα were visualised by Stenoien *et al* (140). These

foci were not removed by CSK with salt alone or after nuclease digestion, and were induced to form with addition of ERα agonists or antagonists. Interestingly, transcriptional co-activators such as SRC-1 did not co-localise with the foci unless a strong agonist was added.

NM resident transcriptional repressors have been described for PPARγ. Scaffold attachment factor B1 (SAFB1) was identified to interact with the DNA binding and LBD of PPARγ by Debril *et a*(141), following transfection of cells with constructs of combinations of PPARγ domains. SAFB1 also interacted with all other nuclear receptors tested and inhibited their transcriptional activity, thereby leading to the conclusion it is a general repressor for nuclear receptors. Expression of SAFB1 was found to be reduced during adipogenesis, leading to speculation that it has an inhibitory role in differentiation of pre-adipocytes. The localisation of a transcription factor and its resistance to extraction can help to identify whether it is resident on the chromatin and if it alters its resistance to extraction after ligand treatment.

3.2 Experimental Aims and Approach

3.2.1 Aims

It was the aim of the work described in this chapter to further investigate the expression and localisation of PPARγ isoforms in NHU cells during *in vitro* differentiation, and where possible to assess their relative contributions to differentiation.

It was hypothesised that PPARγ isoforms in NHU cells would translocate to the nucleus and move from a CSK soluble to insoluble fraction after induction of induction of differentiation with TZ and PD. A further hypothesis was that the use of siRNA specific to PPARγ2 would inhibit the expression of differentiation markers in NHU cells and enable identification of PPARγ2 by reducing expression of PPARγ2 protein.

3.2.2 Experimental Approach

3.2.2.1 *Overview*

Nuclear fractionations, CSK-NaCl and CSK-DNaseI-NaCl extractions were performed on NHU cells, and localisation of PPARγ was assessed by immunofluorescence and western blotting. This was intended to test if PPARγ translocates to the nucleus upon stimulation of differentiation and whether it is part of the nuclease-sensitive chromatin fraction.

Immunoprecipitations targeting PPARγ were performed to assess whether isoforms of PPARγ identified on western blots also precipitated when using anti-PPARγ antibodies. This was performed in order to resolve whether all bands observed by western blotting are genuine isoforms of PPARγ rather than non-specific targets of the antibody.

siRNA was utilised to knock down PPARγ1 and PPARγ2 during *in vitro* differentiation of NHU cells by targeting a sequence of coding RNA common to both isoforms, or PPARγ2 alone was targeted using a sequence in PPARγ2 not contained in PPARγ1. Western blotting was then used to asses which PPARγ isoforms altered in abundance during differentiation, as well as monitoring the upregulation of protein markers of differentiation. The purpose of this experiment was to test if the abundance of any isoforms of PPARγ was affected by the

introduction of siRNA targeting the exon which differentiates PPARγ2 from PPARγ1.

Anacardic acid, a known inhibitor of SUMOylation, was applied to NHU cells at concentrations ranging from 0-10 μM, and western blotting was used to assess if the relative abundance of the 50 and 58 kDa isoforms of PPARγ altered in response. This experiment was designed to test if the abundance of any isoforms of PPARγ was affected by inhibiting the process of SUMOylation.

3.2.2.2 *Anti-PPARγ Antibodies*

All described major PPARγ isoforms share six coding exons, which constitute the totality of the PPARγ1 protein. PPARγ2 has an N-terminal addition of 28 amino acids. Thus, all antibodies which detect PPARγ1 also detect PPARγ2. Three mouse monoclonal antibodies targeting PPARγ were used in this study (Figure 3.2.1). An antibody which binds the C-terminus (E8 clone, Santa Cruz) was used for all western blot analyses, and another antibody targeting the same region (81B8 clone, Cell Signaling) was used for immunoprecipitations. For immunofluorescence studies, an antibody targeting an N-terminal region common to both PPARγ1 and PPARγ2 was used (gift from GSK (142)).
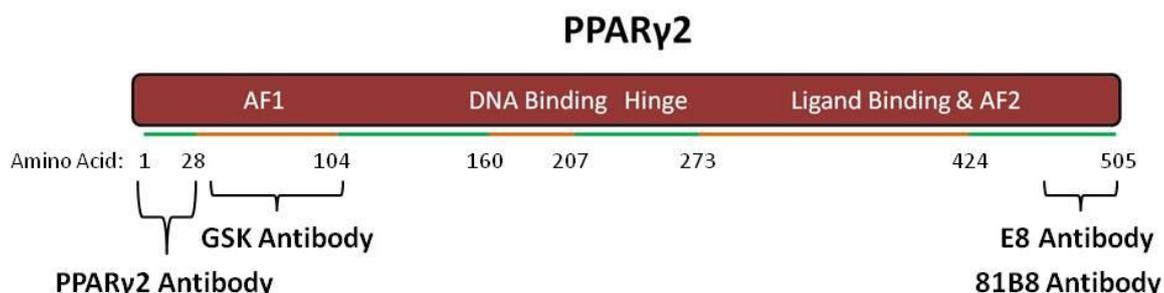
Figure 3.2.1. Anti-PPARγ antibody binding sites. The GSK antibody used for immunofluorescence studies was raised against an N-terminal region of PPARγ common to both PPARγ1 and PPARγ2 (142). The Santa Cruz (sc-7273, E8 clone) antibody used for western blotting was raised using a peptide matching amino acids 480-505 of PPARγ2. The antibody used for immunoprecipitations was the 81B8 clone (Cell Signalling), which is raised using a peptide from the same C-terminal region of PPARγ as the E8 antibody. The PPARγ2 antibody was raised using a peptide matching the N-terminal 28 amino acids of PPARγ with an added cysteine at the N-terminus of the peptide to improve immunogenicity.

3.2.2.3 *Nuclear Extractions and Immunoprecipitations*

A "nuclear complex Co-IP" kit (Active Motif, Cat#54001) was used to hypotonically lyse scrape-harvested NHU cells in the presence of phosphatase inhibitors. Nuclei were then collected by centrifugation and supernatant retained as the cytoplasmic fraction. Nuclei were resuspended in a low salt (0.075 M) buffer and DNA enzymatically fragmented to release intact protein-DNA complexes. Nuclei were collected by centrifugation and supernatant retained as the nuclear extract. All extraction steps were carried out on ice or at 4°C. Proteins extracted by hypotonic lysis, extraction after DNA fragmentation, and the remaining fraction were compared by western blot for presence of PPARγ isoforms.

To provide additional evidence that PPARγ isoforms observed by western blots in fractionated cells were genuinely PPARγ, 200 μg (as measured by Coomassie assay) of NHU cell nuclear extracts obtained using the nuclear Co-IP kit were subjected to immunoprecipitation with anti-PPARγ (81B8) antibody using two different agarose immunoprecipitation kits (Pierce Direct IP kit, Cat#26148 and Millipore Catch and Release 2.0, Cat# 17-500). The Millipore resin precipitates protein-antibody complexes directly onto the resin by binding to the antibody. The approach used in the Pierce kit is to first cross-link antibody to the resin before mixing the antibody-resin mixture with the extract. Immunoprecipitations using both kits were performed overnight at 4°C with 200 μg of protein from the nuclease-sensitive fraction of extracts performed with the Nuclear-Co-IP kit. Anti-rabbit IgG immunoprecipitations were also performed to assess the non-specific retention of protein in the agarose-antibody system.

3.2.2.4 *CSK Extractions*

CSK buffer extractions were performed using cells 6 to 144 h after induction of differentiation. Both stepwise salt extractions with increasing concentrations of (0.1 to 2 M) NaCl, and single salt extractions (0.5 M NaCl) followed by digestion of DNA before re-extraction (0.5 M NaCl) were performed.

Extractions were performed on both control and differentiating cultures at several time-points after induction of differentiation. Where DNase digestion was performed between extractions, identical extractions were carried out in parallel

without DNase. This was intended to control for proteins released non-specifically during the incubation period.

PPARγ relative abundance in extracts and pellets was determined by western blot. PPARγ localisation after DNase extractions was also investigated by indirect immunofluorescence in fixed cells with maintained 3D structure imaged on a single z-plane using confocal microscopy. This was intended to allow visual assessment of PPARγ localisation.

3.2.2.5 *PCR and siRNA to PPARγ2 and PPARγ1/2*

cDNA libraries were constructed using RNA isolated from control and differentiation induced cells at 24 h and 144 h after treatment. PCR was performed using forward primers targeting the first coding exon of PPARγ1 or PPARγ2, with reverse primer in the final coding exon common to PPARγ1 and PPARγ2. Reverse transcriptase negative cDNA samples were also subjected to PCR to control for presence of contaminating DNA. The presence of full-length PPARγ1 and PPARγ2 mRNA was assessed using RT-PCR targeting the extreme 5' and 3' exons of each mRNA. PCR products were separated on 0.75% agarose gels and imaged on a UV-imaging system (GeneGenius, Syngene).

siRNA was designed to target PPARγ2 and PPARγ1/2 and transfected into cells before induction of differentiation (Materials and Methods 2.4.6). Expression of PPARγ and differentiation markers was assessed by western blot of whole cell lysates in transfected NHU cells treated for 24 or 72 h with TZ and PD153035 or 0.1% DMSO vehicle control. Transfection with siRNA targeting the non-human luciferase gene was included alongside anti-PPARγ siRNA to control for any none gene-specific effects the transfection procedure may have had on the NHU cells.

3.2.2.6 *SUMOylation inhibition*

PPARγ activity has been demonstrated to be regulated by SUMOylation in multiple cell types (119, 128-132). To investigate if any PPARγ-reactive bands on western blot were potentially SUMOylated, NHU cells were treated with anacardic acid, a known inhibitor of SUMOylation (143). NHU cells were cultured as described in Materials and Methods 2.3 to 70% confluence and medium changed to include 1, 5 or 10 µM anacardic acid or 0.1% DMSO as vehicle control. After 8

hours, cells were harvested in CSK buffer with added 2% SDS and sonicated as described in Materials and Methods 2.4.

Protein content of lysates was quantitated using the BCA assay and 20 µg subjected to western blot as described in Materials and Methods 2.4. Abundance of PPARγ isoforms and histone H3 acetylated at lysine 4 (H3K4me3) was assessed by densitometry using the Odyssey software (Li-CoR version, 1.2.15) as described in Materials and Methods 2.4.4. As anacardic acid also inhibits histone acetyltransferase activity, the relative levels of H3K4me3 were assessed to determine whether anacardic acid was having an effect on this in NHU cells. Relative abundance of bands on western blots were normalised to beta-actin as a control to account for potential differential overall protein content of lanes.

3.3 Results

3.3.1 Nuclear Extractions and Immunoprecipitations

Before proceeding with investigation of PPARγ isoform behaviour, successful induction of differentiation was demonstrated by detection of upregulation of the differentiation marker FOXA1 in whole cell lysates prepared from parallel cultures to those used for nuclear extraction and immunoprecipitation (Figure 3.3.1).
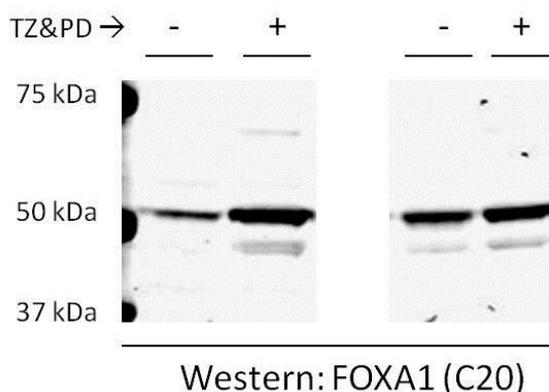
Figure 3.3.1. FOXA1 upregulation in response to induction of differentiation. 2 NHU cell lines used for nuclear extractions and immunoprecipitations were subjected to differentiation-inducing (+) or control (-) culture conditions for 24 h before harvesting. 20 µg whole cell lysates were subjected to SDS-PAGE and western blot, and membranes labelled using anti-FOXA1 (C20 clone, Santa Cruz) and relevant secondary antibodies. Cell line Y933 (left blot) and Y1086 (right blot) showed apparent increase in amounts of FOXA1 labelling in response to differentiation.

NHU cells were fractionated into cytoplasmic, nuclease-sensitive and non-solubilised fractions using the Active Motif Nuclear Co-IP kit. Western blots of cytoplasmic extracts obtained after hypotonic lysis showed weak presence of 50 kDa PPARγ in control extracts, the relative abundance of which was reduced in differentiated extracts (Figure 3.3.2). In both differentiation-induced and control cell extracts, the majority of the PPARγ at 50 and 52 kDa was in the nuclear fraction extracted after nuclease fragmentation, with some PPARγ at this molecular weight remaining in the pellet of material not solubilised by nuclease fragmentation. A third distinct PPARγ was clearly evident above the 50 and 52 kDa isoforms in the non-extracted pellets, with a faint band visible at the same height in the nuclear extracts. Other prominent bands at around 42, 32, 30 and 25

kDa were present, of which the majority of the protein was in the cytoplasmic and nuclear extract fractions, apart from the 42 kDa fraction which was of similar abundance in the nuclear extracts and pellets. These smaller bands were variable in abundance between cell lines, as was the relative abundance of the isoform of PPARγ at 52 kDa (Appendix Figures 7.1.2.1 and 7.1.2.3).
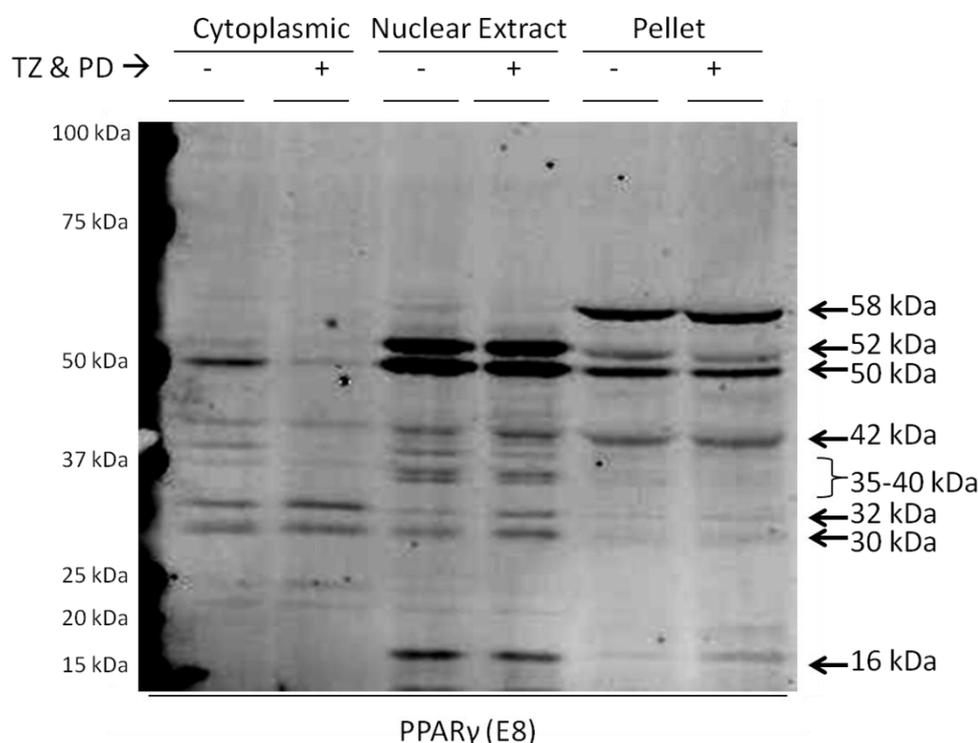
Figure 3.3.2. Fractionation of NHU cells using Nuclear Co-IP kit with (+) and without (-) induction of differentiation. NHU cells contain multiple PPARγ-reactive proteins, which appear similarly distributed in differentiation-induced and control cells apart from a small amount of PPARγ at 50 kDa in the cytoplasmic fraction. Proteins extracted from hypotonically lysed cells are the "cytoplasmic" fraction. Nuclear proteins isolated after nuclease digestion and low-salt extraction are labelled "extract", and remaining non-extracted proteins are labelled "pellet. The doublet at 50 and 52 kDa and the higher band at 58 kDa resemble those previously observed in extracts from NHU cells (31, 122). Cell line This blot is representative of a total of three separate extractions performed on three separate NHU donor lines, although the detection of any PPARγ in the cytoplasmic fraction was inconsistent (Appendix 7.1.2.1 and 7.1.2.3).

Immunoprecipitations were then performed of the nuclease-treated 'nuclear' extracts using a different anti-PPARγ antibody in order to provide additional evidence that the isoforms of PPARγ observed by western blot were not products

of non-specific interaction of the antibody used to label blot membranes. Western blots of bound and non-bound fractions from immunoprecipitations showed that most PPARγ isoforms present in nuclear extracts were present in the bound fraction of immunoprecipitations (Figure 3.3.3). An additional band at 64 kDa was visible in the bound fraction which was not detected in the input nuclear extracts. A proportion of the PPARγ at 50 and 52 kDa did not bind to the affinity resin, which may be PPARγ which is part of a complex wherein the antibody binding epitope is masked, or the proportion of PPARγ that was in excess of the capacity of the antibody-resin mixture included in the IP. Further experiments were performed using differing amounts of antibody which appeared to give similar retention of PPARγ, suggesting that the capacity of the antibody had not been reached (Appendix Figure 7.1.2.1). As the resin was used at a fraction of its 10 µg antibody capacity, this suggested that the resin was also not being saturated. This strongly suggests that the binding capacity of the system was not reached, and that in some complexes PPARγ epitopes were masked.

IP: PPARγ (81B8)
Western: PPARγ E8
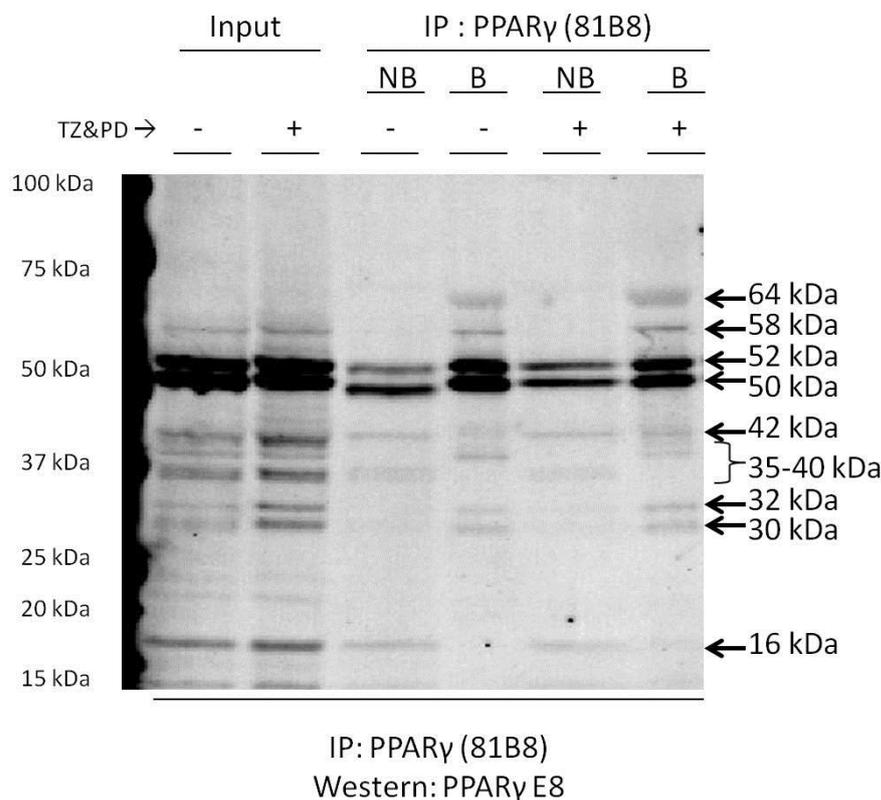
Figure 3.3.3. Immunoprecipitation of multiple PPARγ isoforms from nuclear Co-IP extracts. NHU cell line Y933 treated with TZ&PD for 24 hours was subjected to extraction with the Nuclear Complex Co-IP kit. 200 µg nuclear extracts were immunoprecipitated against PPARγ (81B8 clone) using either the Millipore Catch and Release or Pierce Direct IP kits. 20 µg of extract (input) was compared by western blot with 10 % of non-bound (NB) and 50 % of bound (B) fractions from immunoprecipitations. Most of the PPARγ-reactive bands visible in the input fraction were present to some degree in the bound fraction from the immunoprecipitation. The exceptions to this were the bands at 35 kDa and 17 kDa which did not precipitate, and the band at 64 kDa which did precipitate but was not detected in the input.

To confirm the specificity of PPARγ immunoprecipitation, the Millipore catch and release immunoprecipitation system was compared to that of another supplier: the Direct IP kit from Pierce. Similar results were seen, whereby most isoforms of PPARγ observed in nuclear extracts were successfully immunoprecipitated (Figure 3.3.4). Control anti-rabbit IgG immunoprecipitations showed some background levels of PPARγ in the Millipore, but not in Pierce immunoprecipitation systems. Further experiments using resin only immunoprecipitations showed that some background was inherent to the Millipore system, but that the Pierce system was almost completely devoid of non-specific retention (Appendix Figure 7.1.3). This

showed that the isoforms of PPARγ isolated in the both the Millipore and Pierce system were genuine, but that the Millipore system suffered from some background non-specific pull down.
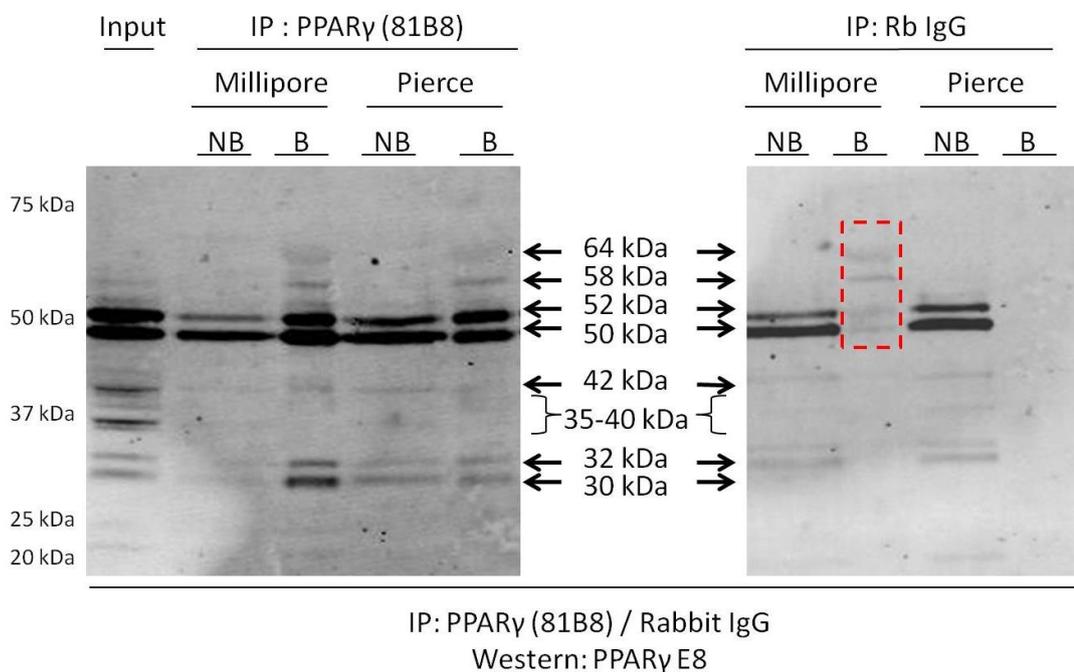
Figure 3.3.4. IP to PPARγ from nuclear extracts using Millipore and Pierce immunoprecipitation kits. IP with anti-PPARγ antibody (81B8), incubated with Nuclear Co-IP extracts from cell line Y933 treated with TZ&PD for 24 h (Y933). Both immunoprecipitation systems resulted in capture of multiple PPARγ-reactive bands observed in the nuclear extracts (input). Some non-specific retention was detected in the Millipore system when control precipitations were performed with anti-Rabbit IgG (red dashed box). No background PPARγ was detected with control immunoprecipitations using the Pierce system.

3.3.2 Sequential CSK Extractions

To further investigate the solubility of PPARγ isoforms during differentiation, CSK-NaCl extractions were used to sequentially extract protein from cells using stepwise increasing NaCl concentrations, with the extracts visualised using western blotting with the E8 clone anti-PPARγ antibody. Most PPARγ-reactive bands at 50 kDa and below which were present in the cells at 6 h were extracted by 0.5 M NaCl (Figure 3.3.5). Most of the 58 kDa isoform remained in the insoluble fraction, along with traces of the isoforms at 50 and 42 kDa. Most strikingly, there was no appreciable difference in extraction profile between

control and differentiation-induced cells. Histone H3 was included as a control to observe the concentration of NaCl which began to disrupt histone occupancy in the nucleus and to confirm the stringency of the extraction. pERK was almost absent from whole cell extracts in differentiation-induced cells indicating successful blocking of EGFR signalling by PD153035.
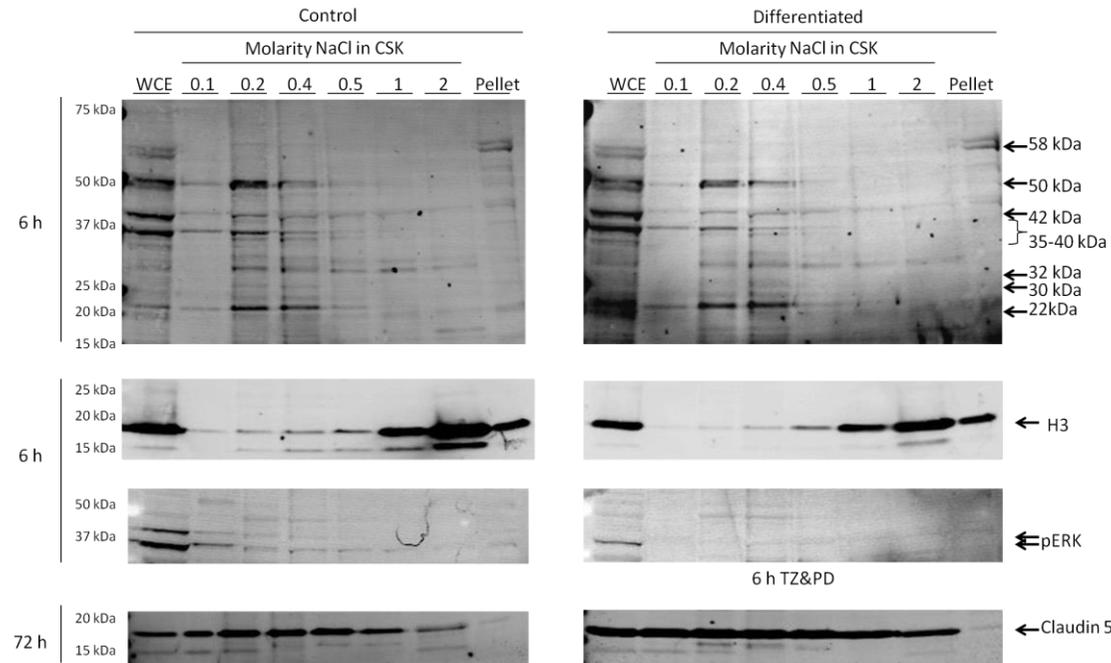
Figure 3.3.5. PPARγ and pERK in sequential CSK salt extracts of cell lysates 6 hours after induction of differentiation. Control and differentiation-induced cells were sequentially extracted in CSK buffer with 0.1% Triton-X100 and NaCl from 0.1 to 2 M. Material remaining in pellets after extraction was solubilised in 2% SDS lysis buffer. Parallel cultures were used for whole cell extracts (WCE). PPARγ extraction was not affected by induction of differentiation. Most Histone H3 did not extract until addition of 1 M NaCl, showing that chromatin integrity was intact up to at least 0.5 M NaCl. H3 also serves as a loading control, showing comparable protein amounts between equivalent extracts of control and differentiated cells. pERK was almost absent in cells treated with TZ & PD, demonstrating EGFR inhibition. Claudin 5 was more abundant in parallel cultures at 72 h time point in TZ&PD treated cells, showing successful induction of differentiation over the time-course of treatment. NHU cell line Y1152.

The sequential salt extraction procedure was also carried out on differentiated and control lysates of cells 144 h after treatment (Figure 3.3.6). PPARγ immunoreactive bands ≤52 kDa were more resistant to NaCl extraction than at 6 h, yet extraction profiles were similar for both differentiated and control cells. Most notable was the increased relative abundance of the PPARγ band at 52 kDa in differentiated and control cells, which was both more abundant and resistant to extraction in differentiated than control cells. The 37 kDa band was slightly more abundant in the control than differentiated cells in the 1 M fraction. Most of the PPARγ-reactive bands appeared to be represented in the pellet, although the resolution of the bands was poor, likely due to the high salt content, as was the reproducibility of the observation of 58 kDa isoforms of PPARγ in whole cell extracts. These observations were repeated in later experiments (Chapter 4), where changing the buffer from SDS to LDS for pellet solubilisation greatly increased the blots of CSK extraction pellets – and the 42 and 58 kDa PPARγ were consistently observed in the non-extracted fraction.
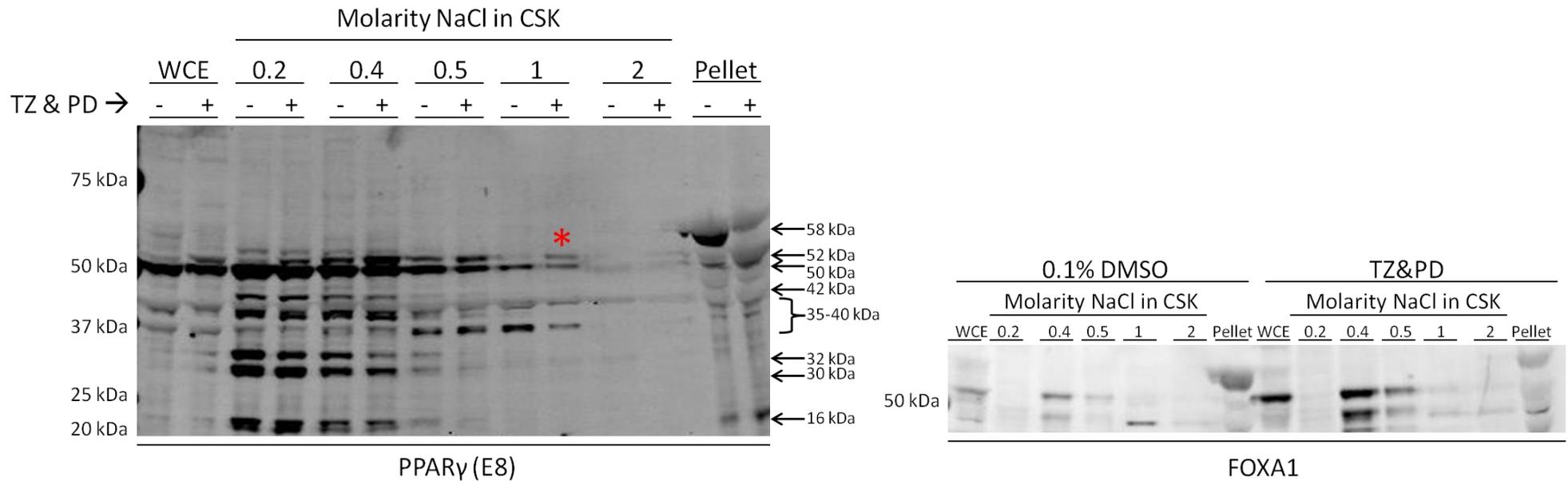
Figure 3.3.6. PPARγ localisation in sequential CSK salt extracts 6 days after induction of differentiation. (Top blot) undifferentiated (TZ&PD -) and differentiation-induced (TZ&PD +) NHU cells were treated for 144 h and harvested before re-suspension in 100 µL CSK buffer with 0.1% Triton-X100 containing sequential NaCl concentrations from 0.1 to 2 M. Extracts were mixed with 4X LDS and reducing agents before blotting against PPARγ. Extraction profiles show that the 52 kDa band was slightly more resistant to extraction in differentiation-induced cells, remaining up to 1 M. PPARγ-reactive bands <50 kDa all showed profiles indistinguishable between differentiated and control cells. Bands in pellet lanes are poor quality but show the retention of the 58 kDa PPARγ-reactive band. Bottom blot: upregulation of FOXA1, which resides in the 0.4-0.5 M NaCl fraction, can be seen in differentiated cells. NHU cell line Y1052.

3.3.3 CSK-DNase Extractions – Western Blots

To assess the sensitivity of NaCl-resistant PPARγ isoforms to nuclease digestion in the CSK extraction system, NHU cells 24 and 144 h after differentiation were subjected to extraction in CSK buffer with 0.5 M NaCl before and after incubation with DNaseI. In this setup, whole cell extracts and insoluble pellets remaining after CSK extractions were lysed in CSK buffer with added LDS as for sequential CSK extracts (Materials and Methods 3.6.1), rather than using 2% SDS as previously. This was because it was observed that use of CSK buffer greatly increased the quality of the appearance of bands in the pellets. Whole cell extracts were included to allow equal loading.

In whole cell extracts lysed in CSK buffer, the PPARγ-reactive band at 58 kDa was clearly present in control and differentiated cells (Figure 3.3.7) after 24 and 144 h. The main change in extraction profile of PPARγ over time was the increased relative abundance of PPARγ at 52 kDa in the 0.5 M pre-DNase extracts in the 144 h sample. DNase treated cells released the same amount of PPARγ at 50 and 52 kDa in both time-points. The 58 kDa isoform was not extracted at either time-point.
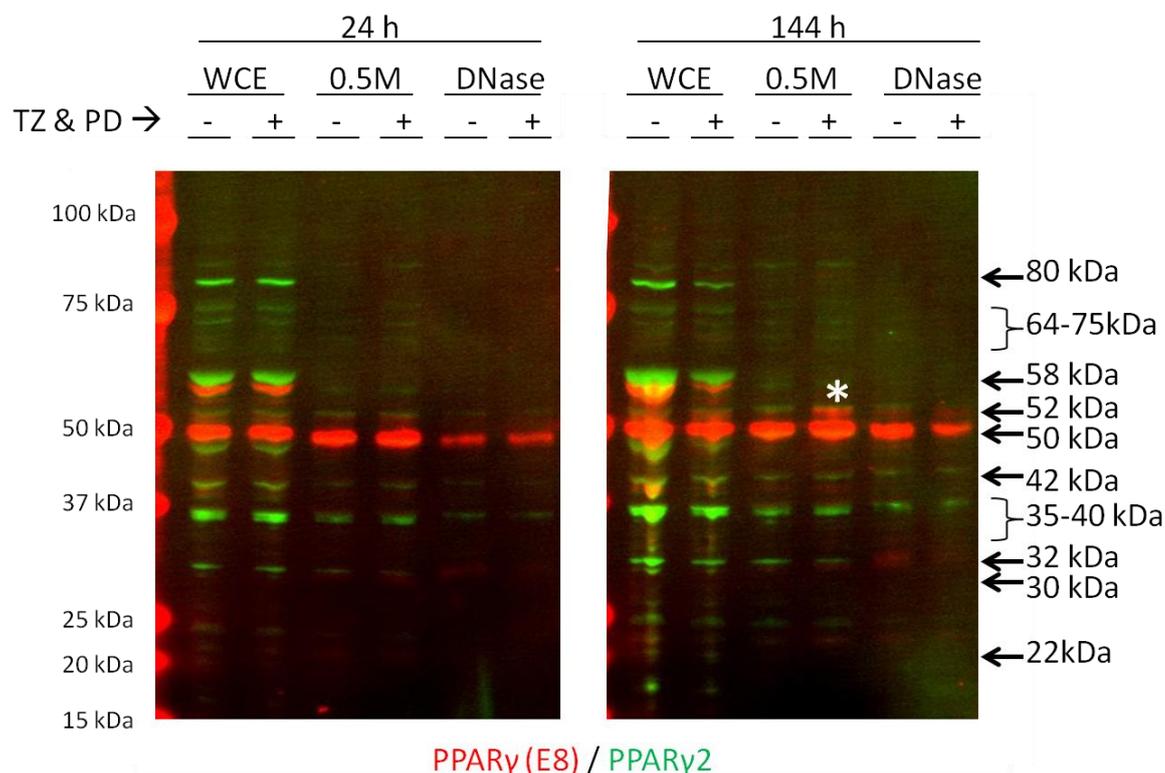
Figure 3.3.7. Western blot of CSK-DNase extracts at 24 and 144 h. Western blots of whole cell lysates (WCE), pre-DNase extracts (0.5 M) and post-DNase extracts (DNase) immunolabelled for PPARγ. Similar extraction profiles are seen for PPARγ in control (TZ&PD -) and differentiated (TZ&PD+) cells at 24 h. 52 kDa PPARγ was relatively more abundant in 0.5 M pre-DNase extracts taken from differentiated cells after 144 h than at 24 h (asterisk). Multiple 64-75 and an 80 kDa PPARγ2-reactive isoforms were also visible in both WCE at 144 h. NHU line Y1120. Western blot labelled with E8 clone anti-PPARγ antibody (red) and anti-PPARγ2 antibody.

To test whether the PPARγ released after DNase treatment was specific to the addition of the DNase and not just a result of incubation conditions, extractions were carried out on differentiated and control cells at 6, 24 and 144 hours with and without DNase, and the material remaining behind in the pellets was blotted. As with the whole cell extracts, solubilising the pellets in CSK buffer rather than SDS showed a consistent presence of the 58 kDa isoform of PPARγ. Similar to the Nuclear-Co-IP extractions, pellets remaining after salt extractions showed higher relative abundance of bands at 58 and 42 kDa compared to 50 and 52 kDa (Figure 3.3.8).

The relative abundance of PPARγ isoforms remaining behind in pellets did not appear to change with the addition of DNase, suggesting that PPARγ isoforms

resistant to the initial extraction with 0.5 M NaCl were not more susceptible to extraction after the addition of DNase. PPARγ at 50 kDa was more abundant in pellets from differentiated cells at early time-points (6 h) than in those from non-differentiated cells. However, this trend reversed at 24 h, where there was less PPARγ at 50 kDa relative to 58 and 42 kDa in differentiation-induced cells. As DNase extractions should contain either the same or less bands than control extractions due to the digestion of DNA, it was noted that at 6 h and 24 h there seemed to be greater abundance of the 50 kDa PPARγ in DNase extracted cells. As the PPARγ at these time points was so low in abundance, it is likely that the same difficulties experienced solubilising the pellets as before may have hampered the reproducibility of low abundance bands. Immunoblots of DNase extracts showed that histone H3 was extracted in much greater abundance when cells were incubated with DNase (Figure 3.3.9), showing that the enzyme was effective.
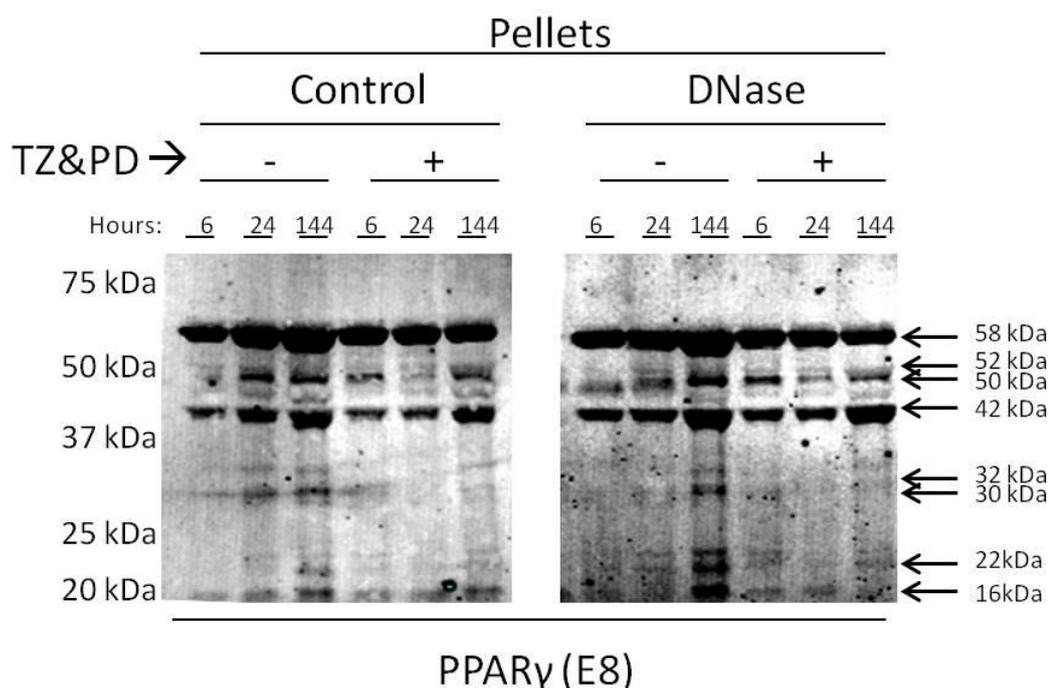
Figure 3.3.8. DNase extracted pellets. Western blot showing presence of PPARγ-reactive bands in DNase and control extracted pellets. Cells were cultured for 6-144 h with TZ&PD (+) or DMSO control (-). Cells were then pre-extracted with 0.5 M NaCl and incubated in DNaseI buffer with (DNaseI) or without (Control) enzyme before re-extracting with 0.5 M NaCl. The remaining pellets were solubilised by sonication in CSK buffer and LDS, prior to SDS-PAGE separations and western blotting. All time-points show presence of 42, 50 and 58 kDa PPARγ in all pellets.
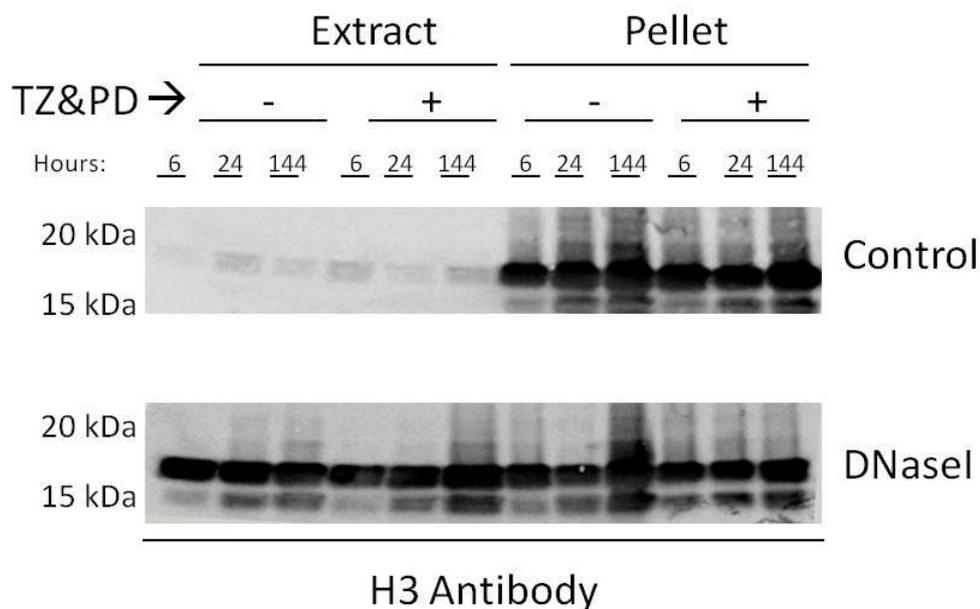
Figure 3.3.9. Histones in DNase extracts and pellets. DNase extracted cell pellets show extraction of histones after addition of DNaseI, indicating enzyme was active. Cells were cultured for 6-144 h with TZ&PD (+) or DMSO control (-). Cells were then pre-extracted with 0.5 M NaCl and incubated in DNaseI buffer with (DNaseI) or without (Control) enzyme before re-extracting with 0.5 M NaCl prior to analysis of the extract or pellet using SDS-PAGE and western blotting.

3.3.4 DNase extractions – Confocal Immunofluorescence

Indirect immunofluorescence was performed before and after the different stages of CSK-DNase extraction (with 0.5 M NaCl) in order to assess if any differences in PPARγ cellular localisation and resistance to extraction could be identified between control and differentiated NHU cells. Cells were kept hydrated after fixation in a 1:1 mixture of methanol and acetone to allow proteins to maintain a more native three-dimensional conformation and enable localisation to be assessed by confocal microscopy (Materials and Methods 2.6.3).

Differentiated and control NHU cells at 24 and 144 h showed some evidence of nuclear foci when labelled with anti-PPARγ antibodies in cells before extraction. Foci were more prominent following CSK-NaCl extraction up to 0.5 M NaCl and subsequent CSK-DNase treatment in cells at 24 h, and of similar appearance at 144 h before and after extraction. PPARγ labelling on the outer and inner nuclear periphery was much more distinct in cells after DNase digestion, especially in cells at 144 h. The distribution patterns of PPARγ were similar between TZ&PD treated and DMSO control cells at each time point. Similar results were obtained from independent cell lines (Figures 3.3.10 A and B).
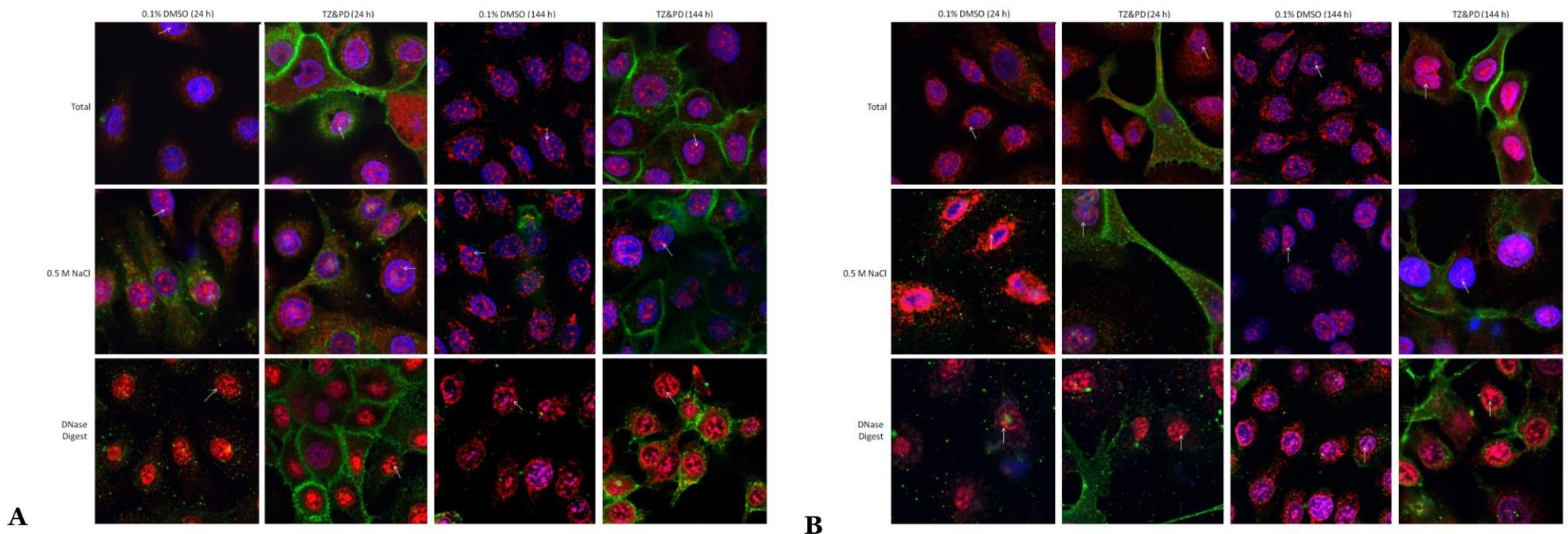
Figure 3.3.10 A, B. Immunofluorescence in two NHU donor lines showing nuclear localisation and resistance to extraction of PPARγ (red) in NHU cells with and without induction of differentiation. NHU cells were treated with control or differentiation-inducing agents as annotated. Cells were either not extracted (Total), extracted for 5 minutes using CSK (0.5 M NaCl), or CSK extracted followed by DNase digestion and repeat CSK extraction (DNase Digest). Extraction did not disrupt foci (white arrows) of PPARγ in the nucleus, even after treatment of cells with DNaseI enzyme which degraded DNA (blue). Differentiation marker aquaporin 3 (green) relocates to the membrane in differentiation-induced NHU cells (144), confirming differentiation induction. Cells were fixed in 1:1 methanol and acetone for 30 seconds and buffer-exchanged into PBS before incubating with primary antibodies overnight followed by relevant secondary antibodies for 1 h and Hoecsht 33258 staining. Images taken on Zeiss LSM-710 confocal microscope at 60 X magnification. 3.3.10 A = NHU donor line Y1185, 3.3.10 B = NHU donor line Y1085.

3.3.5 PPARγ mRNA Expression in Differentiation

Expression of PPARγ transcript was investigated by RT-PCR to assess the extent to which PPARγ 2 was expressed over the course of differentiation of NHU cells.

PPARγ2 expression was investigated using a forward PCR primer in the upstream exon possessed by PPARγ2 (exon "B", Figure 3.3.11). PPARγ4 and PPARγ5 transcripts also contain this exon and could be amplified alongside PPARγ2, although these would be 200 bp longer than the expected PPARγ2 amplicon. PCR targeting PPARγ exons 1-6 was performed in parallel to gauge overall levels of PPARγ transcript and as an internal control.
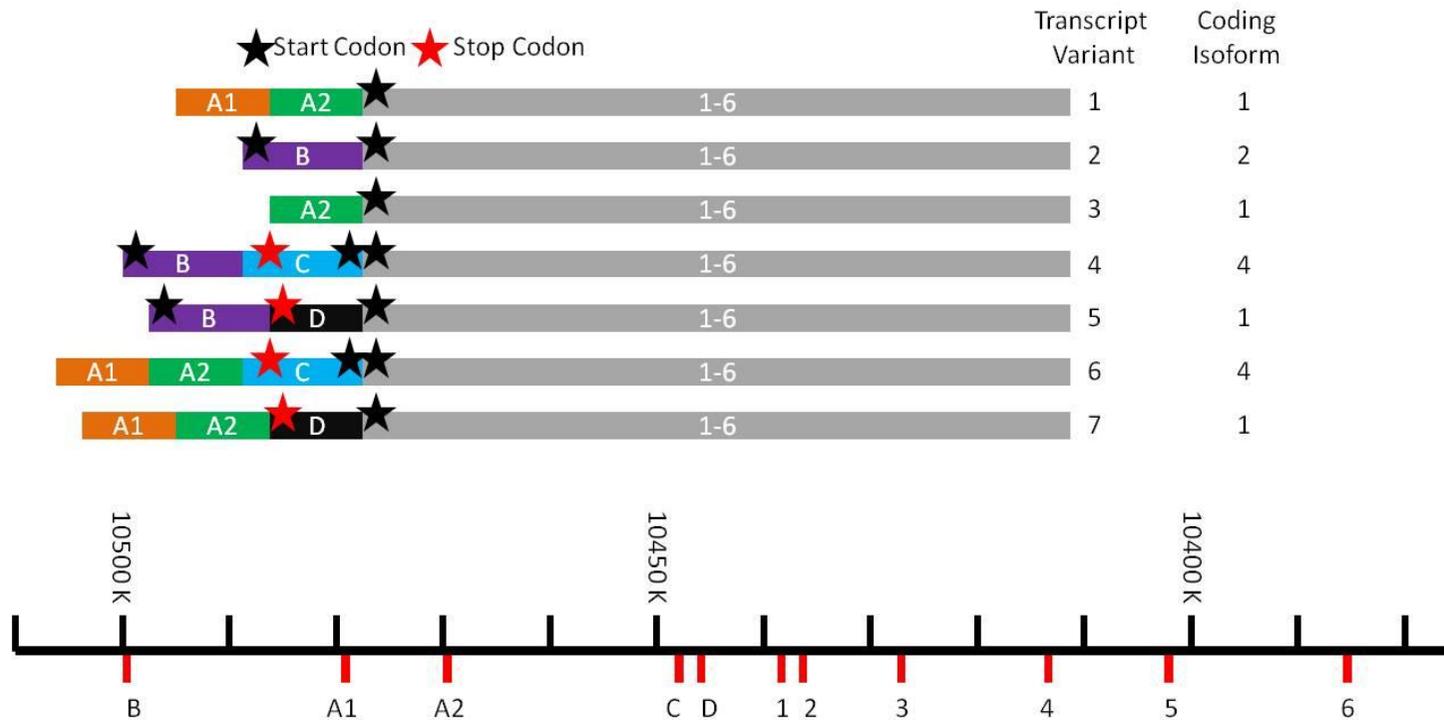
Figure 3.3.11. PPARγ transcript variants. Exons 1-6 are common to all PPARγ isoforms, with variants gaining exons from upstream regions. PPARγ transcripts 1, 3, 5 and 7 all result in PPARγ1 protein due to absence of a start codon in exons A1 and A2 or the stop codon in exon D. Transcripts 4 and 6 result in PPARγ4 protein due to exon C containing both a stop and start codon. Adapted from (83).

PPARγ expression increased over time in both differentiated and control cells, but had greater relative abundance in differentiated cells at each time point (Figure 3.3.12). PCR products from PPARγ exons 1-6 were detected at all time-points, with possible splice variants visible underneath the expected full-length PCR product. PPARγ2-specific PCR product was much more weakly expressed, needing 40 cycles of PCR to give a product in control and differentiated cells at 144 h (Top right, Figure 3.3.12). The presence of shorter than expected PCR products when amplifying from cDNA using primers targeting exon 1 and exon 6 of PPARγ1 was repeated using cDNA from an independent NHU cell line (Figure 3.3.13). The presence of these bands suggests that in addition to the full-length canonical transcript, there are PPARγ transcripts variants present that amplify when using primers targeting the 5'- and 3' terminal exons of canonical PPARγ1 mRNA.
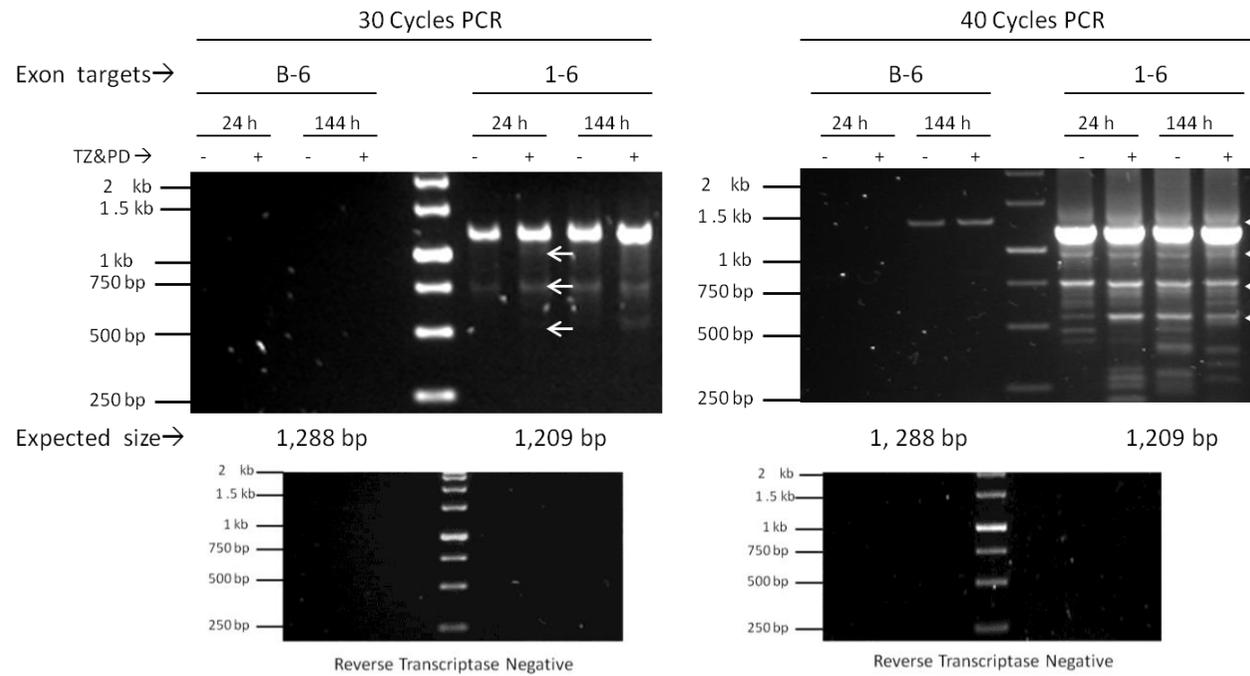
Figure 3.3.12. Full length PPARγ mRNA PCR. PCR product spanning exons 1-6 was present at the expected size of 1,209 bp after 30 cycles (top left), along with smaller products visible at around 1 kb, 750 bp and 600 bp (white arrows). Use of a forward primer in PPARγ2 specific exon "B" required 40 cycles of amplification before a product was detected, which was only present in 144 h samples (top right). cDNA was synthesised from 1 μg total RNA from cells treated with differentiation or control agents for 24 or 144 hours, diluted to 20 μL and 1 μL used per PCR reaction. PCR from cDNA syntheses without reverse transcriptase (bottom left, bottom right) produced no amplicons. Primers; exon "B" forward: 5'-TCCTTCACTGATACACTGTCTGC-3', exon 1 forward: 5'-ACTTTGGGATCAGCTCCGTG-3', exon 6 reverse 5'-GGGCTTGTAGCAGGTTGTCT-3'. All primers had a T_m of 60°C (-/+ 1°C). Thermal profile: 95°C 2 min then cycles of 95°C for 30 s, 60°C for 30 s and 72°C for 2 min followed by final 72°C for 5 minutes. GoTaq (Promega) PCR kit used for amplification.

Expected size: 1,209 bp
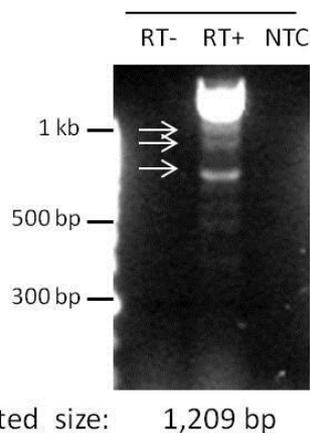
Figure 3.3.13. PPARγ full length PCR independent cell line. PCR targeting PPARγ exons 1-6 shows amplicons present below expected length, with two around 1 kb and one around 750 bp (arrows). This suggests the presence of splice variants missing exons totalling around 200 and 450 bp in length. RNA harvested from cell line Y1085 after culturing to 70% confluence and treating with 0.1% DMSO for 24 h. Primers and PCR as for Figure 3.3.10, 34 cycles PCR. Reverse transcriptase negative cDNA transcription (RT-) and water only no template control (NTC) included as negative controls for amplification of background DNA.

3.3.6 siRNA to PPARγ

siRNAs were generated based on shRNA targets of Strand *et al* (46). PPARγ was targeted in order to assess the effects of PPARγ expression levels and the ability of NHU cells to differentiate. PPARγ2 was targeted using siRNA to the first coding exon of PPARγ2. siRNA was also used which targeted the first coding exon of PPARγ1. As this would also bind the PPARγ2 transcript, it was referred to as PPARγ1/2.

NHU cells exposed to control siRNA (luciferase) and induced to differentiate had >5-fold increased expression of PPARγ at 52 kDa, as well as the differentiation markers cytokeratin 13 (CK13) and claudin 4 (CLDN4) after 72 h differentiation (Figure 3.3.14, Figure 3.3.15). Upregulation of all of these proteins was attenuated by the addition of PPARγ2 siRNA, whereas PPARγ1/2 siRNA slightly increased the relative levels of differentiation markers.

Compared to luciferase siRNA controls, PPARγ2 siRNA reduced expression of PPARγ2 by >50% at every time-point during differentiation. There was a 25% reduction in upregulation of CK13 expression in cells exposed to PPARγ2 siRNA,

but not in response to PPARγ1/2 siRNA. CLDN4 upregulation was reduced >60% in differentiated cells at 72 h when exposed to PPARγ2 siRNA.

In both differentiation-induced and control NHU cells at both time points, PPARγ1/2 siRNA increased the amount of CLDN4 expression relative to siRNA treated cells. A similar effect was seen for CK13 at 72 h.

The PPARγ isoforms at 42, 50 and 58 kDa were only slightly downregulated by the PPARγ1/2 siRNA in differentiation-induced cells at 72 h (Figure 3.3.14). The upregulation of the 42, 50, 52 and 58kDa PPARγ bands seen in differentiation-induced cells exposed to control siRNA was stopped by PPARγ2 siRNA. As evidenced by PCR, PPARγ1 transcript was very abundant relative to the amount of PPARγ2 transcript detected. It is therefore likely that siRNA to PPARγ1/2 was much less efficient than that targeting PPARγ2.
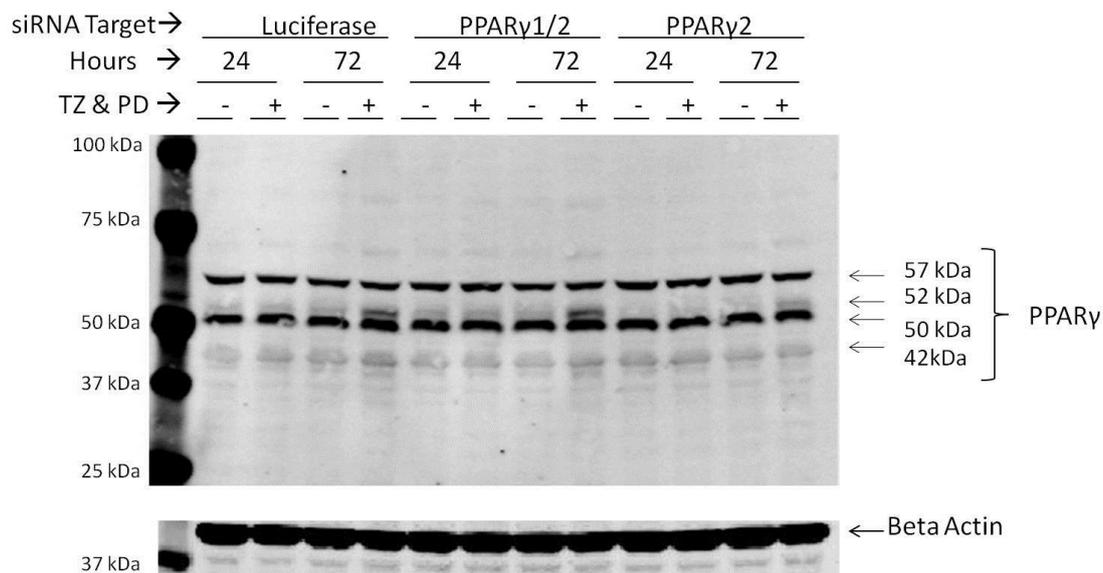
Densitometry Measurements of Western Blot
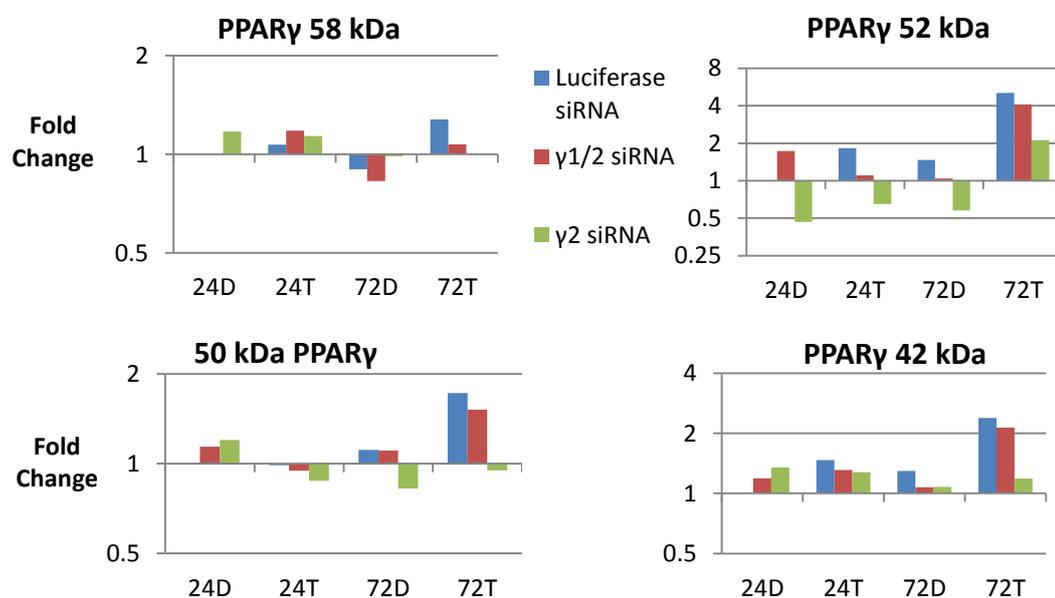
Figure 3.3.14. Effects of PPARγ siRNA on expression of PPARγ isoforms.
In densitometry graphs, 24D, 72D and 24T, 72T designate control (D=DMSO) and differentiated (T=TZPD) cells at 24 and 72 h. Upregulation of PPARγ isoforms observed during differentiation in the presence of luciferase control siRNA was attenuated by both PPARγ1/2 and PPARγ2 siRNA at 72 h, with the greatest effect from PPARγ2 siRNA. NHU cells were transfected with 200 mm siRNA for 7 hours before addition of control (0.1% DMSO) or differentiation-inducing (TZ & PD) agents. Cells were harvested in CSK buffer with added 2% SDS, and reducing agents omitted to enable quantitation by BCA assay and loading of 20 µg per lane. All densitometry measurement areas were of the same size for each measured band, with background correction taken as the median pixel intensity in a region 2 pixels wide adjacent to the measured region.

Densitometry Measurements of Western Blot
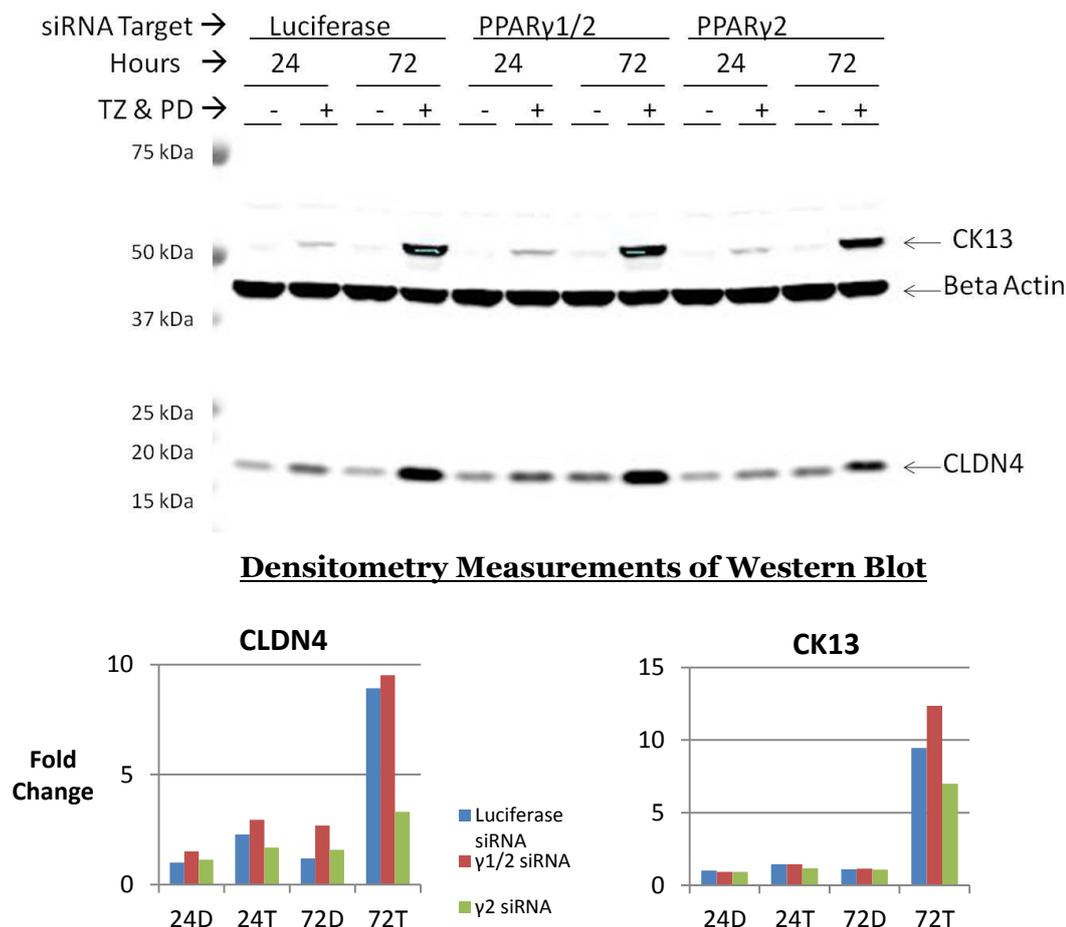
Figure 3.3.15. Effects of PPARγ siRNA on differentiation marker expression. 24D, 72D and 24T, 72T designate control (D=DMSO) and differentiated (T=TZPD) cells at 24 and 72 h respectively. PPARγ2 siRNA attenuated upregulation of cytokeratin 13 (CK13) and claudin 4 (CLDN4) at 72 h differentiation as compared to luciferase siRNA. CLDN4 upregulation was reduced by PPARγ2 siRNA as compared to luciferase siRNA, but slightly increased in PPARγ1/2 siRNA in all time-points. CK13 upregulation at 72 h in response to differentiation was also attenuated by PPARγ2 siRNA as compared to luciferase control, but an increase was observed in differentiated cells exposed to PPARγ1/2 siRNA. Blots were from replicate aliquots of the same samples in Figure 3.3.14, and thus prepared and analysed identically.

3.3.7 Anacardic Acid

To investigate the possibility that any PPARγ isoforms present in NHU cells are a post-translationally modified isoform of PPARγ, relative levels of PPARγ-reactive bands were monitored in response to treatment of cells with 1-10 μM anacardic acid. Anacardic acid can prevent the formation of the SUMOylation precursor complexes, preventing ligation of SUMO to target proteins (143). NHU cells were cultured to 70% confluence and medium changed to include 1, 5 or 10 μM anacardic acid or 0.1% DMSO vehicle control. Cells were harvested after 8 hours in CSK buffer with added 2% SDS, and protein quantitated by BCA assay. Western blots were loaded with 20 μg immunolabelled with antibodies to PPARγ (E8 clone), PPARγ2, histone H acetylated at lysine 4 (H3K4Ac) and beta actin.

 Anacardic acid is also a known histone acetyltransferase inhibitor (145). As such the H3K4Ac abundance was investigated to monitor the effect of anacardic acid. Treatment with 10 μM anacardic acid slightly reduced relative amounts of 58 kDa PPARγ and similarly slightly increased abundance of 50 kDa PPARγ. Blots were also labelled using an antibody targeting PPARγ2 (Figure 3.3.16).
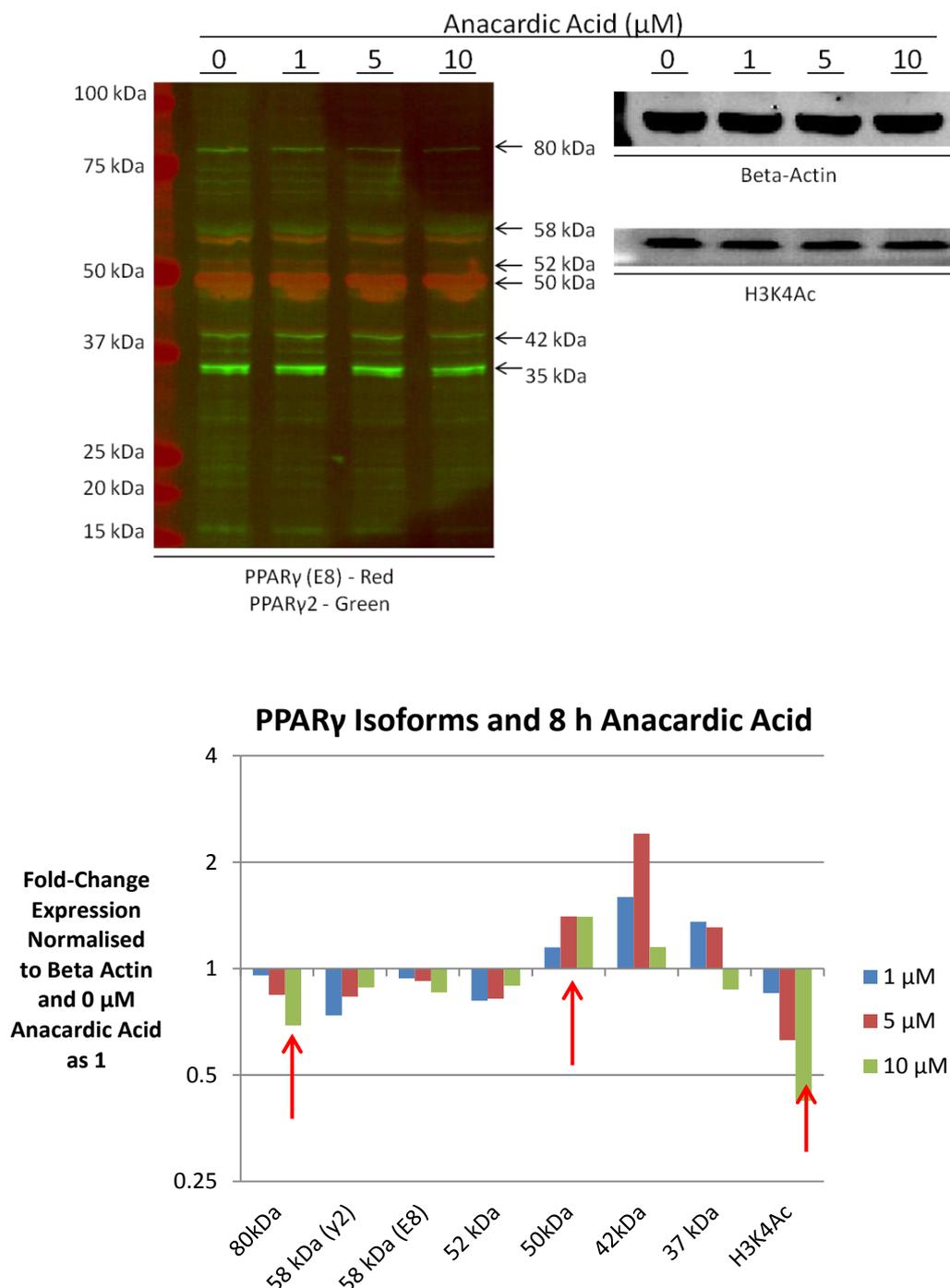
Figure 3.3.16. Modulation of PPARγ isoform abundance in undifferentiated NHU cells treated with 0, 1, 5 or 10 µM anacardic acid for 8 h. Lysates blotted for PPARγ (red), PPARγ2 (green) (left blot), H3K4Ac (right bottom blot) and beta-actin (right top blot). Graph of densitometry measurements normalised to beta-actin shows PPARγ-E8 immunoreactive bands at 50 kDa increase and 80 kDa isoform decrease, in relative abundance in response to anacardic acid. H3K4Ac reduction was dose-dependent in response to anacardic acid.

3.3.8 Overview of Results

- The original hypothesis that PPARγ would translocate to the nucleus in response to differentiation was not fully supported, as the majority of the PPARγ at 50 and 52 kDa was located in the nuclear fraction of nuclear co-IP extracts from NHU cells cultured to at least 70% confluence, independent of induction of differentiation.

- Not all PPARγ protein was released from nuclei after nuclease digestion, in either nuclear co-IP fractionation or CSK-DNase extractions. The presence of non-extracted PPARγ-reactive proteins observed in non-extracted portion of cells after CSK extraction was corroborated by the observation of nuclear by immunofluorescence after CSK-DNase extractions.

- The two PPARγ isoforms at 50 and 52 kDa in the nuclear extract fraction from nuclear co-IP extractions corresponded with those observed in identical extracts of previous studies of NHU cells. Inspection of the non -extracted pellets revealed another, distinct isoform at 58 kDa, agreeing with other observations of three major PPARγ bands between 50 and 60 kDa in NHU cells.

- Some PPARγ isoforms in nuclear extract fractions were successfully immunoprecipitated when using anti-PPARγ antibodies, including the 50, 52, 58 kDa isoforms as well as a low abundance 64 kDa isoform.

- Most of the PPARγ-reactive bands were soluble to some degree up to 2 M NaCl in CSK extractions, although most of the 58 kDa band and a significant proportion of the 42 kDa band remained in the insoluble fraction. The extraction profiles of PPARγ between control and differentiated NHU cells were strikingly similar at all time points observed.

- PCR amplification of full-length PPARγ1 mRNA resulted in one larger than expected, and multiple smaller than expected amplicons, suggestive of the possible presence of splice variants All observed isoforms appeared to be qualitatively upregulated to a similar degree at each time-point.

- An increase in the 52 kDa isoform over time was seen across different experiments. siRNA targeting PPARγ2 dramatically attenuated the upregulation of the 52 kDa PPARγ isoform, and inhibited the upregulation of the 58, 50 and 42 kDa PPARγ isoforms as well as differentiation markers CK13 and CLDN4.

- Treatment with 10 μM anacardic acid for 8 h slightly reduced the relative abundance of 58 kDa PPARγ, whilst the 50 kDa PPARγ slightly increased in

abundance. Using an antibody targeting PPARγ2 labelled the 58 kDa isoform as well as resulting in the presence of 37 kDa and 80 kDa bands. The 80 kDa PPARγ2-reactive band and H3K4Ac decreased in abundance in a dose-dependent manner in the presence of anacardic acid.

3.4 Discussion

3.4.1 Identity of PPARγ Isoforms in NHU Cells

The current study clearly demonstrated the presence of three high abundance PPARγ-reactive bands at 50, 52 and 58 kDa in NHU cells, strongly resembling the pattern of bands observed in NHU cells by Georgopoulos *et al* (31). When Fleming (122) compared PPARγ in whole cell lysates from undifferentiated NHU cells prepared in SDS with the 3T3-L1 nuclear co-IP extracts two PPARγ-reactive bands were observed in both extracts. This led to the assumption that the two PPARγ-reactive bands in nuclear co-IP extracts from NHU cells were equivalent to the two bands observed in the NHU whole cell lysates (Appendix Figure 7.1.1).

Observations made in the current study demonstrated variable presence of the 52 kDa band in undifferentiated NHU cells and the solubilisation of the 58 kDa band in whole cell lysates prepared in SDS. It therefore seems plausible that the two PPARγ-reactive bands in NHU whole cell lysates as observed by Fleming were in fact the 50 and 58 kDa isoforms, and not the 50 and 52 kDa isoforms (Appendix Figure 7.1.2). Following this line of evidence would lead to the conclusion that the 58 kDa isoform in NHU cells is equivalent to the PPARγ2 in 3T3-L1 adipocytes. This conclusion was supported by the observation that the PPARγ at 58 kDa was reactive with an antibody targeting PPARγ2, although this antibody reacted with other isoforms. Based on this line of reasoning and the evidence presented in this study, this leads to the conclusion that PPARγ2 (58 kDa) is constitutively present in NHU cells, yet the vast majority is associated with the a nuclear sub-compartment which is resistant to extraction by salt and detergent (2 M NaCl, 0.1 % Triton-X100), as well as salt and detergent (0.5 M NaCl, 0.1 % Triton-X100) extraction after digestion of DNA.

In the current study, PPARγ2 siRNA was more effective than PPARγ1/2 at attenuating the upregulation of all PPARγ isoforms in both differentiation induced and control NHU cells, but did not significantly reduce the presence of any of the abundant PPARγ isoforms. PPARγ has been demonstrated to drive its own expression in adipogenesis (102). As such, knockdown of PPARγ protein is likely to affect expression of all isoforms through a general downregulation of PPARγ. PPARγ2 siRNA treatment of NHU cells also hampered upregulation of differentiation markers CK13 and CLDN4. That the PPARγ2 siRNA did not affect the basal expression of the PPARγ2 but did prevent its upregulation may be indicative that the protein is rapidly and stably sequestered to the non-extractable

nuclear compartment once it is translated. The presence of PPARγ2 and the attenuation of expression of differentiation markers by siRNA targeting PPARγ2 in NHU cells seems to contradict the conclusions of Strand *et al* (46), who suggested that PPARγ2 drives urogenital-derived cells towards a prostatic phenotype. From the results in this study it appears that PPARγ2 has a role to play in urothelial differentiation, but that the activity of the protein is tightly controlled.

3.4.2 Localisation of PPARγ During *in vitro* Differentiation of NHU Cells

PPARγ has been reported in the literature to be mainly nuclear, but can be cytoplasmic in mitogenically-stimulated cells as a result of binding with ERK1/2, with which it is co-exported from the nucleus (reviewed in (100)). Due to the evidence in the literature that activation of PPARγ and inhibition of ERK1/2 are essential for NHU differentiation (2, 34, 56), and observations of cytoplasmic to nuclear translocation upon EGFR inhibition (26), it was anticipated that such a ligand-inducible development of resistance to extraction would be observed during differentiation of NHU cells *in vitro*.

The evidence presented in this study demonstrated that PPARγ was present in the nucleus and cytoplasm of both non-differentiated and differentiated NHU cells, as shown by immunofluorescence of whole non-extracted cells. Cell fractionation studies using the nuclear co-IP kit showed only minor amounts of PPARγ in the presumed cytoplasmic fraction. As a large fraction of the pool of PPARγ in the cytoplasm was associated with the region around the nucleus and was not affected by 0.5 M NaCl extraction, it is likely that the nuclear co-IP kit, which utilises low salt (0.075 M NaCl) extraction to maintain protein-protein interactions, was not extracting this portion of the PPARγ. This observation explains why relatively low amounts of PPARγ were seen in the cytoplasmic fraction of nuclear co-IP kit by western blot, but was present in the immunofluorescence after extraction with 0.5 M NaCl.

The previous study which reported EGFR inhibition-dependent nuclear translocation of PPARγ in NHU cells made the observation after treating cells with PD153035 for 4 h after cells attached post-seeding (26). Other studies which have reported a mixed nuclear and cytoplasmic localisation of PPARγ (56, 95, 121, 133)

all used undifferentiated NHU cells which were cultured to 70% or near-confluence before fixation. This suggests that the translocation of PPARγ after the induction of differentiation may be due to processes affected by the confluence of the cells influencing the nuclear localisation of PPARγ, and treatment of proliferative cells with PD153035may help to speed the re-localisation to the nucleus at early time-points. This observation could help to explain why NHU cells respond variably to TZ alone, and why PD153035 is able to increase the efficiency of induction of differentiation.

3.4.3 Extraction-Resistant Isoforms of PPARγ

In this study, the difficulties in solubilising the PPARγ isoforms associated with the non-extractable fraction were observed when using SDS rather than CSK buffer. The main difference between the two methods was that the CSK method did not require chilled centrifugation to remove SDS and enable Coomassie assays. Thus, when using the SDS buffer, the variability in efficiency of solubilisation before chilled centrifugation may explain the differential observations of the 58 and 42 kDa PPARγ in the literature as well as those in this study.

Many nuclear receptors are partitioned between the CSK-NaCl labile fraction and the NaCl and DNaseI resistant nuclear matrix, and become more extensively associated with the nuclear matrix upon ligand binding (62, 134, 140, 146). This was not observed for PPARγ in NHU cells, although novel observations about the extraction-resistance and distribution of several isoforms were made. The PPARγ at 50 and 52 kDa remained mostly labile and extractable by salt with and without nuclease digestion. The 42 and 58 kDa isoforms appeared to be most associated with the insoluble fraction of CSK extractions, which is considered to be the fraction of protein associated with the structural portion of the nucleus, the nuclear matrix.

Henikoff *et al* (147) used similar salt-extraction and nuclease digestion approaches followed by next-generation sequencing of extracted DNA and assessment of genome-wide transcription and found that both salt-labile and resistant fractions were enriched in similar transcriptionally active regions of the genome. Their observations suggest that transcribed regions of the genome undergo dynamic processes, which leaves them either susceptible or resistant to

extraction depending on which complexes they are bound to. Combined with the observations that active and repressed ERα can exist in the DNase insoluble fraction (140), this suggests that no assertions can be made about the potential for transcriptional activity of PPARγ based solely on observations of resistance to extraction.

3.4.4 SUMOylation of PPARγ

Treatment of undifferentiated NHU cells with 10 µM anacardic acid slightly reduced the abundance of the isoform of PPARγ at 58 kDa, and concurrently marginally increased the relative abundance of the 50 kDa isoform.

Anacardic acid was chosen as it is an inhibitor of SUMOylation, which has been observed to be associated with PPARγ in various cell types; inhibition is achieved through inhibiting formation of the E1-SUMO intermediate, an essential precursor in the SUMOylation pathway (143). However, anacardic acid also inhibits histone acetyltransferase activity (145). As such, some effects of anacardic acid could be produced through changes to gene expression.

The most striking observation of this experiment was PPARγ2 reactivity with both the 58 kDa isoform of PPARγ and the reactivity with other proteins at 80 and 35 kDa. PPARγ-reactive bands at 80 kDa have been reported in the literature as SUMOylated PPARγ (132) and have also been observed in NHU cells (Appendix Figure 7.1.4.1). However, with this in mind, the only isoform of PPARγ which reduced in abundance in a dose-dependent manner in the same way as H3K4Ac was the 80 kDa isoform, suggesting that SUMOylated PPARγ is present in NHU cells in small amounts.

3.4.5 Potential Transcript Variants of NHU cells

The smaller isoforms of PPARγ which were observed in nuclear fractionations and CSK-extractions are potentially novel splice variants of PPARγ, breakdown or cleavage products, or a mixture. PCR targeting full-length PPARγ1 mRNA (exons 1-6) showed a general trend for increasing levels of PPARγ mRNA over time, with greater relative amounts of amplicon in differentiation-induced cell extracts. In addition to this, multiple smaller, and one larger, PCR products were visible under the expected amplicon. PPARγ2 was barely detectable by RT-PCR, yet there was

an isoform present at the expected size for PPARγ2 (58 kDa). This suggests that either the PCR was not efficient due to restrictions on primer design targeting a small exon (~84 bp) or that the protein is expressed at low levels and is a long-lived protein due to its sequestration on the nuclear matrix.

PPARγ exons 2-5 are 170, 139, 200 and 451 bp respectively. Splicing of either single exons or combinations of these exons could account for the three most abundant potential splice variants observed which were around 200, 300 and 450 bp smaller than the expected amplicon. In terms of molecular weight of the peptides they code for: exons 2-5 are 6.1, 5.6, 7.9 and 17.9 kDa respectively. Loss of some of these exons could account for some of the isoforms of PPARγ <50 kDa which remained immunoreactive with both C-terminal (E8 clone) and N-terminal (PPARγ2) antibodies, although further work would have to be done to confirm this.

Alternative explanations exist as to the identity of the smaller isoforms, although none of the evidence in this study can be used to assess these. Non-apoptotically induced caspase-cleaved isoforms of PPARγ have been reported in adipocytes at 44 kDa, where PPARγ is cleaved in the N-terminal AF1 domain (148, 149) in response to treatment with tumor necrosis factor. This isoform is then targeted to the proteasome for degradation, as assessed through its accumulation when proteasome activity was blocked (148). If this cleavage mechanism were active in NHU cells and the 42 kDa isoform is the result of this cleavage, then it would require that the association with the insoluble nuclear fraction reduces the rate at which it is degraded. C-terminally truncated isoforms of PPARγ1 have also been reported in the literature to be expressed in tumour-derived cell lines (115, 116, 150). However, the C-terminal truncations are missing the target domain of the antibody used for western blotting, and so would not have been detected. One intriguing possibility for the ~42 kDa isoform is a splicing of exon 5 (17.9 kDa) from the full length 58 kDa protein. This would result in a PPARγ protein lacking a ligand-binding domain which could act as a dominant negative repressor of PPARγ activity, perhaps dampening basal transcriptional activity of full length proteins in the absence of strong agonist PPARγ ligands.

The observed PPARγ isoforms <30 kDa would likely only contain the C-terminal ligand-binding domain. Little is known about the possible function of such fragments, other than a potential for suppression, or "squelching", of full-length receptor activity by binding to ligand and co-factor proteins. Such squelching has

been reported for C-terminally truncated non-DNA binding disease variants of PPARγ (117), but no known N-terminal truncations or cleaved PPARγ <40 kDa are reported in the literature. These smaller isoforms may also be non-functional splice variants or breakdown products resulting from proteasomal degradation.

3.5 Conclusions and Future Work

This study revealed a complex series of PPARγ-reactive proteins were observed across a series of extracts in NHU cells (summarised in Figure 3.5.1). PPARγ at 58 kDa was at the expected molecular weight for PPARγ2 and reacted with an anti-PPARγ2 antibody with greater affinity than other isoforms. PPARγ1 at 52 kDa and the isoform at 50 kDa previously observed in adipocytes were also present at relatively high abundance. A 42 kDa isoform was also observed which was abundant in non-extracted fractions. All isoforms measured upregulated with differentiation, and the presence of 52 kDa PPARγ was variable in undifferentiated cells. Evidence from PCR supported the likely presence of multiple splice variants which would explain the large number of antibody-reactive bands. The only two isoforms of PPARγ observed by western blot which reacted to the PPARγ-E8 and not the PPARγ2 antibody were the 50 and 16 kDa isoforms.

γ2 / E8	Whole Cell	Co-IP Nuclear Fr.	IP	0.5 M NaCl CSK	CSK-DNase	CSK-DNase Pellet	Possible Identity
80 kDa	++	-	-	+	-	-	2xSUMO-PPARγ (58kDa)
64-75kDa	+	+	+(64 kDa)	-	-	-	2xSUMO-PPARγ (42, 50, 52 kDa)
58 kDa	++++	+	+	-	-	++++	PPARγ2 + Phospho-PPARγ2
54 kDa	+	-	-	+	-	-	PPARγX
52 kDa	+++	+++	++	++	++	+	PPARγ1
50 kDa	++++	+++	++	++	++	+	PPARγY
48 kDa	+	-	-	-	-	+	Splice Variant
42 kDa	+++	+	+	+	+	++	Splice Variant
35-40 kDa	+	+	+(40 kDa)	+	+	-	Splice Variant
30-32 kDa	+	+	+	+	-	+	Splice Variant
22kDa	+	+	-	+	+	+	Splice Variant
16 kDa	+	+	+	+	-	+	C-terminal fragment PPARγ

Figure 3.5.1. Summary of potential PPARγ isoforms observed in NHU cells. Green = reactivity with PPARγ2 antibody, red = reactivity with PPARγ-E8 antibody. Observations of qualitative relative abundance (-/+/++/+++/++++) in lysates, extracts, immunoprecipitations and pellets across all western blots depicted. PPARγX and PPARγY denote possible transcript variants.

The potential < 50 kDa isoforms of PPARγ observed by western blot represent potential splice variants missing some of exons 2-5 as they are immunoreactive to both the N and C-terminal portions PPARγ2 antibodies. Further work will be needed to determine this, and the role, if any, they play in differentiation.

- As many isoforms were observed by western blot which were reactive with PPARγ2 and PPARγ-E8 antibodies, further work to establish the identity of these isoforms could use with 3' rapid amplification of cDNA ends (RACE) from the terminal exon of PPARγ and 5' RACE from the 5' untranslated region of PPARγ2 to comprehensively amplify all PPARγ transcripts for amplification.

- Alternately, PCR approaches using primers tiled across exons of PPARγ could be used to detect if any exons are consistently spliced out of the PPARγ gene in NHU cells. PCR targeting exons 1-3, 2-4, 3-5 and 4-6 should all produce single amplicons if no splice variants of the canonical exons 1-6 are present. If unexpected amplicons were present, these could be sequenced to identify the skipped exon(s).

- If the presence of splice variants is demonstrated, they could be specifically knocked down by siRNAs which span the exon-exon boundaries. Effects of this knockdown on differentiation will determine if they play a role in the development of the differentiated phenotype.

This study has shown that PPARγ isoforms are constitutively present in the nucleus of near-confluent NHU cells independent of the induction of differentiation. As no significant changes in PPARγ localisation were observed between time-points during differentiation which correlated ERK-activation or TZ treatment, further research into the mechanisms of PPARγ activation in NHU cells is required. Blocking either ERK1/2 or AKT aids TZD-driven differentiation of NHU cells, and PPARγ phosphorylation has been observed to be altered after inhibition of EGFR signalling in NHU cells (26).

- As the evidence presented in this study shows that some PPARγ is present in the nucleus and the salt-resistant fraction at all time-points, use of antibodies specific to the phosphorylated serine at position 112 of PPARγ2 could be informative with respect to the extent to which PPARγ is phosphorylated in NHU cells throughout the six day time-course of differentiation. If a significant proportion of PPARγ remains unphosphorylated in undifferentiated NHU cells, it is likely that PPARγ transcriptional activity is repressed in NHU cells other than phosphorylation by ERK1/2. In this case, other PPARγ repressive

mechanisms such as phosphorylation by cyclin-dependent kinase 5 (CDK5) have been reported which may warrant further investigation in NHU cells (151).

- The nuclear re-localisation of PPARγ in untreated NHU cells could be studied by observing actively proliferating cells over time as they reach confluence to assess the effects of confluence on PPARγ localisation.

The study also demonstrated that upregulation of PPARγ isoforms and differentiation markers were significantly attenuated by the presence of PPARγ2 siRNA without significant effect on basal levels of PPARγ at 50 and 58 kDa. 52 kDa PPARγ was variably present in non-differentiated control cells, but its upregulation was significantly attenuated by the presence of PPARγ2 siRNA.

- Future studies could extend this work by over-expressing PPARγ2 in NHU cells, and monitoring the cells for urothelial and prostatic markers of differentiation. This will assess whether urothelial cells can transdifferentiate into prostate cells via PPARγ2, or if they remain as differentiated urothelium.

The 80 kDa isoform of PPARγ was shown to reduce in abundance in the presence of the SUMOylation and histone acetyltransferase inhibitor anacardic acid. This PPARγ is therefore a potential SUMOylated isoforms of PPARγ, although further work will be needed to demonstrate this and to decipher if any transrepression of NF-κB bound genes occurs in NHU cells (119).

- Although the evidence is not unequivocal it would be interesting to investigate if the presence of this band alters in response to inflammatory signalling (e.g. IFNγ, as it could be that a SUMOylated PPARγ functions to resolve inflammation after an immune response as it does in macrophages. This would be interesting to study in urothelial cells from of interstitial cystitis patients, who have chronic inflammation of the bladder (121).

4 Label-Free Mass Spectrometric Investigation of Changes in Chromatin-Associated Proteins in *in vitro* Differentiated NHU Cells

4.1 Introduction

4.1.1 Mass Spectrometry-Based Shotgun Proteomics

Mass spectrometry-based shotgun proteomics utilises a liquid chromatography (LC) column to separate complex peptide mixtures generated by enzymatic digestion of proteins. A system such as an electrospray ionisation needle is then used to exchange and ionise the peptides eluting from the LC column into the gas phase, and subsequently inject them into the inlet of the mass spectrometer. The mass spectrometer is able to simultaneously measure the mass-to charge (m/z) ratios of all the injected eluting peptides (within the duty cycle and working m/z range of the mass spectrometer used). As unique peptides can have identical elemental composition, and therefore $m/z,$ this makes it impossible to assign confident identification from this single m/z value. To improve confidence of identification, possible peptides are isolated based on their m/z and induced to fragment within the mass spectrometer, allowing the m/z of the fragments to be determined. The masses of the parent ion and its fragments calculated from their m/z can be matched to a database of expected fragment masses for all possible peptides from a given proteome. The probability of the observed fragments matching to a known peptide within the proteome database is then determined using parameters such as the accuracy of m/z measurements, the number of fragment ions matched to the potential parent peptide which fall within the m/z error tolerance, and correction for probability of false discovery.

4.1.2 Label-Free Quantitative Mass Spectrometry

Label-free (LF) relative quantitation utilises data obtained at the detector in the mass spectrometer (MS) during shotgun-based proteomics experiments to estimate changes in protein abundance between samples, without chemically tagging (or labelling) peptides specifically for quantitation. Two approaches to

have been developed for LF-MS: spectral counting and area under the curve or intensity-based.

Spectral counting does not use peptide properties such as intensity of ions at the detector to measure abundance, instead using the empirical observation that numbers of acquired spectra in automated liquid chromatography (LC)-MS/MS experiments and peptides identified per protein increase with protein abundance. The simplest form of spectral counting is expressed in the protein abundance index (PAI), which uses the ratio of measured peptides to peptides theoretically observable in the experimental setup, which depends on factors such as the enzyme used to digest the proteins and the *m/z* acquisition range of the mass spectrometer (152).

Using the PAI as the exponent in the calculation 10^{PAI} -1 Ishihama *et al* (153) showed the exponentially modified PAI (emPAI) value calculated for a protein was approximately proportional to the known relative abundances of peptides spiked into complex whole cell lysates. This emPAI value created a linear scale in which the relative abundance of proteins within a complex sample could be estimated as molar fractions.

One major limitation of approaches such as PAI is associated with how the peptide data are acquired. Standard automated approaches use data dependent acquisition (DDA), whereby the most abundant ions eluting from the LC column and passing into the mass spectrometer at any one time are selected for fragmentation. Once fragmentation spectra are acquired, the mass spectrometer returns to scanning parent ions with the *m/z* window that was just analysed excluded for a period of time, the length of which depend on the experimental setup. Because this approach targets the most abundant ions, this inherently introduces bias towards identifications of proteins that are more abundant.

The issue of estimating peptide abundance is a particular problem for complex samples. Cultured human cell lines are estimated to contain 8,000-10,000 different proteins, which are present at relative abundances spanning many orders of magnitude (154). Detecting lower abundance proteins such as transcription factors among highly abundant structural proteins poses a serious challenge to the dynamic ranges of mass spectrometers. The dynamic range of a mass spectrometer is the range of ion intensities over which the ion signal is linear with analyte concentration. In practice, this means that reliable detection and

quantification of analytes is dependent on the intensity of each analyte relative to those which are most abundant. Useful spectra are generally obtained within 3 orders of magnitude of intensity below the most intense peptide (155, 156).

Signal intensity-based differential label-free approaches use integrated peptide ion peak intensities to estimate relative protein abundances between samples. Highly reproducible chromatography is required in order to enable reliable comparisons of signal intensities of analyte peaks between samples. High resolution mass spectrometry is advantageous in identifying co-eluting peptides with similar masses that would otherwise be deemed to contribute to the same peak. Signal intensities of ion peaks can be acquired in the same experiment or separately from fragmentation spectra. Acquiring technical replicate MS-only runs to obtain averaged values for the peak intensities devotes more time to accurate measurement of peak heights. Peaks which are determined to vary between samples can then be targeted for MS/MS identification in later injections, based on retention time and *m/z* value. Combining MS and MS/MS identification in the same run reduces the time available for acquiring MS data, and so reduces the accuracy of peak measurements compared to performing replicate MS-only analyses (157, 158), and may still require sample reinjection if particular signals are found to vary between samples and were not automatically selected for fragmentation.

4.1.3 Proteomic Studies of Chromatin

Many groups have circumvented the problems associated with sample complexity and made detailed catalogues of nuclear proteins by either repeat injections of whole cell samples or injection of highly pre-fractionated samples into the mass spectrometer. Beck *et al* made a "complete" proteome of a human cell line using directed analysis of multiple injections of the same sample and actively excluding compounds which had already been identified in previous runs (154). Once further injections provided no extra identifications, they declared the proteome complete. Takada *et al* stripped chromatin from human nuclei using salt and nuclease extractions in order to catalogue the proteins in the insoluble nuclear fractions. To achieve this, samples were separated on 1D SDS-PAGE and lanes cut into 54 pieces which were all prepared separately for mass spectrometric analysis (159).

Dutta *et al* prepared chromatin extracts by ultracentrifugation of rat liver homogenates (160). The chromatin was then treated with the nucleases deoxyribonuclease (DNase) or micrococcal nuclease (MNase), to assess which chromatin-associated proteins were released after DNA digestion. Each nuclease cuts DNA where it adopts a specific conformation: DNase when it is "open" or in a transcriptionally permissive state, and MNase indiscriminately at the exposed DNA strands in-between the histone-DNA complexes known as nucleosomes. DNase digestion should therefore release proteins associated with transcription when the DNA template they bind to is digested, and the MNase should release proteins such as transcription factors which bind to the sequences in-between nucleosomes. Each supernatant from the chromatin digests, and the chromatin remaining in the pellet, was separated by SDS-PAGE and lanes split into five slices for in-gel digestion before LC-separation with direct injection into a mass spectrometer and data collection by DDA. Using the emPAI spectral counting method to analyse data from DDA of single injections of triplicate biological samples, across all samples they identified the nuclease extraction sensitivity of 160 known chromatin proteins, ranging from histones to polymerases and transcription factors.

Spectral counting was also used by Mosley *et al* (161) for analysis of changes in members of *Saccharomyces cerevisiae* chromatin complexes. Nuclei were subjected to high-salt conditions and extracted proteins further separated using sucrose density-gradient centrifugation. Identification of chromatin complex members in equivalent sucrose fractions was used as an indicator of whether the proteins were in complexes at the time of isolation.

Zhu *et al.* (162) performed label-free ion peak intensity-based comparisons of MCF-7 breast cancer cell lines after activation of the transcriptional regulator ERα. Whole cell extracts were separated by SDS-PAGE and proteins digested in-gel in 16 separate slices. Peptides from each slice were analysed separately by label-free evaluation of MS intensity and MS/MS spectral counting for MS and MS/MS acquired in the same injection. 2000 proteins were detected over all fractions from all samples, of which 60 (28 with >2 peptides) were found to vary in abundance by ≥1.6 fold in relative intensity. The large dynamic range of peptide intensities in proteins from whole cell extracts means that the identification of lower-abundance proteins such as chromatin-binding proteins, by MS/MS would

likely not have been possible without the extensive pre-fractionation performed by Zhu *et al.*

Partly due to the challenges encountered by such groups in obtaining identifications of such low abundance proteins, few studies so far have attempted label-free proteomic quantitation of chromatin proteins without first using extensive pre-fractionation. This requires the analysis of many samples, which necessitates the use of large amounts of instrument time. The pre-fractionation step is included, as the chromatin proteins have wide ranges of abundance; the complexity of the samples dictates that if DDA is used to obtain MS/MS, only the most abundant ions in that region of the chromatogram will be selected for fragmentation. The reliance on DDA can be circumvented by the use of MS-only analysis followed by targeted identification of peaks which are demonstrated to vary in intensity between samples.

Intensity-based label-free MS comparisons have become the analysis method of choice for large scale clinical studies, as multiple patient samples such as serum, can be analysed in an automated manner (157). Replicate MS-only injections of the same sample can be replaced by replicate biological samples and candidate biomarker peptides identified from targeted identification of peaks found to consistently vary between samples (163, 164). No studies currently available in the literature have attempted to use chromatin extracts in label-free intensity-based proteomic studies.

4.2 Experimental Aims and Approach

4.2.1 Cytoskeletal Extractions and Label-Free Mass Spectrometry

It was the aim of the work presented in this chapter to utilise label-free intensity-based mass spectrometric quantitation approaches (outlined in Figure 4.2.1) to identify chromatin-associated proteins which may be involved in influencing differentiation of NHU cells *in vitro*.

It was hypothesised that cytoskeletal CSK cell preparations as described in chapter three could be optimised to enrich for extractable chromatin proteins, and LF-MS used to detect changes in abundance of chromatin proteins in the extracted fractions when comparing differentiated and control cells (Figure 4.2.2).

In working towards LF-MS analysis of transcription proteins, the following were performed:

- CSK extracts obtained using 2 M NaCl after pre-extraction with 0.1 M NaCl (0.1-2 M NaCl fraction) were digested with trypsin and subjected to LC-MS/MS analysis using DDA and the relative content of the nuclear proteins in the ID list was investigated using emPAI spectral counting.
- LF-MS was performed on biological replicates (same donor line) of 0.1-2 M NaCl CSK extract digests to identify if variations inherent in the extraction procedure would give false positive identifications of changes in proteins.
- CSK extractions were combined with nuclease digestion of DNA to release DNA-bound proteins. The content of these extracts was investigated by DDA before subjecting them to LF-MS workflows (Figure 4.2.1).
- Two software-based approaches for handling LF-MS data from CSK-DNase were investigated: LF-Quant (Bruker GmbH, Germany) and Progenesis (Non-linear Dynamics, UK).

1
- DDA Analysis (MS/MS)
- Protein IDs / emPAI

2
- Triplicate MS-only (Each Sample)

3
- Align MS-only data
- Identify MS-events which alter reproducibly using software

4
- Align MS-only with MS/MS from DDA
- Identify events in DDA which align with MS-only data and quantitate changes

5
- Generate scheduled precursor list (SPL) of *m/z and* elution times to target specific ions, obtain MS/MS of targets by re-injection of sample

6
- Align MS/MS obtained with SPL with MS-only and MS/MS from DDA

Figure 4.2.1. LF-MS approach outline. 1. DDA analyses were first run to check sample quality and assess protein content by emPAI quantitation. 2. Triplicate MS-only runs were obtained for each sample. 3. MS-only data were aligned for all samples to allow comparison of intensity of ions in each sample. MS-only was also aligned with MS/MS obtained from DDA samples to obtain preliminary identifications and allow already-identified peptides to be omitted from SPLs. 4. SPL were generated targeting ions which changed reproducibly between samples and MS/MS obtained by targeting said ions. 5. MS/MS from SPL aligned with the MS-only and DDA data to obtain identification of changing peptides.
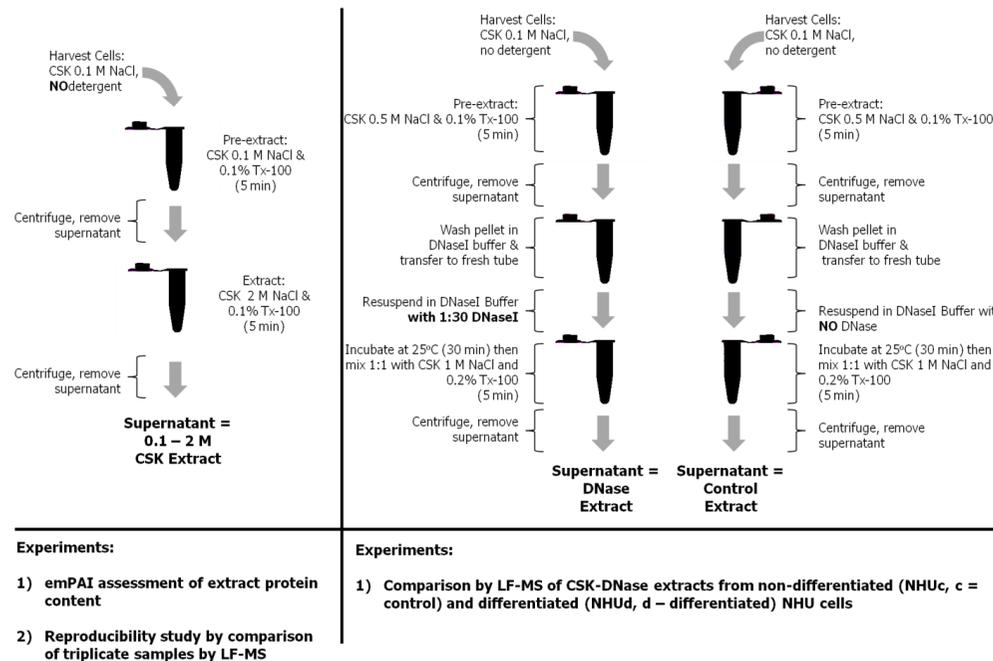
Figure 4.2.2. Experiments outline for LF-MS of CSK extractions from NHU cells. Left panel: Undifferentiated NHU cells were pre-extracted with 0.1 M NaCl, and then extracted with 2 M NaCl (thus containing all proteins extracted between 0.1-2M NaCl). Analysis of protein content by DDA and reproducibility of extractions by LF-MS were performed using the 0.1-2 M fraction. Right panel: subsequent experiments were aimed at using LF-MS to compare proteins released from NHU cells after DNase digestion of DNA. Differentiated (NHUd) and control (NHUc) NHU cells were both subjected to 0.5 M NaCl CSK extraction, then either DNase treatment or an identical enzyme free to control for proteins not specifically released by DNase. Supernatants collected after a second 0.5 M NaCl extraction were subjected to LF-MS.

4.3 Results

4.3.1 Protein Identifications in 0.1 M − 2 M NaCl CSK Fractions

The purpose of this experiment was to determine the similarity of the protein identifications across three biological replicate CSK extracts, obtained by LC-MS/MS using DDA. Biological replicates exposed to the same extraction procedure in parallel should give similar protein identifications, and few peptides which vary in intensity as assessed by LF-MS. Large numbers of proteins altering between extracts from replicate dishes would indicate that CSK extractions, or replicate dishes of cells, were not sufficiently reproducible to allow such samples to be used for LF-MS analysis.

Replicate dishes of undifferentiated NHU cells were cultured to 70% confluence and subjected to CSK extraction with 2 M NaCl, after a 0.1 M NaCl pre-extraction (full methods in Materials and Methods 2.6.1). The resulting 0.1-2 M NaCl fractions and the pellets remaining after extraction were visually compared extracts using Coomassie stained 1D electrophoresis gel and western blotting. 1D gels showed similar protein profiles for all replicate 0.1-2 M extracts and pellets. Western blots of the same extracts showed that the transcription factor PPARγ was qualitatively similar in intensity in all the 0.1-2 M NaCl fractions.
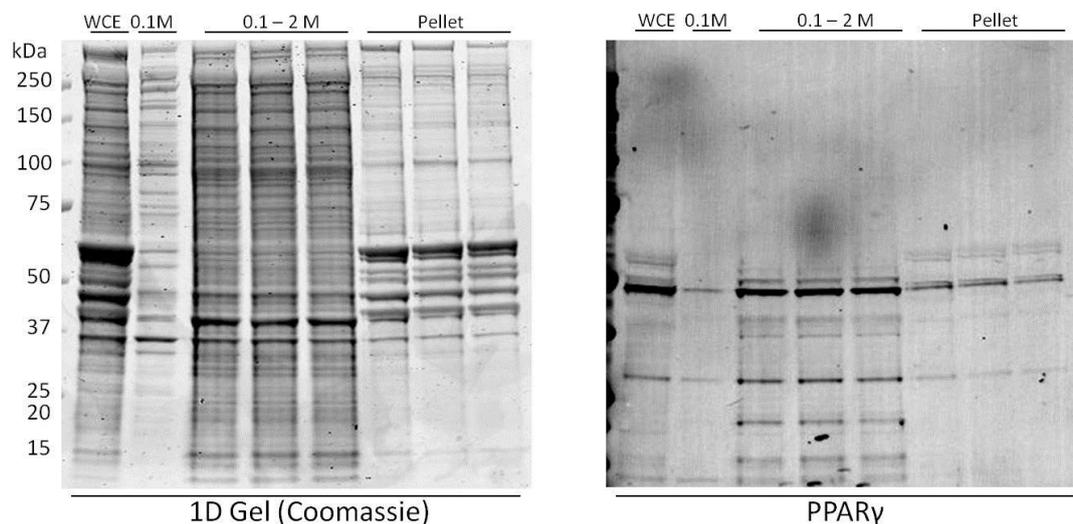
Figure 4.3.1. 1D gel and PPARγ-blot of CSK extracts from triplicate dishes of NHU cell line Y1104. Comparison of pooled non-fractionated whole cell extracts (WCE) and 0.1 M NaCl CSK extracts (0.1 M) with 0.1 - 2M NaCl (0.1 – 2 M) CSK extractions and remaining pellet from triplicate dishes of NHU cells. Lanes of 0.1-2 M NaCl CSK extracts in 1D gels were qualitatively similar between replicates, as was the intensity of labelling of PPARγ across triplicates in western blots.

Separate aliquots of the triplicate 0.1-2 M NaCl fractions were digested using the filter aided sample preparation (FASP) method (Materials and Methods 2.8.1). The resulting tryptic peptide mixtures were separated by liquid chromatography and injected into a Bruker maXis instrument; MS/MS data were acquired in DDA mode (Materials and Methods 2.8.3).

Product ion spectra were submitted to the Mascot server (version 2.3.02, Matrix Science) which was directed to search the human proteome (IPI_human) for peptide identifications. 623-725 accepted protein matches were found per dish, with 522 common to all three dishes (Supplementary Materials 2.1.1). Extracts from dish 1 and dish 3 had similar numbers of protein identifications, with dish 2 having 15% fewer identifications than dish 1 (Figure 4.3.2).
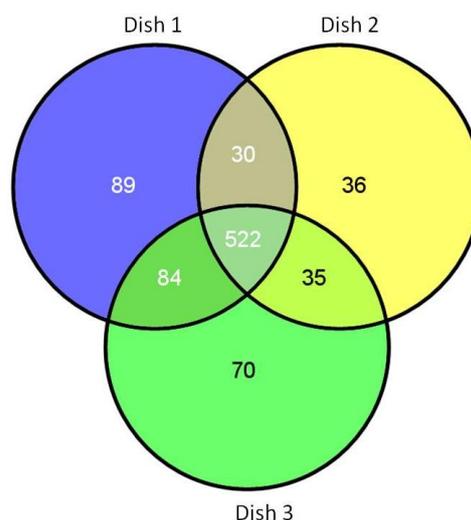
Figure 4.3.2. Proteins Identified in 0.1 – 2 M NaCl CSK Extracts by Data Dependent Acquisition. Venn diagram shows overlap of protein identifications between the three samples. Dishes 1 (blue), 2 (yellow) and 3 (green) gave 725, 623 and 711 individual protein matches respectively. 522 proteins (72 - 83%) were common across the replicate dishes. Each dish gave 5-10% unique protein matches.

Gene ontology (GO) analysis was used to determine which proteins in extracts were potentially associated with DNA transcription. MS/MS data from all three dishes was combined into a single file and a separate Mascot search performed. 872 international protein index (IPI) numbers from identified proteins were converted to Swissprot identifiers using the GeneProfiler g:convert tool (http://biit.cs.ut.ee/gprofiler/gconvert.cgi). The protein identifiers were then submitted to the AmiGO gene ontology enrichment tool (http://amigo.geneontology.org/cgi-bin/amigo/term_enrichment). The AmiGO database collates gene ontology terms associated with proteins, and identifies terms which occur with greater frequency than background in the submitted list. Lists were submitted with default settings of p≤0.01 for enrichment and minimum 2 counts (AmiGO version 1.8).

Proteins annotated with the gene ontology term "Transcription, DNA Dependent" (GO:0006351) or "Nucleus" (GO:0005634) were chosen as the measure for proteins potentially associated with chromatin. Gene ontology terms encompass several "children" terms. There is significant overlap between the chosen terms, as most genes associated with transcription are also nuclear. It was reasoned that most proteins influencing gene expression would be associated with the DNA transcription term, but that this may be too narrow a search term if, for example,

146

the function of the protein was unknown, but it had annotation of nuclear localisation. For this reason, when interpreting the results, proteins were marked with all of their GO terms and those with an annotation of nuclear were only marked as such if they did not also come under the DNA dependent transcription. As such, DNA dependent transcription was the primary term and nucleus the secondary term.

In the CSK extracts, 109 of 872 protein Identifications were linked with the gene ontology term DNA-dependent transcription (Supplementary Materials 2.1.5). Of the remaining protein Identifications, 214 were linked with the ontology term of nucleus and 13 were histone isoforms. The remaining proteins consisted of cytoskeleton (128), endomembrane (165) and mitochondrial (165) proteins which were grouped together as "other" as they were unlikely to contain any proteins involved directly in transcription.

Of the 109 proteins with a GO annotation of DNA dependent transcription, 21 were described as having known transcription factor or co-factor activity (Table 4.3.1). Some of these proteins represented potential targets that would warrant further investigation if it could be shown that they varied in abundance between differentiated and control NHU cells. For instance, beta catenin (CTNNB1) is known to be expressed in NHU cells, but has not been investigated for a role in differentiation, and has been shown elsewhere to have the potential to interact with PPARγ (91). Gene expression of the heterodimerisation partners proto-oncogene c-Fos (FOS) and transcription factor AP-1 (JUN) are known to be reduced in NHU cells upon differentiation (34), and the protein APEX1 identified in these extracts has the potential to modulate their activity (165). Fetal urogenital tissue has the potential to become either urothelium or prostate lineages, and androgen receptor expression is important for regulating differentiation into prostate. Also identified in these extracts was DDX5, which is an androgen receptor co-activator. Negative regulation of DDX5 activity could suppress differentiation of urothelial cells towards the prostate lineage. DDX5 also interacts with Runx2, which is known to influence PPARγ transcriptional activity (96).

APEX1	ENO1,	PDLIM1,
141 (4)	692 (9)	618 (7)
BASP1.	HMGA1,	RAN,
556 (8)	136 (1)	50 (2)
C1QBP,	JUP,	RFXAP,
612 (4)	784 (10)	21 (1)
CSDA,	MTDH,	SND1,
97 (3)	29 (1)	1309 (21)
CTNNB1,	NACA,	SRSF2,
34 (1)	275 (7)	202 (6)
DDX1,	NPM1,	TRIM28,
103 (1)	1373 (8)	673 (8)
DDX5,	PBXIP1,	YWHAB,
671 (11)	62 (1)	282 (8)

Table 4.3.1. Proteins in CSK extracts with known transcription factor or co-factor activity. Proteins in CSK 0.1 – 2 M NaCl fraction with GO annotations of transcription factor or co-factor activity are listed (full description of proteins in Appendix Table 7.3.3.1). Statements about protein function concerning transcriptional activity and associated publications were taken from the UniProt.org database.

4.3.2 emPAI Analysis of Protein Identifications in CSK 0.1-2 M NaCl DDAs

The emPAI is a spectral counting score that uses the linear relationship between the fraction of peptides matched out of the potentially observable peptides from a protein and the mass of the protein, to provide an estimate of the relative percentage of that protein in a sample (153). This measure is included with the Mascot search outputs. emPAI was, therefore, used as a guide to estimate the relative abundance of transcription-associated and nuclear proteins within CSK extracts. To achieve this, proteins with ontology terms DNA dependent transcription and Nuclear from amiGO analyses were converted back to IPI numbers using the g:convert tool. The resulting IPI numbers were then linked those in the emPAI data output from Mascot.

12% of total emPAI score was attributed to histone proteins and 11% to DNA-dependent transcription associated proteins. Other nuclear proteins comprised 30% of the total emPAI and the final 47% was defined as "other" (Figure 4.3.3).

emPAI Contribution of Target Proteins

Figure 4.3.3 emPAI of transcriptional proteins and other nuclear proteins. 27% of emPAI score was attributed to genes associated with transcription. This was achieved by linking proteins with the gene ontology annotation "gene expression" via IPI numbers and emPAI scores in Mascot search results. Remaining proteins with annotations for histone or nucleus were also marked.

To assess the distribution of relative abundances of detected proteins associated with transcription, emPAI scores expressed as a percentage contribution of each protein to the overall emPAI score, were ranked from high to low emPAI percentage contribution, and proteins marked when they were associated with the gene ontology term "DNA dependent transcription" and "transcription factor". The emPAI percentage contributions for proteins involved in gene expression ranged across four orders of magnitude (Figure 4.3.4). No distribution bias could be detected, with target proteins being equally distributed across the top and bottom halves of the detected proteins. This suggests that nuclear proteins were present at a range of relative abundances within the samples.

emPAI Distribution of Proteins with GO Annotations "DNA Dependent Transcription" and "Transcription Factor"

■ Protein with Annotation "DNA Dependent Transcription"

■ Transcription Factor

· emPAI as % Total
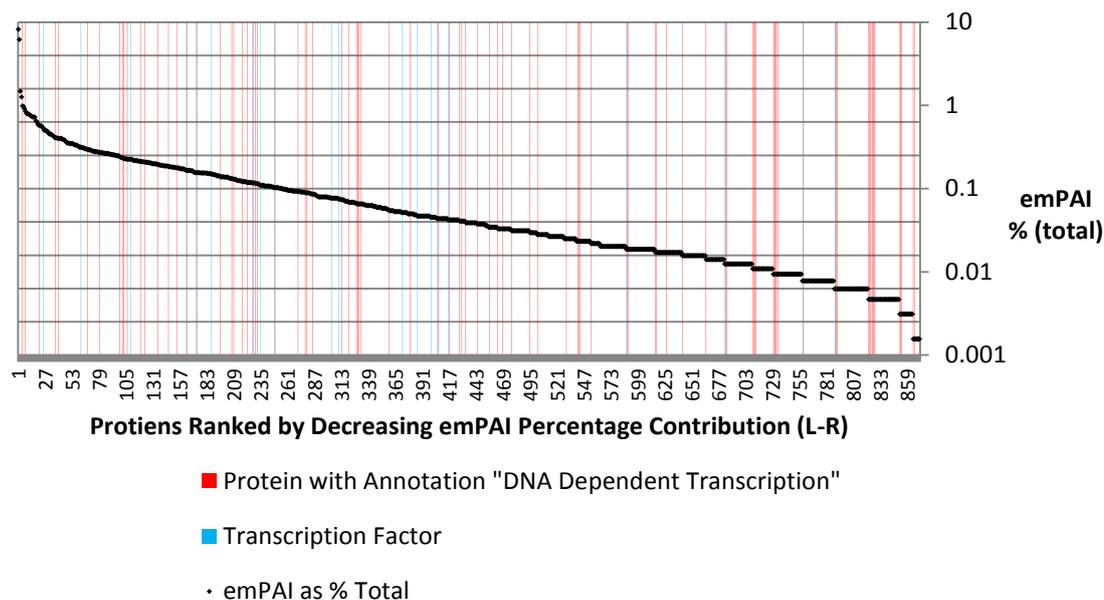
Figure 4.3.4 Distribution of emPAI percentage contributions for proteins with Gene Ontology annotations relating to regulation of transcription. The 109 proteins associated with GO terms 'DNA dependent transcription' or 'transcription factor' were distributed across the entire spectrum of emPAI scores. emPAI scores for all proteins from merged MS/MS of all 0.1-2 M NaCl extracts were plotted as a percentage of total emPAI (right axis). Proteins associated with DNA dependent transcription are marked in red.

4.3.3 Intensity-Based LF-MS of CSK 0.1-2 M NaCl Extracts

Biological replicate samples subjected to the same extraction procedure should display little variability in their protein content. Variability can be introduced when extra sample handling steps are introduced, such as the pre-extraction step in the CSK procedure. If two CSK 0.1-2 M NaCl extracts compared using LF-MS protocols had a significant number of proteins detected as varying then it may not have been fruitful to pursue the use of such extractions for LF-MS studies.

To test compatibility of CSK extracts with LF-MS protocols, triplicate injections of peptides from dishes 1 and 3 of 0.1-2 M NaCl CSK extracts were subjected to MS-only data acquisition. Dishes 1 and 3 had the greatest number of identifications in the DDA experiment, and were chosen as they were likely to give the highest number of features for comparison. Peaks were extracted from MS-data using a "find molecular features" (FMF) script and then MS-profiles time-aligned using the Bruker ProfileAnalysis LF-Quant workflow (Figure 4.3.5).
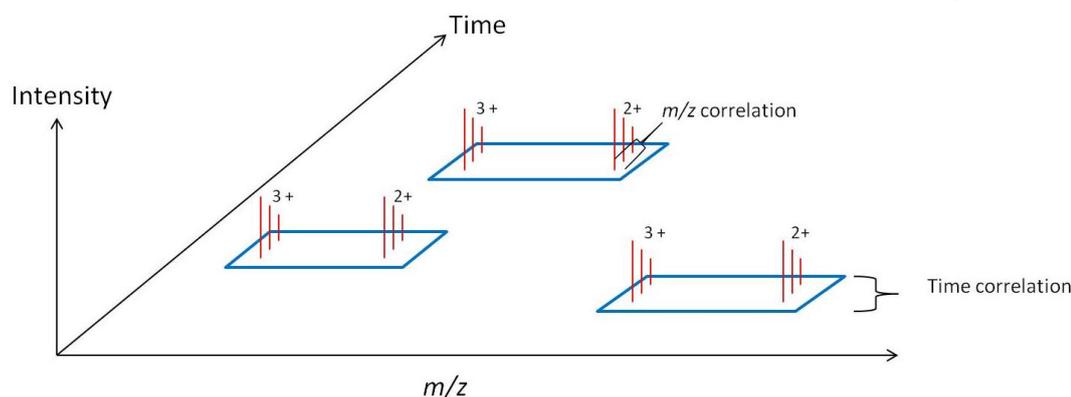
Figure 4.3.5. Find Molecular Features. The FMF script uses the tight time and *m/z* correlation of peaks, as well as the expected distribution of 2+ and 3+ ions to designate features within the MS data. After time-alignment to correct for minor differences in elution times, these features can then be compared between runs of the same sample and between samples.

Potential peptides are assigned feature regions in the chromatogram, which are designated as "buckets". Buckets are then aligned between samples in the time-dimension to account for slight changes in retention times between runs. All six runs (three MS-only replicates of peptides from each dish) were aligned with one another. Buckets designated as containing aligned features were compared between the two sample groups by student's t-test within the Bruker software. The test applied was a two sample, unequal variance, two-tailed t-test with ion intensity from MS-only injections as the input values. Using this approach, a low p-value is obtained when the variance within the groups on both sides of the t-test are low, thus increasing the confidence is gained when intensity values have a narrow distribution within each triplicate MS dataset. To increase stringency, p-values were only calculated when peak values were obtained for the same aligned data point in each run for at least two of the three replicate MS-only datasets in each group.

To obtain protein quantitation without further injections of samples, MS-only data were aligned with the DDA injections for each replicate dish which was merged into a single MS/MS file. The total number of peptides ($p \leq 0.05$) accepted for quantitation was 465 from a total of 5239 in the merged MS/MS file. 98% of peptides which were accepted for quantitation had an intensity value of 10,000 at the detector (Appendix Figure 7.3.1). Of the 874 proteins identified in the Mascot search of the merged MS/MS file, 243 proteins had peptides with peaks which aligned with the MS-only data to enable relative quantitation. Peptides were

rejected from quantitation if they were common between proteins or their p-value was >0.05. 140 of them accepted peptides were the only single matches to a protein, with the other 325 peptides divided over 101 proteins.

46 of the 465 peptides accepted for quantitation (9.5%) had ≥2-fold difference in intensity between the dishes. After each peptide was assigned to its parent protein, 13 of 243 proteins (5.3%) had peptides with median fold change ≥2 (Figure 4.3.6). 58% of all protein fold-change calculations were based on single peptides (Figure 4.3.7), including 10 of the 13 proteins (76.9%) which had a median ≥2-fold change. Within proteins quantitated by single peptide identifications, there was a similar distribution of Mascot scores as compared to the overall Mascot scores from all protein (Appendix Figure 7.3.2). Both these lines of evidence suggest there was no bias for single peptides to be poor quality alignments or outlier measurements, although that can never be ruled out when dealing with single peptides.

Median Fold Change of Proteins in CSK 0.1-2 M NaCl by LF-MS Using in Peptide Intensity Matched between MS-only and DDA Data

Median Fold Change in Peptide Intensity per Protein

Proteins Ranked L-R in Increasing Fold Change (Dish 1 / Dish 3)
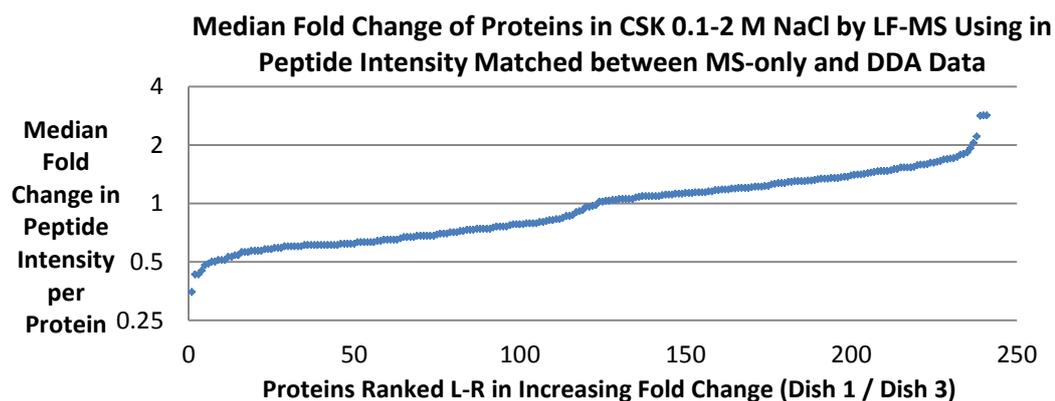
Figure 4.3.6 Fold change of proteins based on median fold-change intensity of constituent peptides. Each protein with peptides matched in the LF-MS data from the DDA runs, was ranked by median fold-change of associated peptides. Out of a total of 243 proteins, 230 and 150 proteins had a median fold change of <2 and <1.5 fold respectively.

Peptides For Quantitation
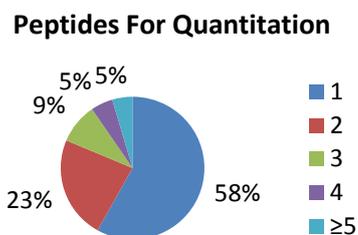
Figure 4.3.7 Number of peptides per protein for quantitation for CSK 0.1-2 M NaCl. The majority of the 239 protein fold-change estimates were based on only 1 peptide (58%).
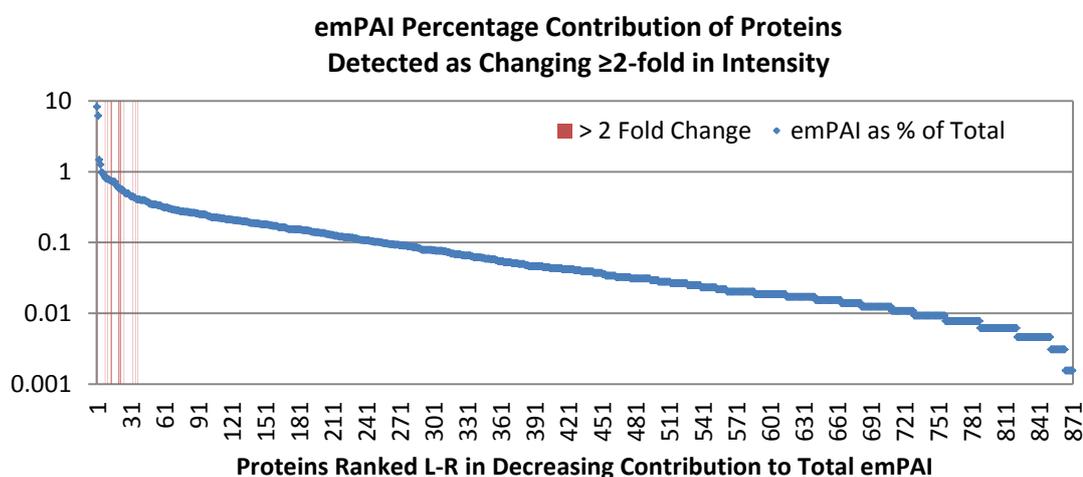
Figure 4.3.8. Distribution of emPAI scores of proteins identified as changing ≥2-fold in CSK 0.1-2 M NaCl. Proteins ranked in descending order of percentage contribution to emPAI score (L-R). Proteins which were detected as changing ≥ 2-fold were all clustered in the top 40 most abundant proteins in the sample and are marked with red lines. Most proteins had only one peptide accepted for quantitation.

Protein Name	Peptides for ID in DDA	Peptides Accepted for Quantitation	Median Fold-Change (Dish1/Dish3)
24 kDa protein	7	1	2.21
ACTB	34	2	0.35
ARL6IP5	3	1	2.82
CFL1	6	1	2.05
DYNLL1	2	1	0.5
ELAVL1	4	1	0.49
HNRNPC	10	1	0.5
KRT2	3	1	2.83
KRT5	6	1	2.83
RPL27	4	1	0.43
RPS13	7	1	0.48
SLC25A6	12	3	0.45
VDAC1	8	2	0.43

Table 4.3.2. Proteins detected as changing ≥2-fold by intensity within replicate 0.1-2 M NaCl CSK fractions. Most of the ≥2-fold changes were based on calculations from single peptides.

The proteins that were detected as changing ≥2-fold were all in the top 40 most abundant detected proteins in the sample, as measured by emPAI (Figure 4.3.9). The detection of only high abundance proteins as changing may have been due to the phenomena of intensity of high abundance peptides at the detector not being linear with protein abundance. Peptide ion intensity measured in the mass spectrometer showed that the peptides which had been accepted for quantitation and which had changed ≥2-fold, were spread out over the range of peptide intensities detected (Figure 4.3.8). However, 75% of peptides detected as changing ≥2-fold in intensity were present in the top 50% most intense peptides, consistent with high abundance peptides having a non-linear relationship with intensity at the detector (153) (Figure 4.3.10).

Intensity of Peptides Changing ≥2 Fold
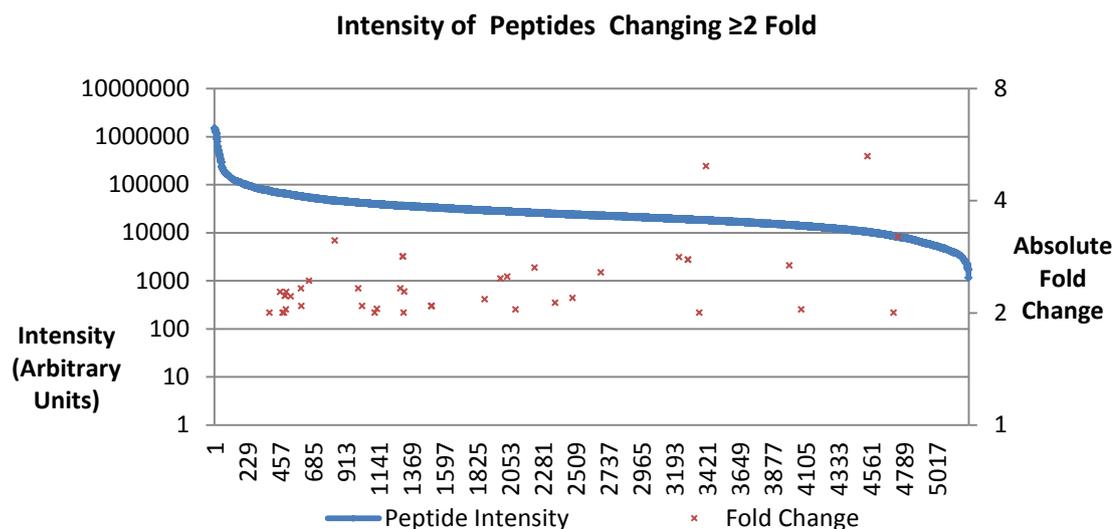
Figure 4.3.9. Intensity of all peptides in CSK 0.1-2 M NaCl versus detected fold change. Peptides were ranked from high to low (left-right on x-axis) in average intensity at the detector (blue line, left hand y-axis) and plotted against and log2 of fold change of peptides accepted by the Bruker software for use in quantitation plotted (red crosses, plotted right hand y-axis). No discernible pattern in changes in detected fold-change across peptide intensity was visible.

Intensity of Peptides Changing ≥2 Fold
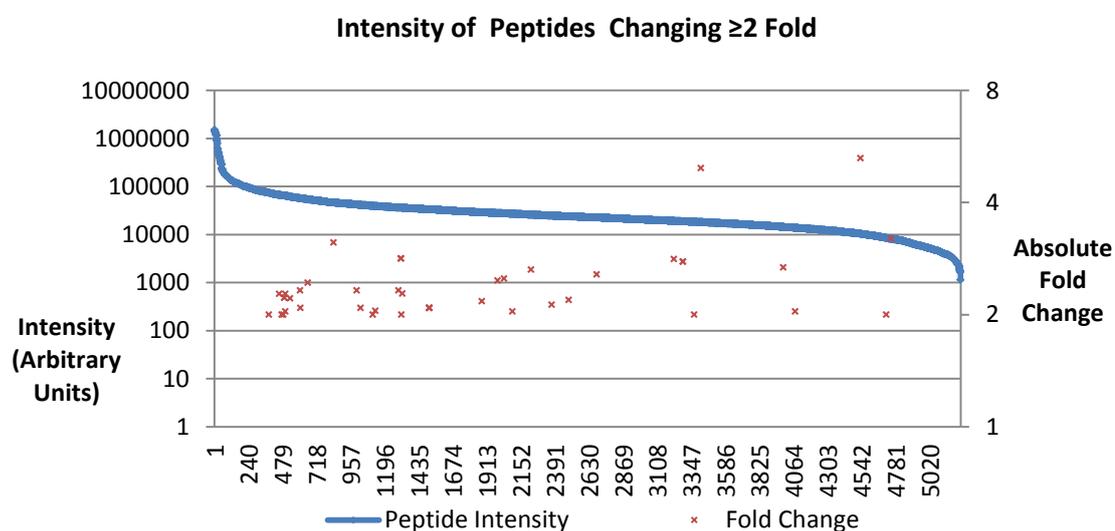
Figure 4.3.10. Intensity of peptides changing ≥2-fold versus detected fold change in CSK 0.1-2 M NaCl. Peptides were ranked from high to low (left-right on x-axis) in average intensity at the detector (blue dots, left hand y-axis) and plotted against and log2 of fold change of peptides accepted by the Bruker software for use in quantitation plotted (red crosses, plotted right hand y-axis). 75% of peptides detected as changing ≥2-fold occurred in the top 50% range of intensity.

4.3.4 emPAI of CSK 0.1 – 2 M Fraction

Although only one DDA injection of each sample was performed, emPAI results were compared to intensity based results to ascertain if similar proteins were detected as changing in abundance as to those in LF-MS. For proteins to be comparable between samples by emPAI, only proteins identified in both samples being compared can be considered, or the fold change is considered infinite. When taking all 624 protein identifications common to both dish 1 and dish 3 of the 0.1-2 M NaCl CSK extracts into account, 104 (16 %) had a percentage emPAI contribution which changed ≥2-fold. Proteins with the large fold-changes in percentage emPAI between dish 1 and dish 3 were proteins which had either one peptide identification in one or both dishes. Removing single peptide hits reduced the number of comparable proteins to 388, of which 47 (12 %) had a ≥ 2-fold change in emPAI contribution (Figure 4.3.11).
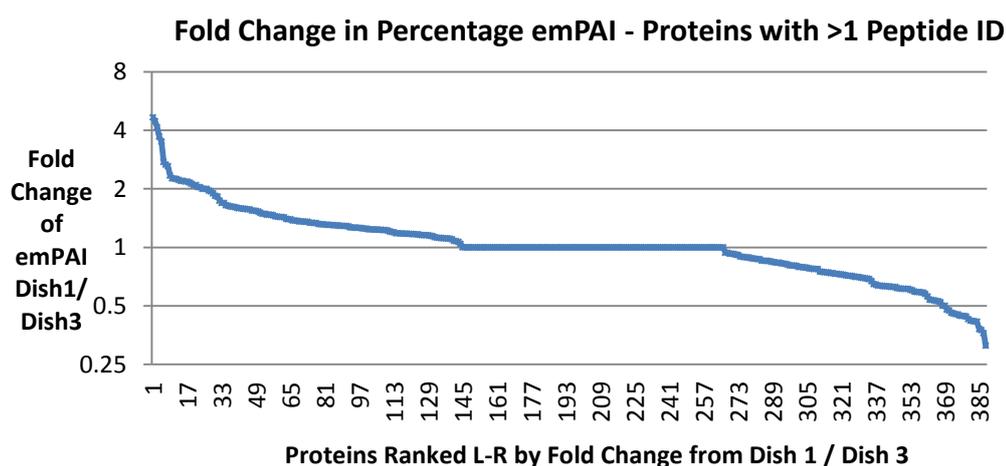
Figure 4.3.11. Fold change of proteins by emPAI between replicate biological extracts CSK 0.1-2M NaCl. The emPAI scores of proteins in dish 1 were divided by corresponding scores in dish 3, and then ranked from high to low in fold-change. Proteins with <2 peptides were omitted from the graph. As emPAI scores can be identical and are only calculated to 2 decimal places, many proteins have an exact 1:1 ratio.

The only protein detected as changing ≥2-fold by both emPAI and intensity-based LF-MS was RPS13 (Table 4.3.3), although it was only quantitated by one peptide in intensity-based MS. Proteins changing ≥2-fold in percentage emPAI contributions were observed throughout the range of relative abundances (Figure 4.3.12).

Proteins Changing 2-fold by Percentage emPAI (>1 Peptide)
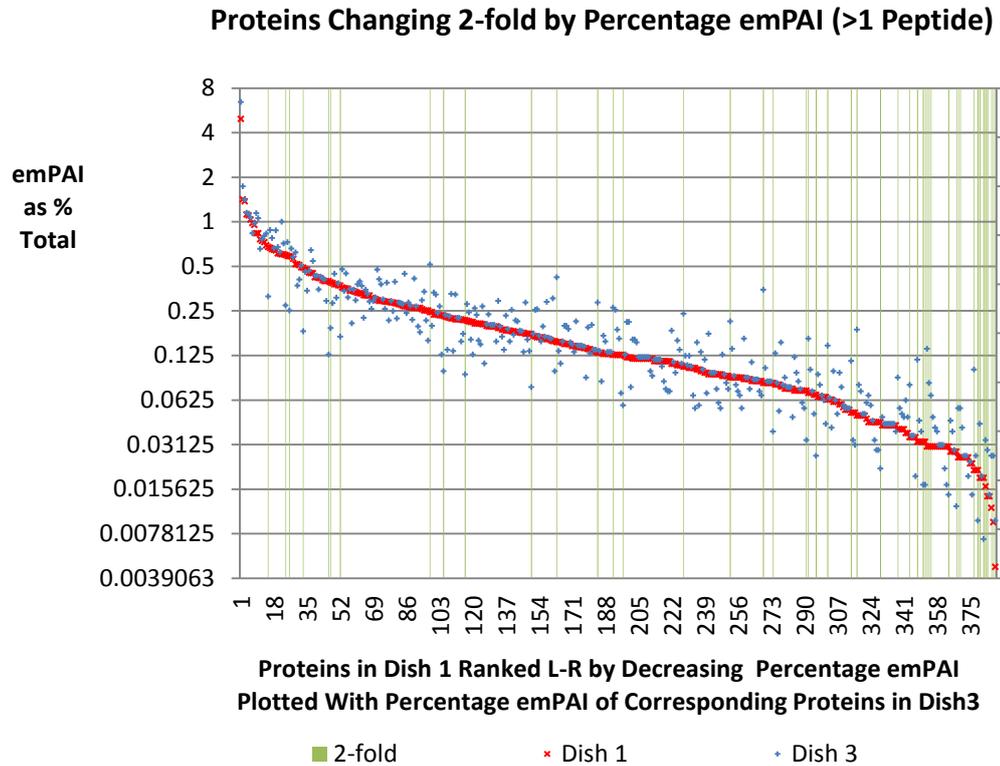
Figure 4.3.12 Abundance of proteins with >1 peptide in both dish 1 and dish 3 of CSK 0.1-2 M NaCl detected as changing ≥2-fold by emPAI. Proteins in dish 1 (red) plotted in descending (L-R) order of emPAI as % total, and emPAI of corresponding protein from dish 3 plotted in the same order (L-R). Differences between the protein emPAI abundances show a bias of those changing ≥2-fold towards low abundance proteins (marked green), with 14 of the 47 changing ≥2-fold being in the bottom 50 % of emPAI scores.

Protein Name	Fold Change (Dish1:3)	Protein Name	Fold Change (Dish1:3)
HNRNPM	4.46	**AP2B1**	0.46
THBS1	2.25	**VAT1**	0.45
PSAP	2.22	**RPS3A**	0.45
EIF2S3	2.15	**TOR1AIP1**	0.45
S100A16	2.66	**S100A14**	0.44
LAMB3	2.61	**DSP**	0.44
HNRNPA3	2.34	**CORO1C**	0.44
SLC25A3	2.17	**RPS4X**	0.44
YWHAB	2.12	**PRDX5**	0.42
P4HA2	2.09	**RPS13**	0.41
RPL18	2.04	**RPS10**	0.41
HADHB	2.03	**GLG1**	0.41
FN1	2	**ARL8B**	0.41
EIF3A	0.5	**KRT5**	0.37
KPNB1	0.5	**PTPRF**	0.37
PYGB	0.5	**RHOA**	0.36
RPS17	0.47	**RPL9**	0.31
TRIM25	0.47		

Table 4.3.3. Proteins detected as changing ≥2-fold by emPAI in CSK 0.1-2 M NaCl. Protein names in bold signify where emPAI calculations were based on >1 peptide in both ID lists. RPS13was (highlighted yellow) was the only protein to be detected as changing ≥2-fold by LF-MS and emPAI.

4.3.5 DNase Extractions

Although 0.1 M – 2 M NaCl CSK extracts resulted in a significant number of proteins associated with transcription being present in extracts, because the pre-extraction step used relatively low salt (0.1 M NaCl), this type of extraction was not selective for proteins bound to the nuclear fraction with high affinity such as transcriptionally active proteins. CSK extractions can be combined with nuclease digestion as a way to release nuclease-sensitive DNA-protein complexes. This option was explored for use with for LF-MS analysis.

In CSK-DNase extractions, cells are pre-extracted with one designated NaCl concentration, and are then incubated in DNaseI buffer with or without the addition of DNaseI enzyme before re-extraction in CSK with NaCl matching or higher than that of the original extraction (Materials and Methods 2.6.2). Parallel control extractions performed without nuclease digestion are performed to allow investigation of whether the release of any detected protein is specific to the DNase treatment.

The aim for pre-extraction is to remove unbound soluble proteins without disrupting the nuclei. 0.5 M NaCl was chosen for extractions, based on western blots of sequential salt extracts in Materials and Methods 3.3.3 which showed removal of the marker of lysosomes and the Golgi, lysosomal-associated membrane protein 1 (LAMP1), and retention of histone proteins at this molarity (Figure 4.3.13). Blots of DNase extractions showed that histones and other proteins were released into the supernatant in greater abundance after DNase treatment (Figure 4.3.14). Transcription factors, such as PPARγ, were also shown to be present in CSK-DNase extracts (Figure 4.3.15).
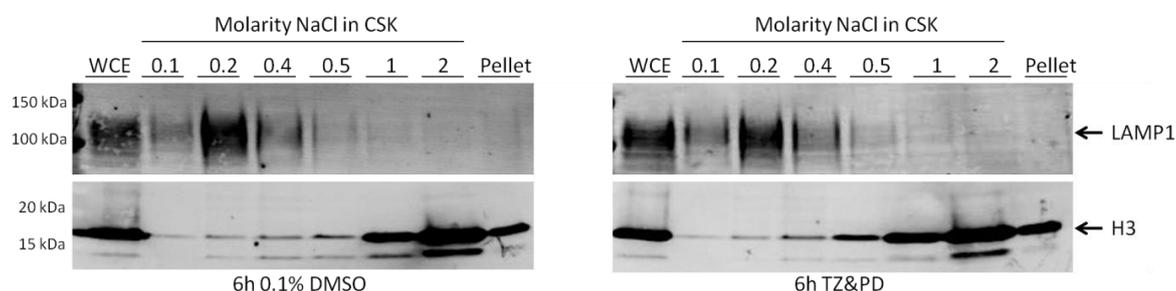
Figure 4.3.13. Western blots of LAMP1 and Histone H3 from sequential salt extracts compared with 20 μg whole cell extract (WCE). Lysosomal and Golgi marker LAMP1 was mostly removed by 0.5 M NaCl. The majority of histones were retained until addition of 1 M NaCl.
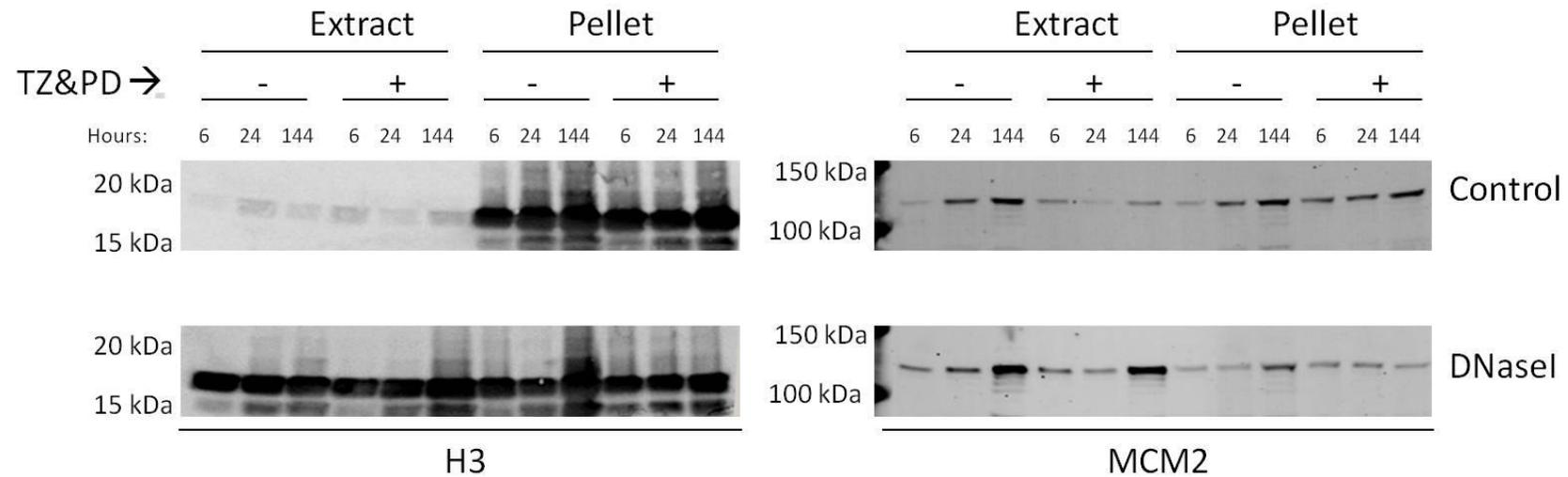
Figure 4.3.14. Histone release after DNase. Western blots show release of histones and the DNA binding cell cycle protein Minichromosome maintenance 2 (MCM2) after CSK-DNase control and DNase extractions. Control extracts released more MCM2 from control (-) as compared to differentiation-induced (+) cells. Cells incubated with DNase released MCM2 and histone H3 into the supernatant in both differentiated and control cells.
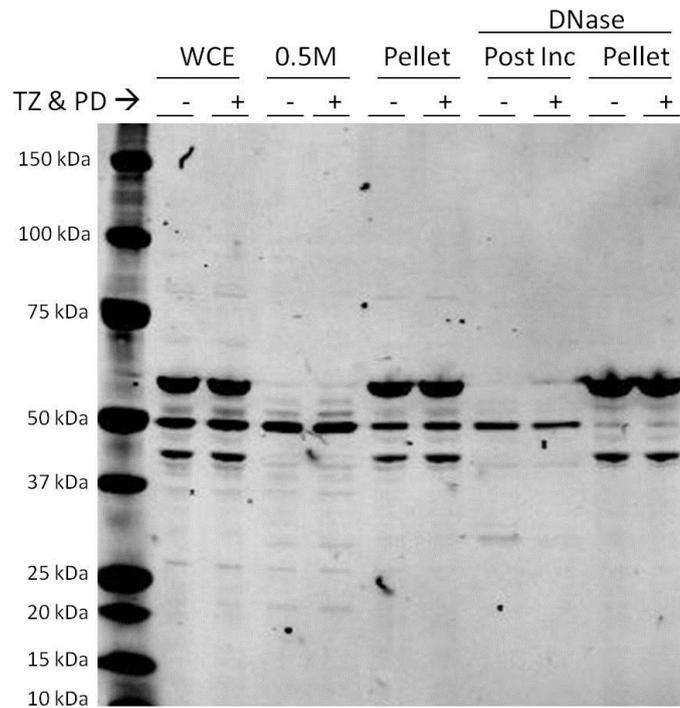
Figure 4.3.15. PPARγ localisation in CSK-DNase extracts of NHU cells. NHU cells were treated for 24 h with differentiation (+) or control (-) agents. NHU cell whole cell extracts taken in CSK buffer show similar intensity of all PPARγ-reactive bands in all lanes. Much of the PPARγ at 50 and 52 kDa is extracted by the initial extraction of 0.5 M NaCl, although some remains in the pellet. Cells which were re-extracted with 0.5 M NaCl after DNase extraction show that the majority of PPARγ at 50 kDa is released by DNase treatment, but that the bands at 42 kDa and 57 kDa remain in the pellet.

4.3.6 DDA Analysis of CSK-DNase Extracts

NHU cells cultured for 144 h with (NHUd) and without (NHUc) differentiation were extracted using the CSK-DNase protocol, both with (DNase) and without (control) DNaseI enzyme (Figure 4.2.2).

Supernatants were prepared for mass spectrometry by in-gel digestion, after running the protein mixture only a very short way into an SDS-PAGE gel (Materials and Methods 2.8.2), as a source of PEG contamination was detected in initial DDA runs from samples prepared by FASP. DDAs of extracts and Mascot searches of MS/MS data were performed as for 0.1-2 M NaCl CSK extracts.

DNase extracts from NHUc and NHUd cells gave 162 and 193 protein identifications respectively. Control extracts from NHUc and NHUd cells gave 96 and 126 Identifications respectively, suggesting DNase extracts were more complex than control extracts, as would be expected. Histone proteins were detected in NHUc and NHUd extracts only if treated with DNase, with similar numbers of peptides identified in the control and differentiated samples.

Proteins identified in all CSK-DNase samples were compared for unique Identifications, bearing in mind that protein ID lists obtained by DDA from complex samples are not exhaustive and absence of ID does not necessarily mean absence of protein in a sample. Venn diagrams of protein Identifications showed that most proteins in the control extractions were also present in the DNase extractions, as would be expected (Figure 4.3.16). DNase extractions for NHUc and NHUd cells gave 78 and 42 unique Identifications respectively.
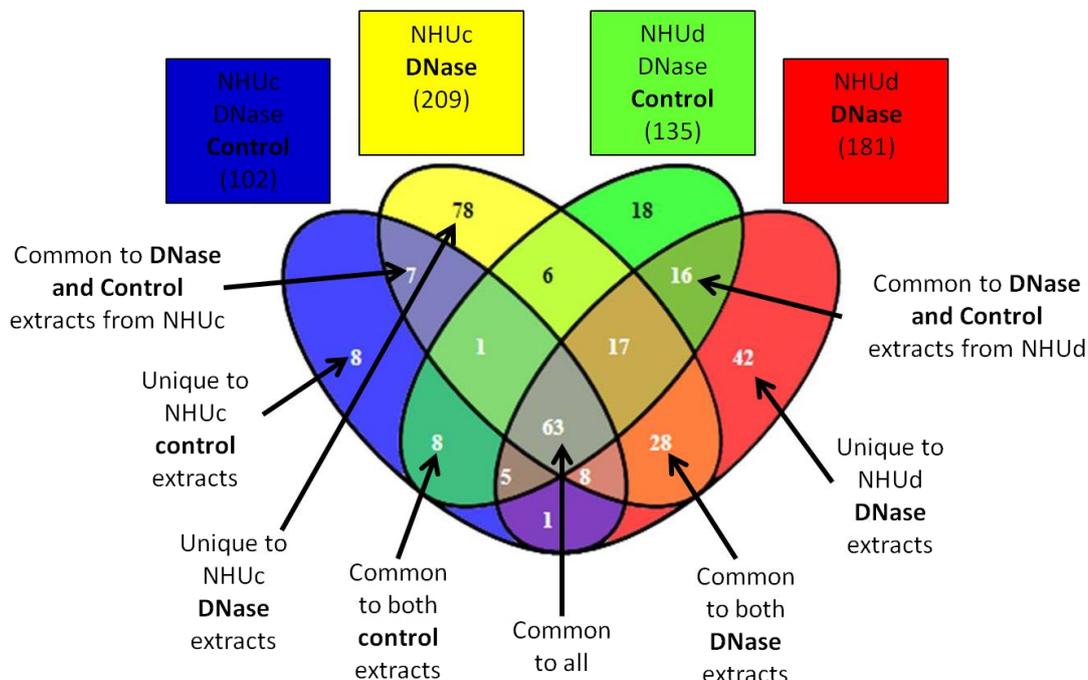
Figure 4.3.16. Venn diagram of Protein Identifications in CSK-DNase DDAs. 144D Control (Blue), 144D DNase (Yellow), 144T no enzyme control (Green), 144T DNase (Red). 63 proteins were common to all samples.

Of the proteins common to all extracts, 24 had GO annotations of DNA dependent transcription (Table 4.3.4). Additionally, 9 and 22 DNA-transcription related proteins were detected that are unique to DNase extracts from undifferentiated and differentiated cells respectively. The range of proteins included wnt-signalling proteins (CTNNB1, CTNNA1, CTNND1, JUP, ZNF326) and sub-nuclear organelle proteins from ribonucleoproteins (ILF2, ILF3, HNRNP and SNRNP proteins), PML bodies (PML), structural proteins (FLNA, LMNA and SAFB), spliceosomes (SRSF), and enhancers of transcription and polymerases (POLR2A/B, SMARCC2, SMARCA4, YBX1).

Common between all extracts from Control or DNase treatment of NHUc or NHUd			Unique to NHUc DNase	Unique to NHUd DNase		
CTNNA1	HNRNPK	PTRF	AGRN	CTNNB1	POLR2A	SMARCC2
DDX5	HSPA8	RBMX	PHB	CTNND1	POLR2B	SMARCA4
EEF1A1	HSPD1	RPS27A	RAD21	DHX9	PSMC5	SMC1A
EIF4A3	ILF2	SP100	RPS3	FAF2	PSMD2	SRSF6
FLNA	JUP	TARDBP	SAFB2	HNRNPAB	RBM14	TMPO
HNRNPA2B1	LMNA	TPR	SHOX	HNRNPD	SAFB	XRCC5
HNRNPF	PHB2	TRIM28	SNRPB	HNRPDL	SF3B2	
HNRNPH1	PML	ZNF326	SRSF7	ILF3	SF3B3	
			YBX1			

Table 4.3.4. DNA-dependent transcription proteins in CSK-DNase extracts. DNA-dependent transcription proteins common to all control and DNase extracts are listed. These included structural elements such as filamin proteins (FLNA, FLNB) and lamina proteins (LMNA). Some ribonucleoprotein proteins are common to all extracts, but others were detected in different extracts, with some being uniquely identified in DNase extracts of differentiated cells. DNA polymerases and members of the switch/sucrose non fermentable (SWI/SNF) transcriptional control complex (SMARCC2, SMARCA4) were also uniquely identified in DNase extracts of differentiated cells. Only two transcription proteins were common to DNase extracts from differentiated and control cells; PHB2 and RBMX.

4.3.7 Comparison of CSK 0.1-2 M and CSK-DNase Fractions

Extracts from CSK 0.1-2 M NaCl and CSK-DNase were compared to determine differences in the types of protein complexes observed. 306 protein identifications were obtained over all CSK-DNase and control CSK-DNase extracts compared to 872 from all CSK 0.1-2M NaCl extracts. The lower number of protein identifications in CSK-DNase extracts was likely due to the more stringent (0.5 M NaCl compared with 0.1 M NaCl) pre-extraction step reducing the complexity of the extracts. Although there were fewer total protein identifications in the CSK-DNase extracts with respect to the CSK 0.1-2 M NaCl extracts, 61% of total protein identifications from all CSK-DNase extracts were also found in CSK 0.1-2 M NaCl extracts (Figure 4.3.17).This suggests a similar sub-cellular fraction was being targeted.

25 proteins associated with transcription were identified in both CSK-DNase and 0.1-2 M NaCl CSK extracts, many of which were ribonucleoprotein complex members (HNRNP, DDX, ILF, RPS3) or ubiquitin subtypes (RPS27A) (Supplementary Materials 2.2.2.1). Also common to both extractions were α, γ (JUP) and β catenin, as well as the transcriptional regulators DDX5, TRIM28 and YBX1. The 60 transcription-associated proteins unique to 0.1-2 M NaCl CSK fractions consisted of multiple variants of similar proteins common to both extracts, such as the ribonucleoproteins (HNRNP, SNRNP, RBMX), ubiquitins (U2AF and spliceosome proteins (SRSF, EIF, U2AF).
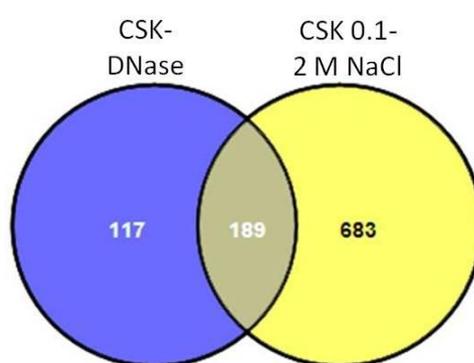
Figure 4.3.17. Venn Diagram comparing summed identifications from CSK-DNase extracts and CSK 0.1 − 2 M NaCl extracts. CSK 0.1 − 2 M NaCl fractions had significantly more overall identifications with 872 identifications compared with 306 from CSK-DNase extracts. 189 proteins were common to both extraction procedures.

Some of the proteins unique to the CSK-DNase extracts included the polymerases DNA-directed RNA polymerase II subunit RPB1 (POLR2A) and POL2RB, the switch/sucrose non fermentable transcription modifier complex members transcription activator BRG1 (SMARCA4) and SWI/SNF complex subunit SMARCC2 (SMARCC2). The presence of polymerases suggests that CSK-DNase extracts had indeed specifically released transcription associated proteins (Table 4.3.5), as seen in a related study using nuclease digestion of isolated chromatin (160).

CSK DNase Only		Common to CSK DNase and 0.1-2 M NaCl CSK	
AGRN	SAFB	CTNNB1	JUP
EIF4A3	SHOX	CTNND1	LMNA
ITGA6	SMARCC2	DDX5	PHB PHB2
MET	SMARCA4	EEF1A1	PTRF
PML	SNRPB	FLNA	RBMX
POLR2A	SP100	HNRNPAB	RPS27A
POLR2B	SRSF6	HNRNPD	RPS3
PSMC5	TPR	HNRNPK	TMPO
RAD21	TRIM29	HNRPDL	TRIM28
RBM14	ZNF326	HSPA8	XRCC5
		ILF2	YBX1
		ILF3	

0.1-2 M NaCl CSK Only				
ANP32A	CSDA	KIAA1967	RAN	SRSF11
APEX1	CSNK2A1	LRPPRC	RBBP4	SRSF2
ASCC3	CSTF3	MTDH	RBM39	SRSF3
BASP1	CTNNA1	MYBBP1A	PARP14	SRSF4
BCLAF1	DDX1	NACA	RFXAP	SUPT16H
BLM	DDX3X	NONO	RHOA	TRIM25
C1QBP	DEK	NPM1	RNPS1	TROVE2
CA9	DRG1	NUDT21	RPS14	TXNIP
CALR	DYNLL1	PA2G4	SET	U2AF1
CAND1	EEF1D	PABPN1	SFPQ	U2AF2
CBX3	ENO1	PBXIP1	SND1	UFL1
CDC5L	FUBP1	PDLIM1	SNRPD3	USP7
CDK9	HMGA1	PRDX5	SNRPE	XRCC6
CHD4	HMGA2	PRKDC	SRRT	XRN2
CHP1	HNRNPUL1	PUF60	SRSF1	YWHAB
	KHDRBS1			

Table 4.3.5. Comparison of identifications of transcription-associated proteins in CSK-DNase and CSK 0.1-2 M NaCl extracts. Common transcription associated proteins identified in all extracts included members of mRNP complexes, transcription regulators and transcription factors. Many of the proteins unique to the CSK 0.1-2 M NaCl fraction were multiple members of protein families, or similar to those common to or unique to CSK-DNase extracts. DNA polymerases and the transcriptional control proteins SMARCC2 and SMARCA4 were unique to the CSK-DNase extracts, suggesting possible enrichment for chromatin-bound transcriptional machinery.

166

4.3.8 emPAI Contribution of Proteins in CSK-DNase extractions

emPAI contribution of proteins was calculated for CSK-DNase extracts as for 0.1-2 M NaCl extracts. Histone proteins were detected in nuclease treated cells and not in control extractions (Figure 4.3.18), in agreement with western blot results which demonstrated their increased abundance in DNase treated cells (Figure 4.3.14). Histone contribution to emPAI was similar to that seen for the CSK 0.1-2 M NaCl extracts, but transcription-associated proteins were lower in CSK-DNase extracts, at 6-9% as opposed to 11%. emPAI representation of proteins associated with gene expression was down from 21% in CSK 0.1-2 M NaCl extracts to 9-16%. Both of these reductions could be accounted for by the removal of proteins by the increased stringency in the pre-extraction step and the use of 0.5 M NaCl for extraction after DNase or control extraction rather than the 2 M used for the CSK extracts.

Although the overall percentage contribution to emPAI score of proteins with ontology of DNA-transcription was the same for NHUc cells subjected to control and DNase extractions, there were more individual transcription-related proteins in DNase extracts which were spread across a wide distribution of emPAI values. This increase was more pronounced in NHUd cells, which also showed an increase in emPAI of transcription-related proteins from 6 to 11% between control and DNase extractions (Figure 4.3.19).
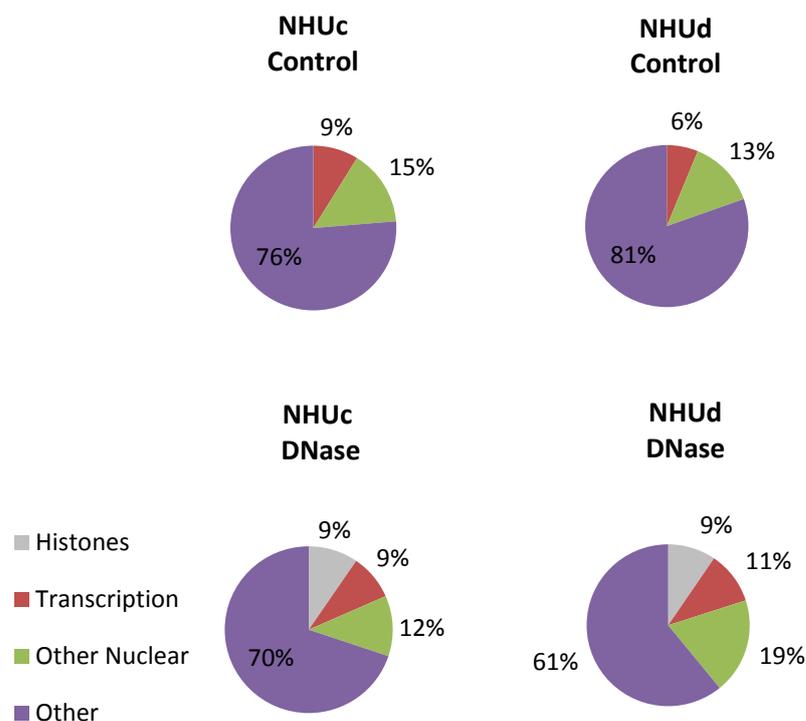
Figure 4.3.18. emPAI contribution of genes with nuclear or transcription related gene ontology annotation in CSK-DNase extracts. IPI numbers from all four protein ID lists were submitted as one to the AmiGo tool. Proteins with GO annotation "DNA dependent transcription" were noted and the remaining proteins with annotations of "nucleus" or "histone" were marked and emPAI summed for each annotation. Control extracts gave no protein identifications for histone proteins, 6-9% of emPAI from genes annotated as "DNA dependent transcription" and 13-15% from other nuclear proteins. emPAI from transcription related proteins stayed the same for DNase extracted non-differentiated cells and increased to 11% for differentiated cells. Nuclear proteins decreased slightly for control extracts and increased for differentiated cells. Nuclease treatment resulted in the same emPAI contribution for both differentiated and control cells.
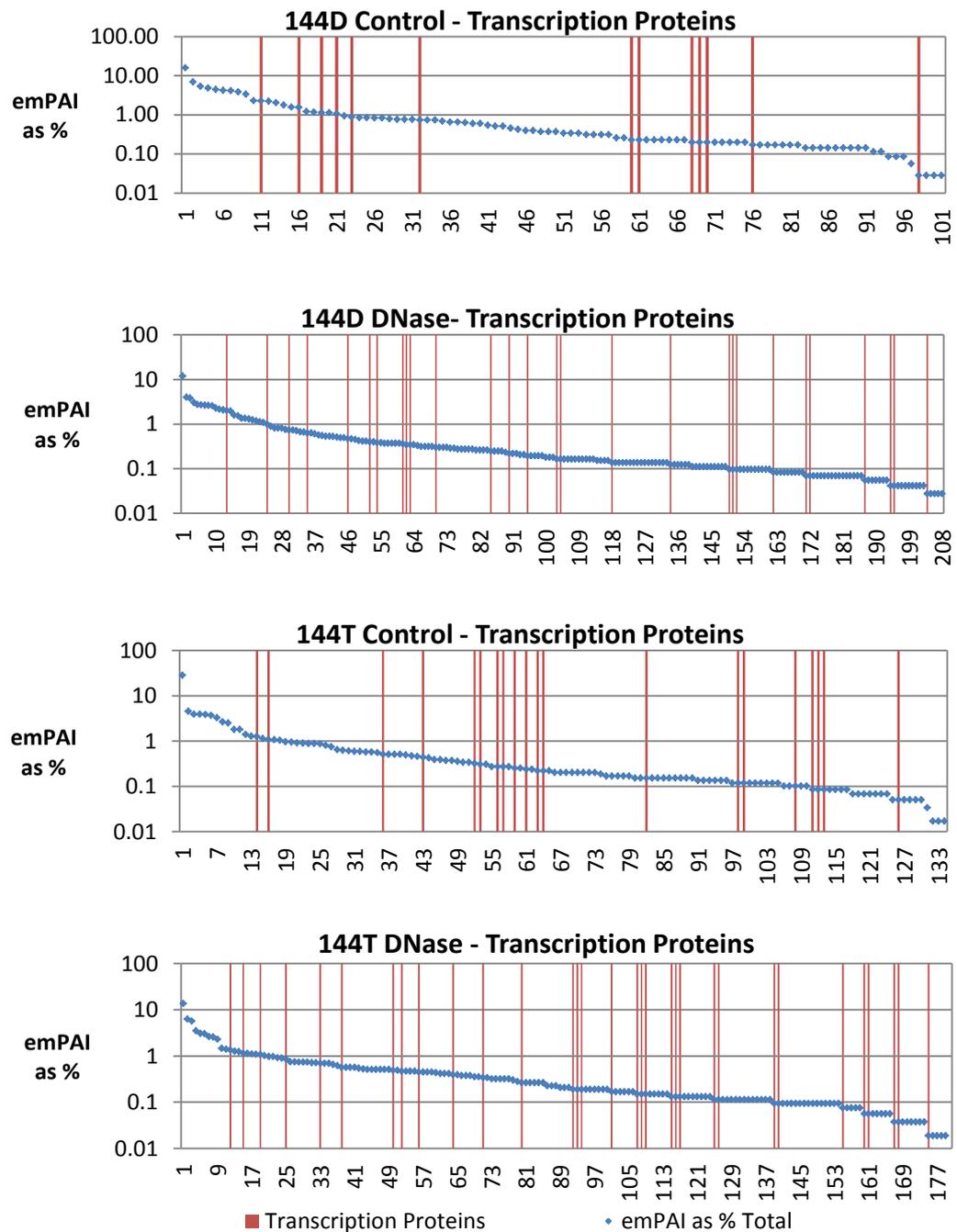
Figure 4.3.19. emPAI distribution of genes annotated as "DNA dependent Transcription" in DNase and control extracts. Protein emPAI scores as a % of total (blue, y-axis) were ranked high to low (L-R, x-axis). Proteins with ontology annotation "DNA dependent transcription" were marked on the graph with red lines. More transcription related proteins were observed in DNase extracted cells.

4.3.9 emPAI Quantitation of CSK-DNase Fractions

emPAI quantitation of the proteins in DNase-treated extractions of NHUc and NHUd cells was performed. When considering all identified proteins, 50 of 117 identifications (42%) common to both samples changed ≥2-fold between the samples. When protein identifications with only 1 peptide were not included this changed to 32 of 82 (39%) proteins (Figure 4.3.20).
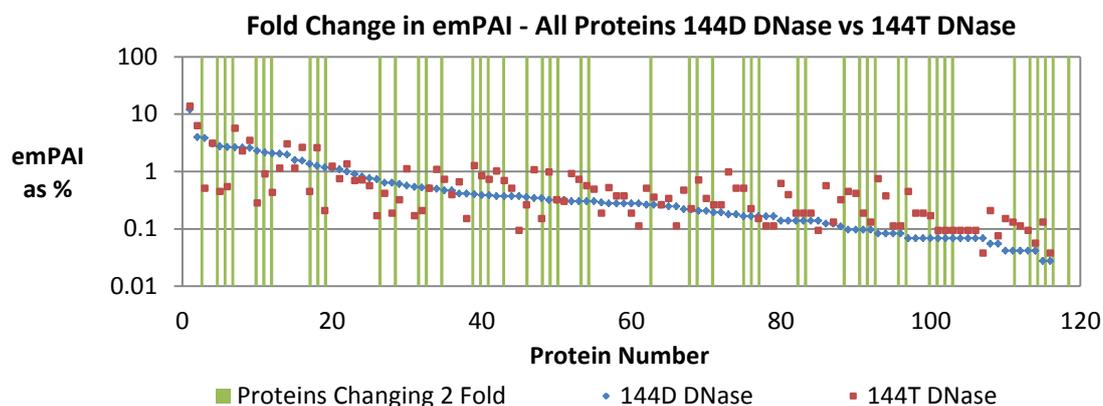
Figure 4.3.20. emPAI distribution of proteins detected as changing ≥2-fold by emPAI in CSK-DNase extracts from DNase extracted control and differentiated cells. Proteins were ranked by emPAI percentage contribution of proteins. 144D DNase extracts (blue) and emPAI contribution of the same proteins from 144T DNase extracts mapped onto the axis. Proteins detected as having emPAI contribution of ≥2-fold are marked with green lines, which shows a uniform distribution throughout emPAI contribution.

The proteins detected as differing between the DNase treated extract samples of NHUc and NHUd cells were spread out across the range of emPAI scores without bias towards high or low abundance proteins. Similar distribution of relative abundance of proteins found to be changing ≥2-fold were obtained when comparing CSK-DNase control extractions from NHUc and NHUd cells to one another, although the number of proteins changing was lower due to fewer identifications and the restriction that, to make comparisons, peptides had to be positively identified in both samples being compared (Supplementary materials 2.2.2.2).

33 proteins were calculated as upregulated in DNase extracts from NHUd cells compared to DNase extracts from NHUc cells, with 22 having >1 peptide in both samples for fold-change calculations (Table 4.3.6). 17 proteins were identified as

upregulated in DNase extracts of NHUc cells as compared to NHUd cells, with 9 proteins having >1 peptide. In control extracts, 14 proteins were detected as upregulated in undifferentiated cells (11 with >1 peptide), and all 8 proteins upregulated in differentiated cells had only one peptide in either sample.

CSK-DNase emPAI				
Control vs Control		DNase vs DNase		
Upregulated in NHUc Control	Upregulated in NHUd Control	Upregulated in NHUc DNase	Upregulated in NHUd DNase	
ATP2A2	ATP1A1	**ANXA2**	**ACTN4**	MAOA
HNRNPK	CLTC	ATAD3A	ACTR3	**NDUFS1**
HSPA8	DSP	BSG	**ATP1A1**	**NDUFS2**
LAMA3	MYH9	**DSP**	**ATP5A1**	NNT
LAMB3	SPTBN1	HIST1H2BL	CTNNA1	**RBMX**
LAMC2	TPR	KRT10	DDX5	**RPL6**
MATR3	UQCRC1	**LAMA3**	**FLNA**	**RPSAP15**
RPN1	UQCRC2	**LAMB3**	**FLNB**	**SLC25A24**
RPN2		**LAMC2**	**HIST1H2AL**	SPTAN1
SDHA		PHB2	HNRNPA3	**SPTBN1**
SLC25A24		**PLEC1**	**HNRNPF**	TACSTD2
TGFBI		PTRF	**HNRNPH1**	**TARDBP**
UBC		**STOML2**	**HNRNPK**	TOMM70A
		TGFBI	HNRNPU	**UQCRC1**
		VDAC1	**ILF2**	**UQCRC2**
		VDAC2	**ITGB4**	VCP
		VIM	**KRT17**	

Table 4.3.6. Proteins identified as upregulated using emPAI in undifferentiated (NHUc) or differentiated (NHUd) cells when comparing control or DNase extracts. Fold change of emPAI contribution (%) was used as a measure to determine which proteins were more relatively abundant in control or DNase extracts. Proteins in bold had >1 peptide meeting the quantitation criteria. Highlighted proteins were observed as changing ≥2-fold in >1 comparison, for example HNRNPK was more abundant in NHUc than NHUd control extracts , but more abundant in NHUd DNase extracts, suggesting it was released to a greater extent NHUd cells in the presence of DNase.

Proteins observed by emPAI to change ≥2-fold in abundance when comparing control extractions of NHUc cells with control extractions of NHUd cells (control vs control), and DNase extractions from NHUc cells and DNase extractions (DNase vs DNase) from NHUd cells were either ribonucleoproteins, structural proteins or metabolic proteins. Comparisons by emPAI of control extractions from NHUc or NHUd cells to their respective DNase extractions produced lists of proteins as changing between both NHUc and NHUd cells (Supplementary Materials 2.2.2.3), but to a much smaller extent smaller extent than the comparison of DNase extracts from NHUc and NHUd cells. This was due to many of the proteins identifications present in DNase extracts being absent in the corresponding control extract; hence an estimate of change in relative abundance could not be made.

Heterogeneous ribonucleoprotein K (HNRNPK) is a ubiquitously expressed ribonucleoprotein which induces cell cycle arrest as part of the p53 response pathway (166). HNRNPK was more abundant in control extractions from NHUc cells than those of NHUd cells, but, conversely, more abundant in DNase extractions from NHUd cells compared to NHUc cells. Proteins which have such an extraction pattern are potentially associated to a greater extent with the nuclease-sensitive fraction in NHUd cells. Protein Atlas showed expression of HNRNPK in all layers of the urothelium and the underlying stroma, with a qualitative increase in labelling in superficial compared to basal urothelial cells.

Desmoplakin (DSP) is an obligate constituent of hemidesmosome cell junction proteins (167). Protein atlas shows DSP as being expressed weakly but specifically at cell junctions in intermediate layers of the urothelium. DSP was detected by mass spectrometry as upregulated in control extracts from NHUd cells (1 peptide), and DNase extracts from NHUc cells.

SLC25A24 is a mitochondrial carrier protein reported in Protein Atlas to be expressed by all layers of the urothelium, with stronger labelling in superficial layers. SLC25A24 was upregulated in NHUc control extractions (1 peptide), and in DNase extractions of NHUd cells.

Three other mitochondrial proteins, ATP1A1, UQCRC1 and UQCRC2, were upregulated in control extracts from NHUc cells (one peptide each), and also upregulated in DNase extractions from NHUd cells. ATP1A1 is reported in Protein Atlas to be moderately expressed by all layers of the urothelium, with strong

labelling in the junctions underlying the superficial layer. UQCRC1 and UQCRC2 are labelled in Protein Atlas as being present throughout the urothelium, with particularly strong labelling at the superficial surface.

4.3.10 Intensity-Based LF-MS of CSK-DNase Extracts using DDAs

Control and DNase extracts from NHUd and NHUc cells were subjected to triplicate injection in the mass spectrometer and MS-only data collected as for CSK 0.1 -2 M NaCl extracts. To prevent saturation of the detector by high-intensity peptides affecting quantitation, one third the amount of sample used for DDA was injected. MS peak data were extracted using the FMF script, and then MS-profiles time-aligned using the Bruker ProfileAnalysis workflow (Materials and Methods 1.8.5).

Data were imported into LF-Quant, and intensities of aligned peaks compared. Comparisons were performed between DNase extracts from NHUc and NHUd cells. To improve stringency of comparisons between groups of triplicates (six samples total), only aligned buckets with intensity counts in at least five of six buckets were considered.

MS-only spectra were aligned with merged MS/MS data from all CSK-DNase DDA runs using Bruker ProfileAnalysis, as was done for the LF-MS analysis of 0.1-2 M CSK extractions. Samples gave good separation by principle component analysis (Figure 4.3.21).
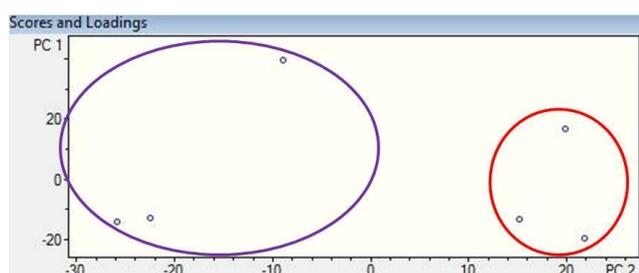
Figure 4.3.21. Principle Component Analysis (Unit Variance) of variables from LC-MS-only runs from DNase Extracts from NHUc and NHUd. PC1 (y-axis) vs PC2 (x-axis) showed separation of NHUc (circled in purple), and NHUd (circled in red). PCA executed in ProfileAnalysis (Bruker).

When comparing DNase extracted samples, 463 of 3078 peptides (15%) identified in the DDA data were successfully aligned with the MS-only data. 133 aligned peptides were detected with ≥2-fold difference between the samples. 31 proteins

had peptides with median fold change in peptide intensity of ≥2. Very few peptides with intensity of <10,000 (arbitrary units) were accepted for quantitation (Figure 4.3.22), which is in line with expectations that quantitation is generally reported to be possible within 2 orders of magnitude of the most intense peptide (149129 arbitrary intensity units).

Peptides Accepted for Quantitation from DDA for NHUc vs NHUd DNase

Peptide MS Intensity (Arbitrary Units)

Peptide Fold Change in MS

—— Intensity   × Fold Change 144D:144T
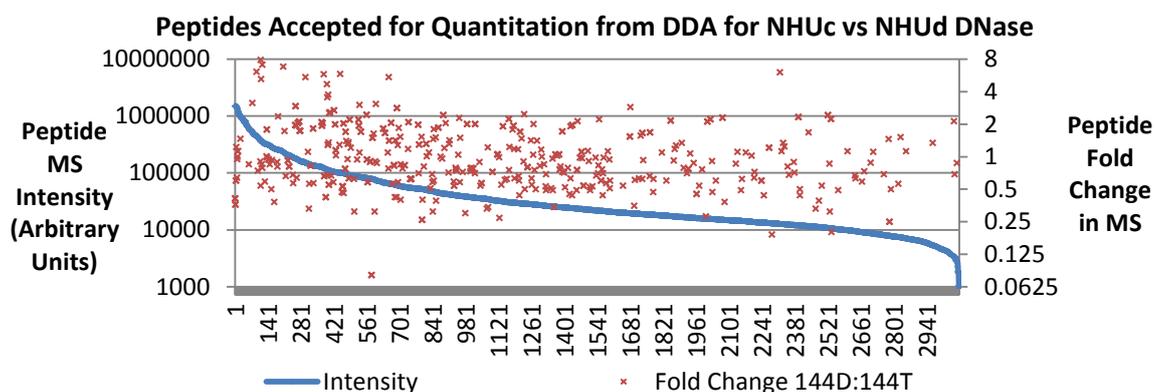
Figure 4.3.22 LF-MS intensity based quantitation of CSK-DNase extracts using peptides aligned by ProfileAnalysis between merged DDA and MS-only data. Peptides were ranked from high to low intensity and fold change of peptides accepted for quantitation plotted on a second y-axis. Peptides p≤0.05 and ≥2-fold change were spread across the top two orders of magnitude of peptide intensity.

When the peptides were mapped to their parent proteins, 30 proteins were identified that contained peptides changing with a median intensity of ≥2-fold between DNase extracts from NHUc and NHUd cells (Figure 4.3.23). When comparing control extracts of NHUc or NHUd to their respective DNase extracts, fewer peptides were successfully aligned than between DNase extracts of NHUc and NHUd cells (Figure 4.3.24 A, B). Comparisons of the control extracts from NHUc and NHUd cells showed similarly low numbers of aligned peptides (Figure 4.3.24 C).

In the comparison of DNase extracts, 19 proteins were upregulated in differentiated cells and 11 upregulated in undifferentiated cells. Proteins detected as upregulated in DNase extracts from NHUd cells as compared to NHUc cells were structural proteins KRT7, SPTBN1 and FLNB, the membrane protein ATP1A1, the mitochondrial membrane protein UQCRC1, the ribonucleoproteins

DDX5, HNRNPAB, HNRNPK, and HNRNPA2B1, histone sub-types H2AFY and H2AFY2, and the dual function ribosomal protein and transcription co-factor EEF1A1. The differential presence of structural proteins in these extracts is difficult to interpret; as they are not the target of extraction it may be that they may be more abundant in the NHUc cells or just more prone to extraction. Looking back at the DDA data, some of the structural and membrane proteins had significantly more peptide identifications in NHUc DNase than NHU DNase extracts, but this appears to be confined to a select few proteins and not all structural proteins, suggesting there was no inherent bias in the extraction procedure (Supplementary Material 2.2.2.1).

Heterogeneous ribonucleoproteins (HNRNPs) are involved with the splicing, storage, transport and degradation of RNA within the cell. Some HNRNPs have been shown to bind to DNA and possess transcription factor activity, specifically towards the c-MYC oncogene in the case of HNRNPK (168). Histone sub-types H2AFY and H2AFY2 are known to preferentially replace the canonical H2A histone in the nucleosome in transcriptionally repressed genomic regions (169).

Proteins detected as upregulated in NHUc DNase extracts as compared to NHUd DNase were the mitochondrial protein SLC25A6, the secreted extracellular matrix proteins FN1, LAMA3, LAMB3 and LAMC2 and the collagen associated TGFBI. The proteins which had a ≥2-fold detected change in abundance between the two samples were a mixture of proteins from the previously mentioned cell fractions, as well as transcription related factors such as CTNNB1 and several types of ribonucleoproteins.

13 of the 30 changing proteins were the same as those seen to change by emPAI (Table 4.3.7). This suggests that the software is successfully aligning high abundance peptides between samples and giving similar estimations of relative abundance as to that which can be obtained by emPAI.

Calculated Fold Change of Proteins with ≥2 Peptides by LF-MS
Comparing DNase Extracts from NHUc and NHUd Cells

Legend: ■ NHUc DNase / NHUd DNase   ✕ % CV

Left axis: Fold Change NHUc /NHUd (8, 4, 2, 1, 0.5, 0.25)
Right axis: % CV (0, 20, 40, 60, 80, 100, 120, 140, 160)

X-axis proteins: KRT7, KRT7, ATP1A1, H2AFY2, DDX5, HNRNPAB, HNRNPK, H2AFY, UQCRC1, HNRNPA2B1, SPTBN1, FLNB, EEF1A1, NDUFV1, UQCRC2, VCP, ACTN4, PTRF, SLC25A12, SLC25A13, ACTN1, HNRNPH1, HNRNPF, JUP, KRT19, LIMA1, PML, MATR3, ILF2, NDUFS1, JUP, KRT17, CTNNB1, HNRNPM, KRT1, ACTA1, ATP5A1, ATP5A1, TRIM29, FLNA, ACTB, BSG, BAT1, HSPA1A, HNRNPUL2, HSPA1L, KRT2, DLST, IMMT, DDX39, DDX21, HSPA8, RPSAP15, CLTC, SERCA2A, HSPA5, HNRNPC, RPN1, OGDH, SP100, DDOST, LMNA, SLC2A1, LMNA, ATAD3A, DSP, CKAP4, THBS1, CANX, SLC25A5, PLEC1, PLEC1, SLC25A6, FN1, FN1, TGFBI, LAMA3, LAMB3, LAMC2

Figure 4.3.23. Calculated fold-change of proteins between DNase extracts from NHUc and NHUd using DDA runs aligned in Bruker ProfileAnalysis. Proteins with ≥2 unique peptides with positive identifications aligned between runs are shown. Most proteins have <2 fold-change between the samples. Coefficient of variance between peptide intensities within groups is represented as a percentage (%CV).
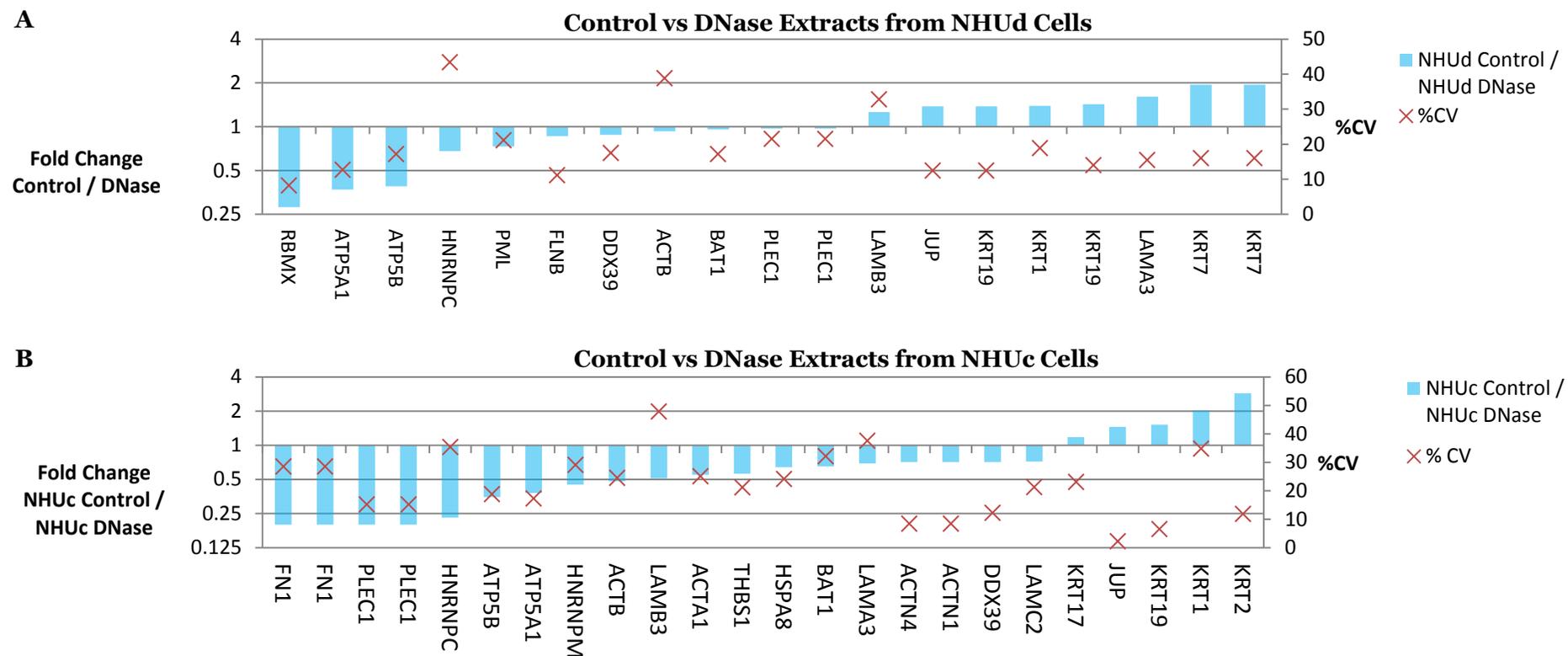
Figure 4.3.24 A, B. Calculated fold-change of proteins between control and DNase extracts of (A) differentiated (NHUd) and (B) undifferentiated (NHUc) NHU cells using MS intensity measurements and MS/MS identification data from DDA runs aligned in Bruker ProfileAnalysis. Proteins with higher median peptide intensity in the DNase extracts above control extracts suggest protein release after DNase digest, whereas the opposite suggests non-specific release. FN1, PLEC1, HNRNPC, ATP5B and ATP5A1 were ≥2-fold more intense in NHUc DNase extracts as compared to control NHUc extracts. RBMX, ATP5A1 and ATP5B had median peptide abundance≥2-fold higher in NHUd DNase extracts as compared to control extracts. Coefficient of variance between peptide intensities within groups is represented as a percentage (%CV).

Control Extracts from NHUc vs NHUd
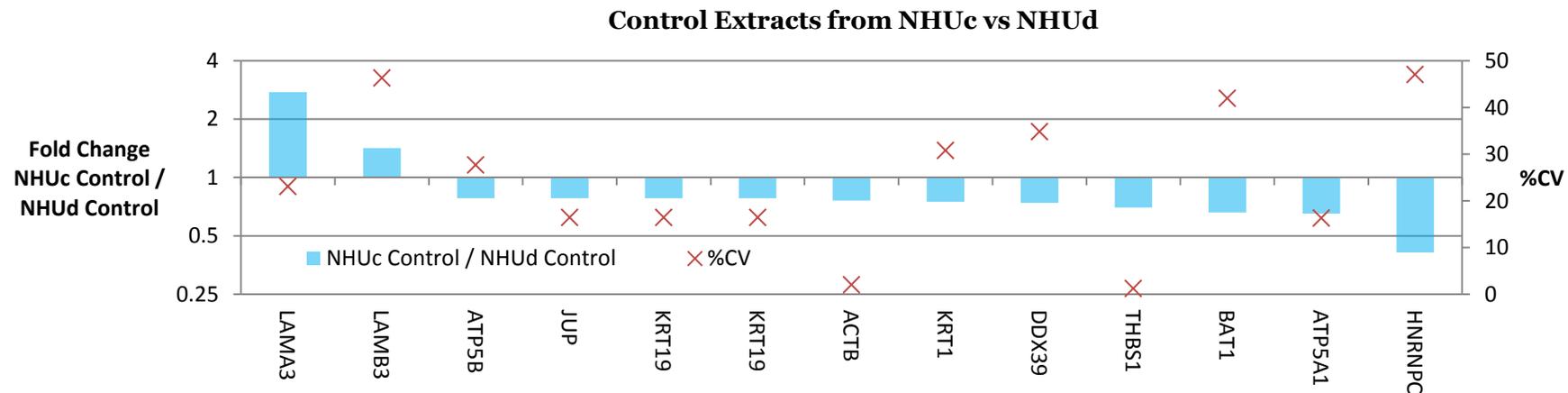
Figure 4.3.24 C. Calculated fold-change of proteins between control extracts of undifferentiated (NHUc) and differentiated (NHUd) NHU cells using MS and MS/MS data from DDA runs aligned in Bruker ProfileAnalysis. Comparison of the two control extractions showed LAMA3 was ≥2-fold more abundant in extracts from NHUc cells and HNRNPC was more abundant in NHUd extracts. Coefficient of variance between peptide intensities within groups is represented as a percentage (%CV).

Upregulated in NHUd vs NHUc DNase Extracts by LF-MS using DDA MS/MS for Identifications				
# Peptides for Quantitation	Peptides in DDA	CV [%]	Name	Fold Change
4	41	21.28	KRT7	3.23
4	35	21.28	KRT7	3.23
3	13	17.02	**ATP1A1**	2.86
3	7	12.26	H2AFY2	2.86
2	4	5.42	**DDX5**	2.78
2	3	18.44	HNRNPAB	2.70
2	7	28.36	**HNRNPK**	2.63
3	17	23.27	H2AFY	2.13
2	8	5.18	**UQCRC1**	2.13
6	8	25.41	HNRNPA2B1	2.08
2	12	1.88	**SPTBN1**	2.08
2	7	10.12	EEF1A1	2.04
13	50	21.95	**FLNB**	2.04

Upregulated in NHUc vs NHUd DNase Extracts				
# Peptides for Quantitation	Peptides in DDA	CV [%]	Name	Fold Change
3	47	2.32	**LAMC2**	5.98
4	57	13.47	**LAMB3**	5.65
8	67	59.49	**LAMA3**	4.05
3	12	21.31	**TGFBI**	2.34
9	46	13.23	FN1	2.28
9	45	13.23	FN1	2.28
3	9	14.89	SLC25A6	2.04

Table 4.3.7. Proteins differing by intensity-based LF-MS using merged DDA for comparison of DNase extracts from NHUc and NHUd cells at 144 h. 19 proteins were detected as upregulated in NHUd and 11 in NHUc cells. The number of peptides use for quantitation did not always reflect the number of peptides seen in the DDA. Coefficient of variance represented as % intensity. Proteins in bold were also detected as changing in corresponding emPAI comparison.

4.3.11 Targeting of Ions Changing by LF-MS not identified by DDAs

As would be expected, not all peaks detected in the MS-only data had corresponding MS/MS data in the DDA file.

257 peaks that were exported from LF-Quant had a p-value of ≤0.05 when comparing NHUc and NHUd DNase extracts. These were split into two SPLs depending on which sample the target ion was the most intense in, to maximise the peak intensity and so maximise the chances of recording useful MS/MS data. Samples were re-injected using the same amount of peptides, as quantified by measurement of absorbance at 280 nm, as used for the original DDA. 2 minute tolerance windows around the retention times of each target *m/z* were used, to account for possible changes in retention time of ions.

Using these SPLs, 43 and 24 peptides were detected in DNase extracts from undifferentiated and differentiated cells respectively (Supplementary Materials 2.2.2.7 and 2.2.2.8). No proteins absent from the DDAs were detected in the NHUd DNase extracts, and 8 previously unidentified proteins were additionally identified in the NHUc DNase extracts. The types of proteins identified were similar to those identified previously from the DDA data, with a mixture of structural and ribonucleoproteins, but a lack of transcription-related proteins.

It was later confirmed that there had been significant retention time drift of peaks in the LC system (Figure 4.3.25). Such retention time drift can occur upon routine replacement of the LC column, which took place between the MS-only acquisition and the SPL being run (a period of several months had elapsed). This meant that the target ions mostly fell outside the target window, which makes it likely that the identifications obtained from the SPLs were unlikely to be the same compounds which had eluted at that time in the original run. These identifications were not pursued further as they were not likely to be useful in giving further identifications.
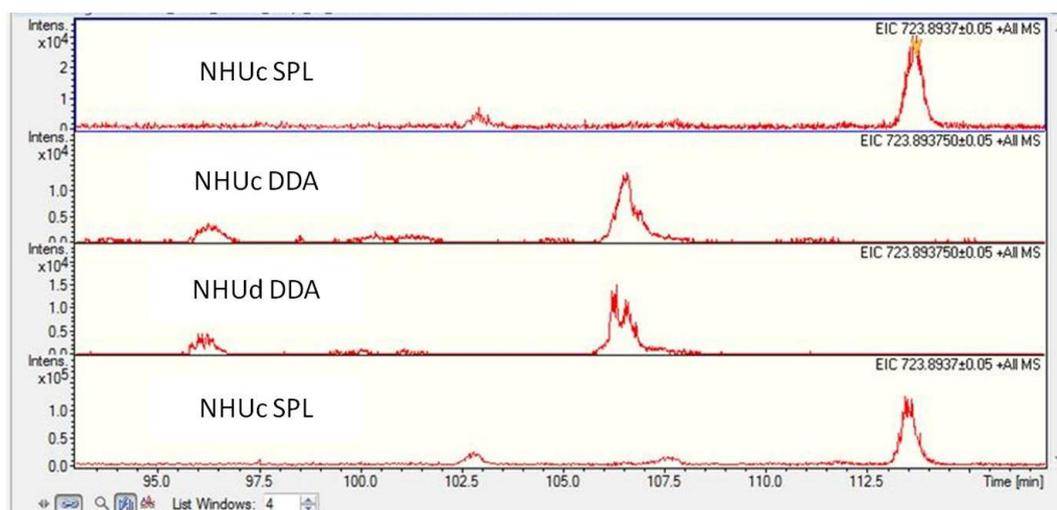
Figure 4.3.25. Elution times of ions between CSK-DNase DDA and SPL.
Extracted ion chromatogram (EIC) of intensity (y-axis) of measurement
at *m/z* 723.893750 (+/- 0.05) from 92.5 to 115 min using MS-data from
DDA of NHUc and NHUd DNase compared with MS-data from SPL
generated using the same two samples. The major peak at this *m/z* was
observed at an elution time of 106.5 min in both DDA samples, and 113.5
min in both SPL samples. As elution times which SPL lists were set to for
this particular target was 106.5 min +/- 2 min, no MS/MS data were
acquired for this *m/z* in SPL.

4.3.12   Nonlinear Dynamics Progenesis LC-MS Software

Progenesis LC-MS (Nonlinear Dynamics) is a software package that, like
ProfileAnalysis (Bruker), can align MS-only and MS/MS datasets for the purposes
of label-free quantitation. Use of this software to compare the CSK-DNase extracts
from NHUc and NHUd cells provided a similar number of compounds identified
as changing as were calculated using all buckets generated by the Bruker software
(including the infinite fold changes for which p-values were calculated separately)
(Supplementary materials). Progenesis LC-MS software does not suffer from one
of the major drawbacks of the Bruker software, which is that if a feature is totally
absent, rather than just reduced in intensity, in one of the datasets, the Bruker
software returns an N/A value; in contrast, the Progenesis software records a
value of infinity. This allows generation of SPL lists even if an ion is absent in one
LC-MS run. As many features in the list from ProfileAnalysis were absent from
one of the two datasets of the DNase extracts from NHUc and NHUd cells, it was

decided that future targeted MS/MS of CSK-DNase samples should be restricted to SPL lists generated using the Progenesis LC-MS software.

Retention times of target ions in the previous SPL were observed to be consistently shifted +5 minutes across the run. SPL target times generated in Progenesis were therefore adjusted by +5 minutes, to account for the shift in LC retention times. The resulting SPLs generated 636 peptide identifications from 136 proteins, of which 215 peptides and 28 proteins were not in the original DDA data (Figure 4.3.26).
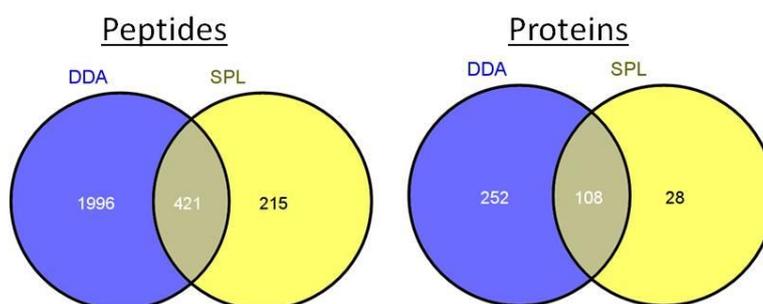
Figure 4.3.26. Venn diagram showing overlap in peptide and protein identifications from combined DDA data from all control and DNase extracts, and those identified in the SPL.

Of the 138 proteins identified in the SPL, 50 were potential nuclear proteins, including transcription factor CTNNB1, chromosomal protein CHD2, transcriptional co-factor SMARCA1, tumour suppressor RUVBL1, RNA-binding proteins RBMXL1 and BAT1, and DNA replication factor MCM2 (Appendix Table 7.4.3.1).

The identification of these proteins was indicative that some of the chromatin proteome was potentially detectable. As the peptides targeted in the SPL were potentially low intensity and unlikely to give good MS/MS based on the observation that in the DDA an intensity of 7000 was the minimum which gave positive peptide identifications in Mascot, ten times more sample (1 µg total) was injected for SPL runs than was injected for MS-only (0.1 µg). To approximate the intensity of the observed peptides with positive identification in the SPL to the approximate intensity in the DDA, the observed intensity in the SPL was divided by ten (Figure 4.3.27). This showed that 506 of the 676 peptides would likely have been intensity 1000-10,000 in the DDA, thus features which would have been below the threshold for MS/MS in the DDA would have constituted many targets.

Intensity of Peptides in SPL
Extrapolated to Original Intensity in DDA
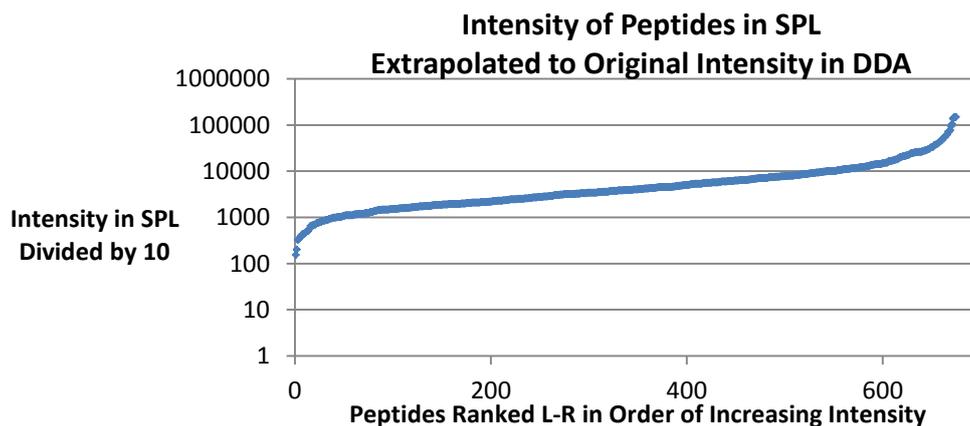
Figure 4.3.27. Intensity of peptides in SPL extrapolated back to likely intensity in DDA by dividing observed intensity by ten.

Alignment of these identifications with the MS-only and DDA data was unsuccessful, making assignment of MS/MS to MS peaks (and therefore quantitation) impossible. Further work on this is in progress as part of an ongoing Progenesis software development project (not reported in this thesis).

4.4 Discussion

It was the aim of this work to identify proteins which differed in relative abundance on the chromatin between differentiation-induced and control NHU cells. Analyses with 0.1-2 M NaCl extracts showed that nuclear proteins could be identified using CSK extractions. The relative label-free quantitation of replicates of the 0.1-2 M extracts showed low numbers of proteins differing in intensity between replicates, suggesting that the experimental setup used did not introduce variation that would mask genuine changes in relative protein amounts when detected using LF-MS. This evidence was considered sufficient to justify the use of LF-MS analysis of extracts in order to obtain relative quantitation.

The CSK-DNase protocol was used to generate samples for LF-MS comparisons to increase the recovery of DNA-bound proteins not soluble in salt. Pre-extraction of the soluble fraction of the cells using 0.5 M NaCl was used to remove weakly chromatin-bound proteins and reduce the abundance of cytoplasmic factors such as lysosomes in the extracts. A second extraction of 0.5 M NaCl after DNA digestion was performed to extract proteins which were bound to DNA. 2 M NaCl was not used at this stage as it was anticipated that the abundance of histones recovered would have been interfered with identification of lower abundance proteins. CSK-DNase extracts contained histones which demonstrated that the DNase digestion had been successful, and also contained transcription factors and transcription-related proteins. Control extractions which did not utilise DNase were introduced to allow the specificity of the extraction to be tested.

The proteins released by DNase treatment after pre-extracting the cells with 0.5 M NaCl consisted of many ribonucleoproteins and other transcription-related proteins which are important to the control of gene expression. Although structural proteins were present in CSK-DNase extracts, if histone proteins are excluded, nuclear and transcription-related proteins accounted for 20-23% of the 162-193 protein identifications by number and 21-23% by emPAI in DNase extracts from NHUc and NHUd cells. This is in line with other studies which have used more extensive purification. Dutta *et al* (160) used sucrose-gradient centrifugation to purify chromatin from homogenised rat liver cells, and over five separate injections into the mass spectrometer managed to identify 160 chromatin-associated proteins from a total of 694 (23%) proteins present in nuclease digestions of the isolated chromatin.

One expected advantage of this type of extraction was that highly abundant structural and cytoplasmic proteins would be reduced in quantity compared to their presence in whole cell extracts, without having to resort to lengthy nuclear fractionations. This was true for some proteins such as keratins, which had reduced Mascot protein identification scores in the DDA analyses of CSK-NaCl and CSK- extracts compared to those of keratins identified in DDA analyses that were run using the pellet of material remaining behind after CSK-NaCl extractions (Supplementary Materials 2.2.2.2.4 and 2.2.2.2.7).

The presence of ribonucleoproteins in the DNase-treated extracts suggested that a specific sub-nuclear proteome was being extracted, with some elements which were similar to those observed by Dhutta *et al* upon DNase digestion of purified chromatin (160). Dhutta *et al* detected proteins from the SMARC and ribonucleoprotein families, as were also observed in the DDAs in the current study. By aligning the MS-only data and performing MS/MS analysis on signals not previously submitted to MS/MS in DDA mode, it was anticipated that deeper penetration into the proteome could be achieved and compounds identified that would not have been chosen for MS/MS by standard DDA protocols.

Peptide identification by Mascot is based on computational interpretation of masses observed in the mass spectrometer after peptide fragmentation. This information allows the Mascot algorithm to search a nominated protein database for amino acid sequence which codes for a peptide with matching mass within the chosen error tolerance which could theoretically produce the same fragmentation patterns. As fragmentation spectra are so complex that they will always contain peaks which are the result of chemical noise or unexpected peptide modifications which could generate random matches for *de novo* sequencing, the use of databases greatly decreases the number of potential matches. As the search will always identify a "best" match, the critical function of the Mascot program is to provide a probability based score which denotes how likely it is that best match occurred by chance (170).

The Mascot score for a MS/MS match which is significant depends on the number of peptides in the database which match the mass of the precursor and the chosen significance threshold. If 1500 peptides in the queried database match the precursor mass within the accuracy to which the mass was measured, to pass a significance threshold of p≤0.05 the Mascot probability (P) score will need to be

45 -10*Log(P) or greater, known as the identity threshold (37). However, even when the score of the best match does not reach threshold, if its score is high relative to the other potential matches then it still has the potential to be a correct match. A second, lower, threshold can be produced by comparing the submitted peak lists with a randomised database which contains amino acid motifs not found in nature, thus assessing the false positive match rate for the peak list. If the peak list does not generate false positive matches, this increases the likelihood that the match is genuine. This second threshold is known as the homology threshold.

Peptides with low ion scores can still contribute to protein identification if multiple matches are achieved, but for it is especially important in large datasets that the peptide matches must pass one or both threshold scores to contribute to overall protein scores in order to reduce the contribution of random matches to protein identification. If a protein is identified by a single low scoring (<30) peptide in a dataset with a large number of searches, there is a risk that the peptide may be a false positive match.

The issue of single peptide identifications is pertinent in the case of some of the proteins associated with transcription which were identified in CSK extracts from NHU cells. If the proteins with single peptide identifications are potential candidates for further work and they have marginal Mascot scores then the data used to generate the match can be further investigated using the existing data in order to better assess the likelihood the match was genuine. If a fragment spectra was matched to more than one sequence, then if there is a large difference between the best and "next-best" score (a delta of ≥10 (171)), then there is a lower probability that the match was due to chance. Repeated searching of the peak lists could be performed using different algorithms, with the rationale that if different scoring algorithms such as MaxQuant (172) or SEQUEST (173) produce the same identifications, the match could be considered more robust. The spectra themselves could be inspected manually and assessed for markers of quality such as high mass accuracy, high signal to noise ratio, high intensity y-ion series below the precursor ion, and several other criteria as summarised by Steen and Mann (174). Other possible approaches include comparing the spectra of the identified peptide sequence directly with those of matching sequence from other experiments which have been deposited on databases such as the global proteome machine (175).

When MS-only replicates were aligned using ProfileAnalysis, PCA showed good grouping of triplicates in all pairwise comparisons. Once the features identified by ProfileAnalysis were aligned with their MS/MS spectra from the merged data from all of the DDA runs, it was possible to use the MS-only ion intensities to estimate the change in relative abundance of proteins which had previously been identified by MS/MS. The scope of this comparison was limited, as proteins had to have a minimum of two different peptides that were present in both samples successfully aligned and identified as belonging to the same protein by the software. Due to these constraints, fewer peptides could be compared between control and DNase extractions than were probably differing, as proteins released by the DNase extraction may not have been present in the control extracts. Although it is not advisable to judge a protein as absent when using only DDA data, which may lack MS/MS for lower abundance peptides, this is a major shortcoming of the Bruker software, because after the MS data were aligned, it should be possible to assess the differential intensities between two samples even if a protein is totally absent from one sample.

The samples which had the greatest number of features designated in the MS-data and which were successfully aligned with the DDA data, were the DNase extractions from NHUc and NHUd cells. These two samples had 463 peptides successfully aligned and 79 of the 116 proteins common between them had ≥2 aligned peptides. Based on the median fold change of the intensity of matched peptides, 50 of 79 aligned proteins did not alter in abundance between the two samples. The proteins which were observed to change in abundance were a mixture of structural proteins and ribonucleoproteins.

No ribonucleoproteins were detected as being upregulated ≥2-fold in NHUc DNase extracts, although 10 were detected as being upregulated ≥ 2-fold in NHUd DNase extracts. This observation could have been attributable to a greater abundance of extracted material in DNase extracted NHUd cells. However, core histone H3 was demonstrated by western blotting to be extracted to a similar degree by DNase treatment in both the 144 h time-point NHUc and NHUd cells used for LF-MS. Peptides from core histone variants H2A and H4, which were aligned in the data from DNase extracts of NHUc and NHUd cells, showed no change, arguing that variability in protein extraction levels is not the explanation.

However, the quantitations of these two proteins were each made based on only one peptide (Supplementary Materials 2.2.2.2). To be able to confidently use the core histones as indicators of equivalent extractions between the cells, more aligned peptides would be needed. Histone variants H2AFY and H2AFY2 were detected as just above and below 2-fold upregulated, respectively, in DNase extracts from NHUd cells above DNase extracts from NHUc cells. These histone variants are not uniformly expressed in human tissues, and cannot be used as an indicator of overall differences of core histones. H2AFY and H2AFY2 are, however, involved in repressing gene expression at genomic sites targeted by SMARC family proteins (169), and could therefore be regulated during differentiation of NHU cells.

LF-MS quantitations showed an abundance of laminin proteins in extracts from NHUc cells above those from NHUd cells. As peptides from all types of lamins detected were consistently more abundant in the extracts from NHUc cells, it seems likely that the lamins were either more effectively extracted in NHUc than NHUd cells or simply more abundant in non-differentiated cultures. Laminins function as part of a trimer of alpha, beta and gamma subtypes which combine together to make a single functional unit, which in the case of LAMA3, LAMB3 and LAMC2 is Lamin 5B (or 3B32) (176). Laminin proteins are components of epithelial basement membranes and form part of the extracellular matrix. Lamin 5B is known expressed at the basement membrane of urothelium in humans, a localisation that is aberrant in urothelial carcinoma (177). Laminin proteins have influence on gene expression through their interaction with integrins, which modulate cell signalling pathways upstream of NF-κB, ERK, AKT and JUN (178). Laminins have also been reported to form part of the nucleoskeleton (179), and are involved in the functional organisation in the nucleus through specific interactions of regions of chromatin marked with particular epigenetic marks and specific laminin variants (180). The detection of laminins in both NHUc and NHUd cultures is likely due to decreased deposition of laminin 5 by cells induced to differentiate. This is backed up by the increased abundance of fibronectin (FN1) in NHUc cells above NHUd cells, which associates with laminins in the extracellular matrix (181).

The proteins identified in the DNase extracts associated with transcription were dominated by members of the ribonucleoprotein family, which are involved in the processing of mRNA transcripts and also exhibit transcription factor-like activity.

HNRNPK, HNRNPAB, H2AFY, DDX5 and EEF1A1 were all detected as increased in abundance in DNase extracts from NHUd cells when compared to DNase extracts from NHUc cells. These proteins all play a part in post-translational control of mRNA and miRNA translation (182, 183). The biology of these ribonucleoproteins is complicated by their diverse functions. For instance DDX5 is involved in mRNA splicing, but can influence gene expression via its interaction with cAMP response element binding protein binding protein (CREBBP) and SMAD3 (184) or RUNX2 (185).

Further identification and quantitation of proteins in the CSK-DNase extractions were pursued. It was proposed that use of SPL lists based on MS-only data should be able to identify some of the compounds which changed in intensity between the two samples which had not been identified using the DDAs. ProfileAnalysis uses the generated p-values and fold-changes to allow the user to filter the full feature list to compile an SPL based on only features which appear to be changing reproducibly between samples. Once the identifications are obtained using the MS/MS data of the ions targeted by the SPL, the MS/MS data is re-aligned with the MS-only data to enable ProfileAnalysis to discern which ions from the SPL should be matched up to which MS feature.

The inability of the software to produce p-values for compounds which were present in one sample and not in the other was problematic, as many compounds were observed to be unique to one condition when the data were filtered less stringently to allow groups of triplicates with zero values through (e.g. 10000, 10120, 10210 vs. 0, 0, 0). It was possible to manually generate p-values for the compounds in these features by exporting the intensity data from the buckets (Supplementary Materials 2.2.2.1), but only outside of the ProfileAnalysis software package. Just from the comparison of DNase extracts from NHUc and NHUd cells, 380 compounds with a manually calculated p-value of ≤0.05 were left out of the SPLs due to this software problem.

Because of the need for features within ProfileAnalysis to be filtered based on p-value or fold-change during automated SPL generation, features for which ProfileAnalysis had not assigned p-values but p-values could be generated manually could not be isolated from the full list of features within the ProfileAnalysis software. Thus, if ProfileAnalysis has generated a p-value for a feature, the SPL can be generated with ease. However, if there is no p-value

associated with the desired feature, for example one which is present in one sample group and absent in another, then it is impossible to filter the list based on any attribute which will include it in a generated SPL other than to export all features. It would theoretically have been possible to use the full feature list as the SPL but this is generally impractical as the number of targets within a given time period exceeds the speed at which the mass spectrometer can acquire sufficient ions to obtain good quality MS/MS (also known as the duty cycle). Multiple injections could have been used to circumvent this, but this would have required excessive amounts of samples and instrument time. One potential work around to this problem was to calculate p-values outside of the Bruker software for buckets with intensities which were present / absent between two samples. This was performed successfully (Supplementary Materials 2.2.2.1). However, although the calculated precursor mass (Daltons) was exported along with the intensity values for the buckets, the charge state of the ion originally observed in the MS-only run was not included in the data exported from the Bruker software. This made it impossible to know what *m/z* was observed in the MS-only data, thus hampering generation of SPLs. One work around for this was to calculate the *m/zs* of the most commonly observed 2+ and 3+ charge states of the precursor ions using the precursor ion mass and to use these two *m/zs* for each precursor ion as the basis of an SPL. The resulting fragmentation data obtained in these SPL and *m/z* windows would then be aligned with to the MS-only data, as for any other SPL, by ProfileAnalysis. Thus, after importing the MS/MS data from the SPL, ProfileAnalysis could filter the identifications based on whether the *m/z* matched what was observed in the original MS-data.

Both standard SPLs based on p-value filters from ProfileAnalysis and manual generation of 2+ and 3+ ions of features with manually generated p-values was attempted (Appendix 2.2.2.4 – 2.2.2.6). However, both results suffered from the same problem of retention time drift, which could have been avoided by performing the SPL before the column was changed. The risk of shifts in retention times occurring between MS-only acquisition and running samples for SPL is a well known issue, whereby chromatography is highly reproducible over several runs, but if elements of the system are changed, there is likely to be an impact on reproducibility.

Better results were obtained when elution times in SPLs generated from Progenesis LC-MS were adjusted by +5 minutes to account for the drift in

retention times. Using the peptide identifications from the MS/MS obtained using these SPLs, 28 proteins not previously observed in DDA were observed. Some of these were well known chromatin remodelling proteins such as the SMARC family proteins and the multifunctional transcription influencing ribonucleoproteins, as well as transcription machinery such as the polymerase proteins POL2RA and POL2RB.

4.5 Conclusions and Future Work

Although the proteome identified in the DDA of CSK-DNase extracted cells in the current study was smaller than the one that Dutta *et al* (160) obtained from DDA analyses of isolated chromatin, the nuclear proteins which were identified appeared to be from a similar sub-nuclear proteome. Some of these proteins were detected to be differentially expressed between the samples when using emPAI analysis of the DDA identifications obtained from DNase extracts. Laminin proteins were much more abundant in control and DNase extracts from NHUc cells, likely because of decreased laminin 5 deposition in the more basal-like NHUc cultures, or increased breakdown as cells were induced to differentiate. The DNase extracts contained many ribonucleoproteins which are known to influence gene expression in a variety of different cell types. Some were identified as upregulated in DNase extracts from NHUd cells as compared to NHUc cells, and others still were identified in the MS/MS from the SPL generated using the Progenesis software.

The identifications of above differentially expressed proteins were all obtained using MS/MS from the DDA runs of samples. When MS-only data was used to identify ions differing in abundance between the two samples, both the ProfileAnalysis and Progenesis LC-MS software indicated that there were ~1000 features which changed reproducibly between the DNase extracts of NHUc and NHUd cells. The number of features differing between the control and DNase extracts of NHUc or NHUd cells was greater than was observed between the comparison of control extracts from NHUc and NHUd. Most of the differences between control and DNase extracts were proteins which were more abundant in the DNase extracted samples. Both these observations are consistent with the expectation that the control extractions from NHUc and NHUd should have contained similar proteins to one another, and that the DNase extractions from NHUc and NHUd cells should have contained proteins not present in the control samples due to the release of proteins by DNase. The potential for identification of these features was at first hampered by the length of time between the MS-only data collection and the generation of the SPL which resulted in a drift in ion retention times that rendered the retention times targeted in the SPL outside of the elution time of the target ions. SPL were then generated using Progenesis LC-MS software and elution times amended to account for the shift in retention times. These SPL lists identified some proteins which were not identified in the

original DDA, indicating a modest increase in the proteome which was targeted. However the shift in retention times, re-aligning the MS/MS data from the SPL did not prove successful. The problems with the alignment could have arisen from the need to inject more sample amount to improve the MS/MS on lower abundance peptides. This may have altered the intensity of the sample to such a degree that the software could no longer successfully time-align the samples. Only when the SPL can be aligned with the MS data will quantitation be possible, and the goal of assigning fold-changes to more of the proteins associated with differentiated or control cells be realised.

Failing further quantitation, proteins already identified as upregulating in DNase extracts from NHUd cells as compared to DNase extracts from NHUc cells warrant follow-up work to assess their function in NHU cells. DDX5, as a known CREBBP/SMAD3 coactivator is particularly intriguing.

Future studies using LF-MS to investigate nuclear complexes may consider:

- Isolating nuclei prior to extraction of nuclear proteins.
- Preparing chromatin as for chromatin immunoprecipitation (ChIP) for analysis by LF-MS.

Such studies should also show whether the purification achieved would be compatible with label free mass spectrometric workflows and whether quantification of low abundance proteins is achievable with these techniques.

5 Epigenetics and Chromatin Dynamics During *in vitro* Differentiation of Normal Human Urothelial Cells

5.1 Introduction

5.1.1 Chromatin Binding Proteins and Next-Generation Sequencing

Binding of transcription factors to DNA during cellular differentiation has been extensively studied, as exemplified by the role of PPARγ in adipogenesis (39, 88, 104, 112, 136). Chromatin immunoprecipitation (ChIP-seq) involves immunoprecipitation of formaldehyde fixed chromatin-DNA complexes followed by high-throughput sequencing of the isolated DNA (Figure 5.1.1).

ChIP-seq has allowed the binding sites of transcription factors such as PPARγ to be mapped in a genome-wide fashion, and the changes in binding sites over time to be tracked (39, 104). Similar to other transcription factors (186-189), ChIP-seq studies of PPARγ have revealed that the model whereby activated PPARγ binds to DNA and expression of surrounding genes increases, is too simplistic, highlighting that interactions between multiple transcription factors are required to bring about large-scale changes. Siersbæk *et al* (104) showed that the transcription factor CCAAT/enhancer-binding protein beta C/EBPβ binds to closed chromatin in pre-adipocytes around genes known to be upregulated during adipogenesis, and within 4 hours of induction of adipogenesis, the chromatin in these regions becomes DNaseI sensitive. The newly opened chromatin is bound by a variety of factors, including RXRα, GR and Stat5a. These factors maintain the open status of the chromatin, allowing PPARγ/RXRα heterodimers to bind. Even PPARγ/RXRα binding is not sufficient to drive gene expression immediately, suggesting still other factors may be involved in upregulation. This type of co-operative behaviour of transcription factors has been demonstrated, such as in the interplay between ERα, RARα and FOXA1 in driving ERα-induced gene expression in MCF-7 cells (187-189).

Recent large-scale studies of binding sites of transcription factors in multiple cell types, such as the one undertaken by the ENCODE consortium, has highlighted that multiple transcription factors bind the genome in site-specific and cell-type

specific combinations (38, 190-193). Multiple groups have used these results to research networking of transcription factors, and infer the influence of the interactions of individual proteins or complexes on gene expression. The key finding of Gerstein *et al* (38) was that transcription factors bind in a "combinatorial and context-specific fashion". The resulting "hierarchy" of transcription factors extracted from the data showed that some factors always appear to have more influence on gene expression than others, and that most bind in complexes consisting of preferred yet dynamic members. The complexity of these networks underscores the difficulty of making interpretations of the effects of transcription factor binding on gene expression, based on knowledge about the binding patterns of only a small number of sequence-specific transcription factors.
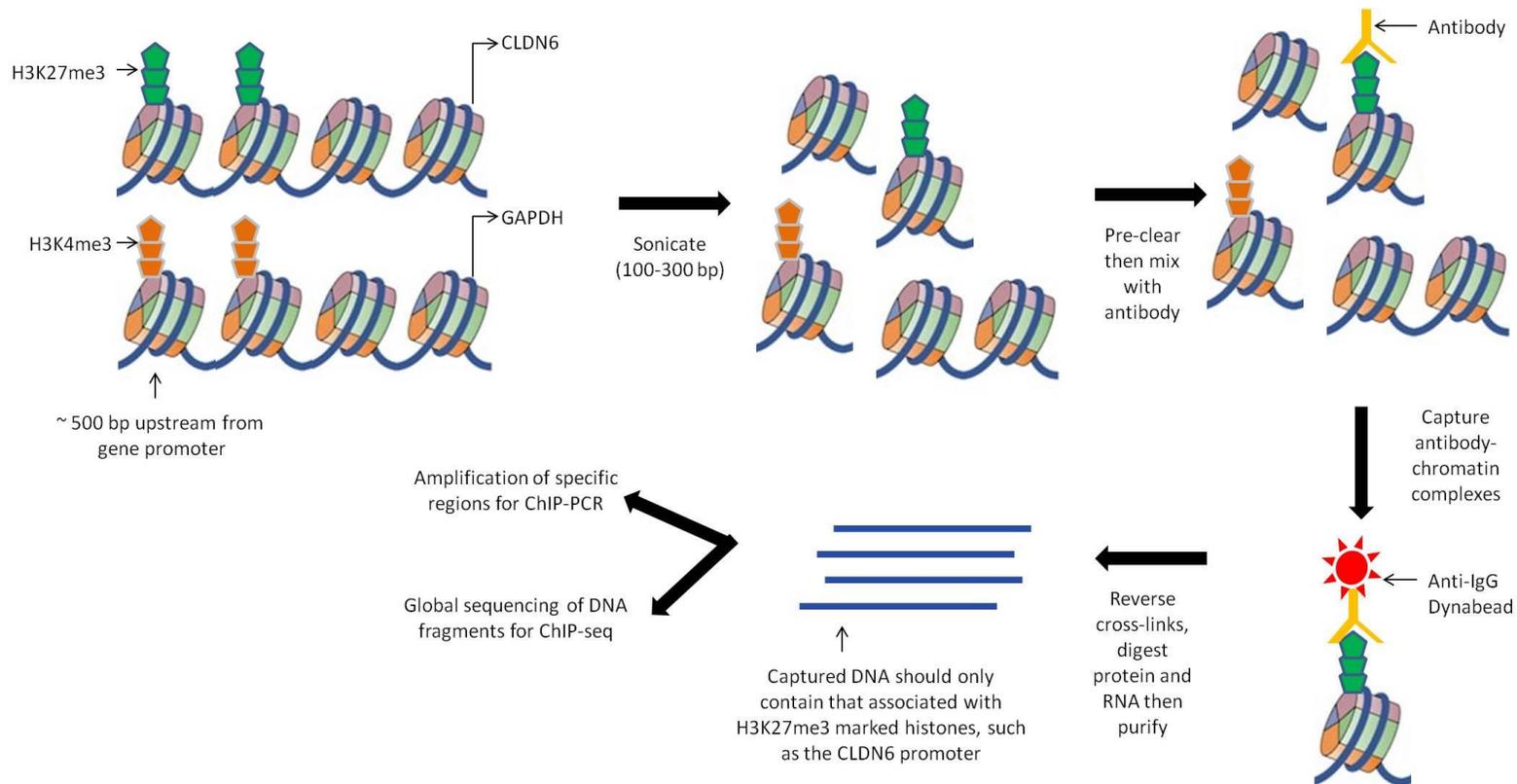
Figure 5.1.1 Chromatin immunoprecipitation workflow. Cross-linked cell samples are subjected to hypotonic lysis and sonication to fragment DNA. DNA-histone complexes are incubated with antibodies raised against specific histone modifications before capture of antibody-chromatin complexes with anti-IgG Dynabeads. DNA captured by antibodies can then be used for PCR (ChIP-PCR) or sequencing (ChIP-seq).

Upon binding to DNA, most transcription factors displace the nucleosomes which package the DNA (194). In addition, DNA-transcription factor interactions are transient, often leaving DNA that is both free of nucleosome and protein around the transcription factor binding site. This so called "nucleosome-depleted" DNA can be separated from nucleosomal DNA by sonication of formaldehyde-fixed cells. The protein-free nucleosome-depleted DNA can then be isolated by removal of protein-bound DNA via phenol purification. This method is known as formaldehyde-assisted isolation of regulatory elements (FAIRE) (195) (Figure 5.1.2).
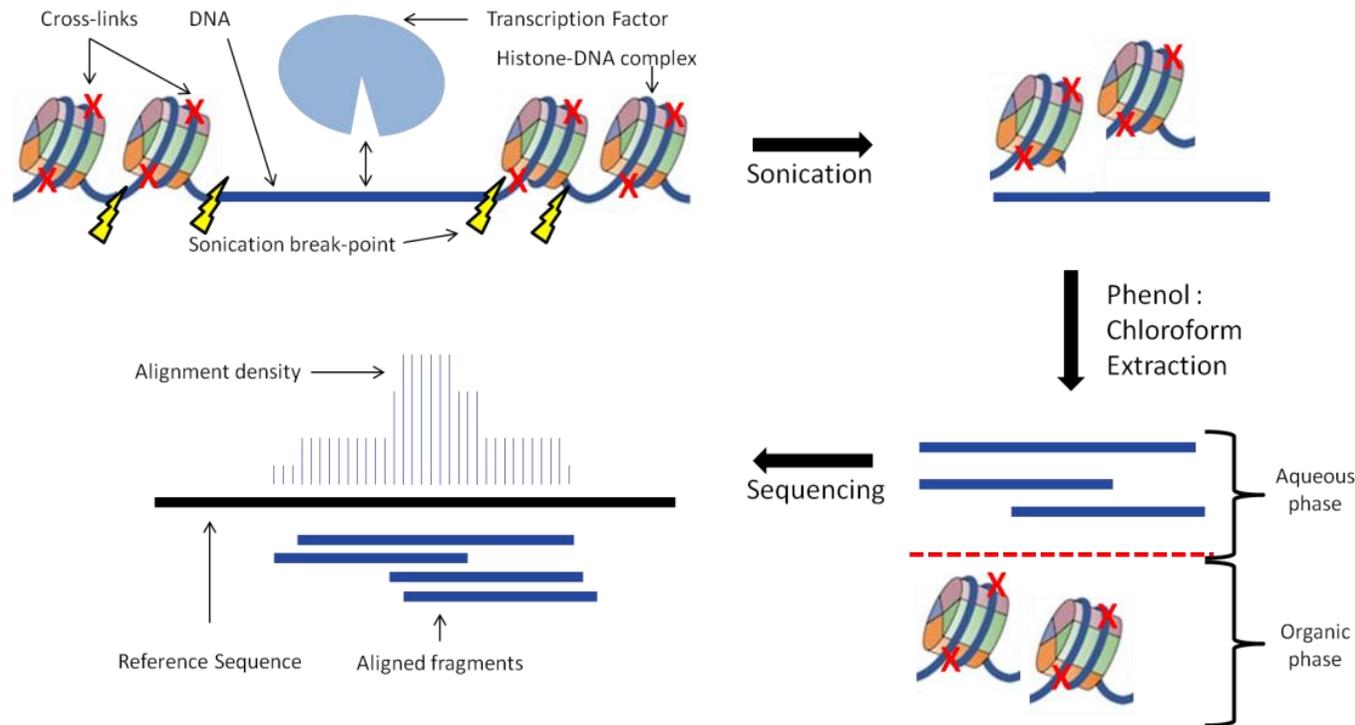
Figure 5.1.2. Formaldehyde-assisted isolation of regulatory elements (FAIRE) workflow. DNA in close proximity to protein at the time of fixation, such as that in nucleosomes, has many sites of interaction for potential cross-linking. Transcription factors have fewer contact points with DNA, and thus less chance for cross-linking. The DNA from nucleosome-depleted regions preferentially fragments from the nucleosomes in response to sonication, enabling it to be purified in the aqueous phase of a phenol: chloroform extraction. Sequencing of the isolated fragments allows identification of regions of the genome with enriched transcription factor binding.

FAIRE-extracted DNA can then be sequenced, and known transcription sequence motifs can be identified in the isolated DNA (195, 196). As DNaseI preferentially cleaves open chromatin or exposed DNA (194), treatment of nuclei with DNaseI also releases nucleosome-depleted regions of DNA. Short fragments where transcription factors were bound can then be isolated on sucrose gradients and sequenced (DNase-seq), allowing the genome-wide identification of regulatory elements in a similar fashion to FAIRE (197, 198).

FAIRE and DNase-seq have been compared in the same study by Song *et al* (40). A total of seven primary and carcinoma human cell lines were compared, with 30-40% of regulatory sites identified being common to each technique. DNaseI sites were found to be enriched within 2 kb of transcription start sites and 5' exons and introns, whilst FAIRE sites were found within genes and non-promoter intergenic regions. Differences between the two techniques are likely a result of complexes which affect the ability of DNaseI to access DNA, or where tightly DNA-bound factors cross-link to DNA as efficiently as histones. Despite these differences, >90% of high-confidence transcription factor motifs identified in other ChIP-seq studies using the same cell types were captured by both FAIRE and DNase-seq. Some specific differences in identification of transcription factor binding sites were seen, in that FAIRE fared better at identifying binding sites of FOXA1, FOXA3 and GATA3, whereas DNase-seq had greater success with ZNF263 sites. One of the major findings of this study was that most transcription start sites (TSS) in all cell types had open chromatin, and that the open chromatin regions which differed between cell types were mostly observed in intergenic regions around genes expressed specifically in each cell type. The authors proposed that transcription factor combinations which bind these open chromatin regions in specific cell types could be key to maintenance of cell identity, echoing the studies in adipocytes discussed previously. As such, understanding the combinations of factors which bind at open chromatin sites unique to specific cell types could help further understanding of the development and maintenance of cellular phenotypes.

Analysis of potential transcription-factor interactions *in silico* from ChIP-seq datasets has been investigated by Giannopoulo *et al,* using co-occurrence of binding in multiple datasets from ENCODE (199). The inferred complexes were composed of known interactors, with some complexes gaining members. Inferred complexes occurred in genomic regions with specific histone marks and were

placed at loci which matched the known functions of some member proteins, such as EP300 histone acetyltransferase enhancer complexes in regions distal to genes.

These studies highlight the potential with large-scale datasets to uncover detailed descriptions of chromatin composition at specific sites throughout the genome. Both FAIRE-seq and DNase-seq can be performed as part of existing next-generation sequencing workflows. Data from such experiments can be compared to existing datasets from the ENCODE project in order to identify unique chromatin features in the cell type studied. This brings the use of such approaches within the reach of the wider scientific community, although utilising either approach requires significant computing power and bioinformatics knowledge or support. One early example of FAIRE utilisation outside the ENCODE project was its application to studying adipogenesis, which allowed Waki *et al* to elucidate, *de novo,* the role of the transcription factor NFIA in development of adipocytes (93).

Combining the ability of FAIRE to identify multiple transcription factor binding sites from a single dataset with the approaches of Giannopoulo *et al* to infer binding complexes would represent a step-change in the understanding of how chromatin-bound complexes maintain cell-specific gene expression patterns.

5.1.2 Chromosome Conformation Capture

Chromatin in eukaryotic cells exists within the tightly packed space of the nucleus, yet this mass of proteins and DNA is not disorganised. Spatial organisation of DNA in the nucleus has been shown to be non-random, whereby chromosomes occupy distinct regions of the nucleus, dubbed "chromosome territories" (200-202). Within chromosomes, regions with similar transcriptional activity preferentially localise either with one another, or with specific structures such as the nuclear envelope (180). Actively transcribed and repressed regions within chromosomes co-localise respectively with active RNA polymerase transcription foci outside the territory of the specific chromosome studied (32), or the nuclear lamina (203). The drive to understand how genomes are organised has resulted in experimental approaches which have since been applied to wider areas of genome biology.

An approach to probing the physical interactions between genomically separated regions of DNA in the nucleus was initially developed for analysis of small

numbers of loci, using a technique called chromosome conformation capture (3C). 3C was developed by Dekker *et al* in order to investigate changes in DNA conformation around a bacterial gene after induction of gene expression (69). 3C isolates interacting genomic regions by cross-linking DNA using formaldehyde. DNA is then enzymatically cleaved using a site-specific restriction enzyme, leaving cross-linked protein-DNA and DNA-DNA complexes intact. Subsequent proximity ligation anneals cut-ends of DNA within the same cross-linked complex, allowing sequences which may have originated from differing genomic regions to form a contiguous sequence (Figure 5.1.3). Polymerase chain reaction (PCR) primers, designed to only give products where successful ligation of genomically separated regions has occurred, are then used to infer interactions.
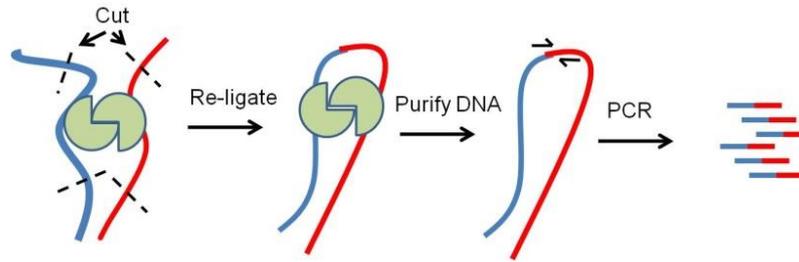
Figure 5.1.3. Outline of 3C procedure. DNA strands from distant genomic regions (red, blue) are brought together via their binding with two interacting proteins. These interactions are preserved during fixation, after which the unbound DNA is cut by nucleases. Re-ligation of the cut DNA ends allows the two strands of DNA to be made into one contiguous sequence. The specific interactions of regions around sites targeted by the nucleases can then be targeted by PCR (black arrows on DNA strands).

Developments of the 3C method have adapted the technique for use with DNA microarrays and next-generation sequencing technologies, allowing genome-wide analysis of interaction profiles of specific loci with the rest of the genome (33), analysis of regions which interact when bound to specific transcription factors (204) or interaction of any two ligated sequences (71).

Genome-wide binding maps of transcription factors show that they only bind the chromatin at a small percentage of their potential binding sites(94). Correlating binding sites with RNA expression can help to identify target genes, but it is not always the case that factors upregulate the gene to which they bind to most proximally (188). Transcription factors are able to interact with distant genomic regions thanks to the compact 3-dimensional nature of the nucleus. In order to investigate the changes in genome organisation imparted by estrogen-receptor alpha (ERα) activation, Fullwood *et al* developed a 3C based technique termed chromosome interaction analysis by paired-end tagging (ChIA-PET)(205). In ChIA-PET, DNA is fragmented and immunoprecipitated as for ChIP. ERα-bound DNA fragments, still in their complexes, are ligated with sequencing adapters before the ends are ligated together (Figure 5.4). These sequencing adapters contain target sites for a nuclease which cuts proximal enough to the site to allow sequencing of some bases of the two original DNA strands that were ligated together.
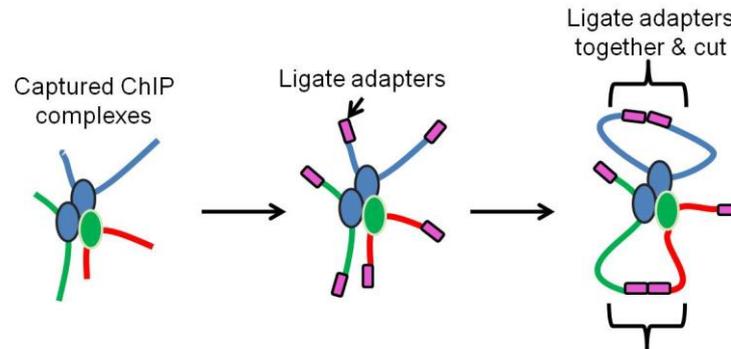
Figure 5.1.4. Outline of ChIA-PET procedure. Chromatin complexes containing DNA from genomically separate regions (blue/green proteins and blue/red/green DNA strands) are captured by chromatin immunoprecipitation (ChIP). Broken DNA ends are ligated with adapters, which are then ligated together. The adapters and some of the flanking sequence are cut by nucleases, making the strands amenable to sequencing with short-read next-generation sequencers.

The interaction between DNA from distant genomic regions mediated by ERα was then inferred by deep-sequencing of the resulting DNA fragments. These data were compared with known binding sites of other transcription factors thought to impact ERα driven expression, such as FOXA1. The analysis from the Fullwood *et al* (205) study indicated that ERα proteins interact chiefly with one another in clusters within chromosomes, and that interactions are greatly enriched between regions where FOXA1 binds within 10 kb of ERα. Modelling of the ERα binding sites and the upregulation of surrounding genes suggested that interacting ERα binding sites may serve as anchors around which to form large DNA loops, with genes in the periphery of the loops (away from the anchor) being less upregulated than those closer to the anchor. It is thought that the ERα proteins may interact with each other at sites where they also interact with the basal transcriptional machinery (206). The ChIA-PET approach is beneficial when studying a single transcription factor and its genomic targets, but provides little information about the portion of the genome not affected by the chosen factor.

High-throughput chromosome conformation capture (HiC) adopts a global approach to analysis of interacting DNA, as opposed to isolating only those in complex with a chosen transcription factor (71). HiC is based on 3C, but instead of direct re-ligation of digested DNA, biotinylated nucleotides are introduced into overhangs prior to ligation (Figure 1.5). Biotinylation facilitates purification of re-

ligated DNA junctions, greatly increasing the number of sequenced fragments which arise from ligated DNA (70). Genome-wide maps of interaction frequencies can be produced from HiC data, showing regions with enriched and depleted interactions. The HiC approach was successfully used by Rickman *et al* to identify the regions of the genome which altered conformation in response to over-expression of the oncogene ERG in the normal benign human prostate cell line RWPE1 (207). ERG binding sites, as detected by ChIP-seq, were significantly enriched within regions which altered their HiC interaction profiles. In addition to this, 65% of the 1,266 genes differentially regulated in cells over-expressing ERG were in regions which interacted with one another. Within this gene set were the HOXA, HOXB and HOXC gene clusters, which are important in urogenital development and known to be downregulated in prostate carcinoma cells under the influence of the polycomb protein EZH2 (208). Conversely, some genes interacted less in ERG-transfected cells, yet had increased expression, indicating a possible release from repression complexes. HiC data allowed Lin *et al* to observe the shift in interaction profiles of genes associated with B-cell development during lineage progression (209). Genes known to be involved in orchestrating B-cell fate were observed to change genomic position, inferred from a shift in interaction partners, early in the differentiation process. Lin *et al* also observed collaborative binding of transcription factors which appeared to bind in distinct patterns that changed during differentiation.
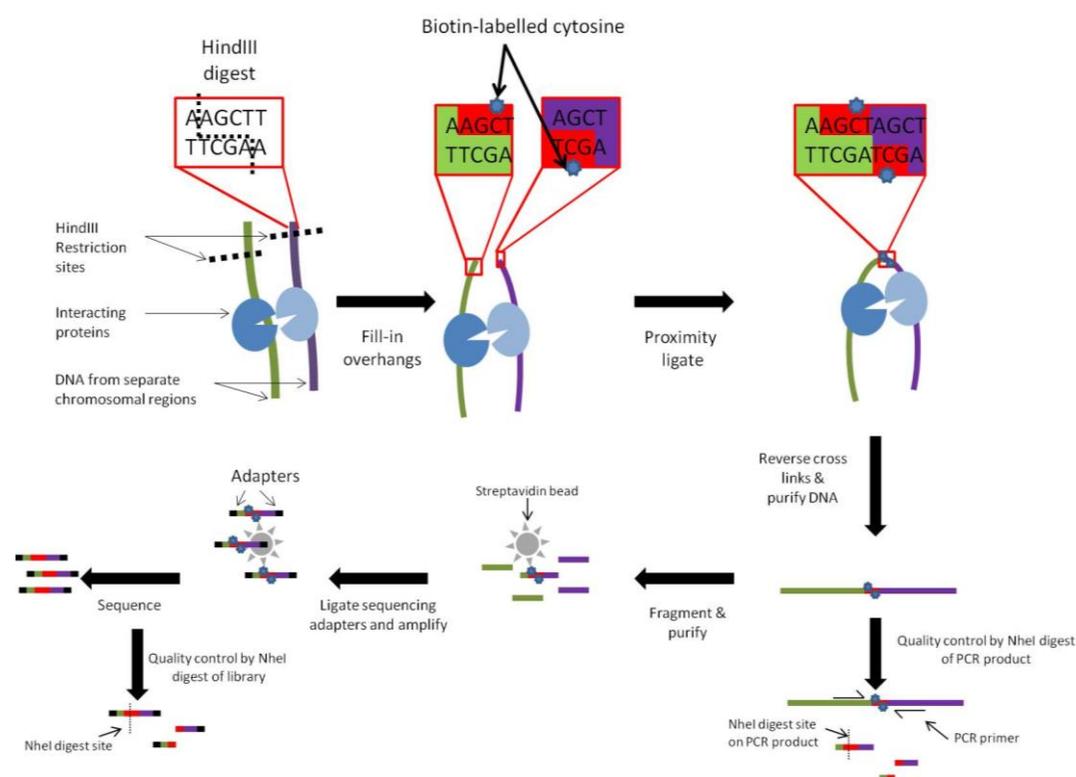
Figure 5.1.5. Construction and quality control of HiC libraries. HindIII digestion allows interacting DNA regions captured by fixation to be excised. The DNA overhangs left by the digestion can then be filled in, with the inclusion of biotinylated cytosine. Subsequent ligation allows DNA within complexes to be ligated together. PCR amplification of ligated junctions and NheI digestion of PCR products serves as quality control for incorporation of nucleotides at the original HindIII digest site. HiC libraries can then be fragmented and biotinylated ligation junctions streptavidin purified prior to ligating next-generation sequencing primers and library amplification. As a final quality control, an aliquot of the amplified library can be digested using NheI restriction enzyme which recognises the sequence present at ligation junctions which were successfully filled in.

One disadvantage of genome-wide 3C is that the data represents an average of the whole population of cells studied. The 3C derived "enhanced 4C" (e4C) technique was developed by Schoenfelder *et al* (33) to probe interactions of a specific locus during transcription. e4C entails pre-enriching for actively-transcribed genes using ChIP with an RNA polymerase II antibody before digestion and ligation of DNA ends. A specific locus is then targeted by using a primer extension step with a biotinylated primer to allow recovery of the target DNA before amplification and hybridisation to microarray. e4C showed that the interactions between genes in the mouse globin locus had preferred interactors, but the sheer number of interactors observed for a single locus suggested a large plasticity in the conformation of the genome across the cell population. Alone, high-throughput 3C technologies such as HiC may be limited to describing interactions of DNA without any functional inferences beyond changes to transcriptional activity. However, combined with maps of chromatin-binding proteins it can provide significant insight into the key regions of the genome which change their activity in response to physiological stimuli.

5.1.3 Epigenetic Studies in Urothelium

Most work carried out so far regarding epigenetics in urothelium concerns carcinogenesis, in which there is evidence for disruption of both histone modifications and DNA methylation. Nishiyama *et al* carried out DNA methylation studies comparing urothelia obtained from normal urothelium, non-cancerous urothelium from patients with urothelial carcinoma (UC) and samples of urothelial carcinoma (210). DNA was enriched for methylated sequences by digesting non-methylated DNA using methylation-sensitive nuclease SmaI prior to hybridisation to microarrays. Most non-cancerous urothelia from patients with UC had methylated regions of chromosomes similar to UC samples, but distinct from normal samples. Methylation patterns within UC samples clustered into two groups, corresponding to higher and lower invasive potential. The low resolution of the arrays only allowed identification of regions spanning whole chromosomal bands as being differentially methylated. As such, although the arrays were able to predict the presence of cancer and determine the invasive potential, little about the specific genes that were methylated could be inferred.

Radvanyi *et al* used RNA microarrays to profile gene expression in bladder cancers with a range of stages and grades (211). Following previous studies which showed contiguous regions of chromosomes with aberrant gene expression in cancers (212, 213), Radvanyi *et al* sought to identify chromosomal regions which contained neighbouring co-regulated genes expressed differentially in carcinoma samples compared to normal samples. Clustering resulted in identification of seven regions consistently silenced in 46% of tumours. This sub-group of tumours was identified as having "multiple regions of epigenetic silencing" (MRES), and represented 100% of the original cohort which had the previously defined carcinoma *in situ* (CIS) gene expression phenotype associated with muscle invasive disease (214). In a cell line derived from one of the MRES tumours, the DNA demethylating agent 5-aza-deoxycytidine had no effect on expression of genes in the silenced regions, whereas the histone deacetylase (HDAC) inhibitor trichostatin-A (TSA) increased their expression. Repressive histone methylation marks (trimethylation of histone H3 on lysines 9 and 27) were found to decrease in MRES regions upon TSA treatment of cell lines. The increase in acetylation and decrease in methylation after TSA treatment indicated that inhibition of HDAC proteins allowed histone demethylase and histone acetylase proteins to associate with the chromatin in these regions.

Several groups have described differential methylation of promoters of genes known to be altered in expression in CIS such as p53 (215, 216), and linked this to the altered expression of the DNA methyltransferase 1 protein (DNMT1) identified by Nakagawa *et al* (217). Dudziec *et al* , having previously associated DNA methylation with carcinogenesis (218), sought to identify a relationship between DNA methylation, histone methylation and gene expression (219). This was achieved by profiling gene expression and DNA methylation in cultured normal urothelial cells and two malignant urothelial cell lines using microarray approaches. Expression of genes was compared with genome-wide profiles of histone H3 trimethylated at lysine 9 (H3K9me3) or at lysine 27 (H3K27me3). H3K27me3 was found to occur around genes with low expression in all cells, whereas H3K9me3 was only weakly associated with repression in a subset of genes with DNA methylation. DNA methylation itself was only weakly associated with repression of expression in malignant cell lines.

The above studies in urothelial carcinomas focus on DNA methylation, histone modifications or both. However, no two groups were studying the same sets of

genes or chromosomal regions, or were comparing similar samples. There is, however, some correlation in their evidence, which shows the separate influences of histone modification at H3K27me3 and DNA methylation on gene expression in carcinoma. Although Dhawan *et al* and Nishiyama *et al* showed that DNA methylation may be predictive of carcinoma in apparently normal tissue (218), Dudziec *et al* showed that repressed genes in urothelial carcinoma have hypomethylated promoters, as corroborated by Radvanyi *et al*. The lack of DNA methylation at genes in MRES regions seen by Radvanyi *et al* was corroborated by Nishiyama *et al* who observed hypomethylation in carcinoma samples at two of the three MRES chromosomal regions they also identified as differing between normal and carcinoma samples (17q21 and 14q11.2). Major oncogenes found by Dhawan *et al* to have hypermethylated DNA in promoters in CIS samples (p53, E cadherin) are not associated with MRES silenced regions. Interestingly, Rassf1a, which was found by Dhawan *et al* to be hypomethylated in CIS patients, lies within 3p21.31, adjacent to an MRES region (218).

One missing aspect from all of these studies is the genome-wide study of epigenetics in differentiated normal tissues. Proliferating and confluent cells, as used by Dudziec *et al* (219) and Dhawan *et al* (218), will likely have different gene expression profiles from differentiated NHU cells. It would therefore be informative to understand the change in gene expression, and associated epigenetic changes, undertaken by urothelial cells during differentiation. The difficulty of performing such studies in the past has in part been due to the difficulty of obtaining sufficient DNA for sequencing, as ChIP-seq isolates relatively small percentages of the genome and typically requires a minimum starting quantity of 2×10^7 cells per experimental condition to obtain a yield sufficient to allow sequencing. This is prohibitive if the target cells are rare, or the cell line is finite, as is the case with NHU cells. However, building on protocols developed for the unbiased amplification of total genomic DNA (220), ChIP-seq performed on DNA recovered from as few as 10,000 cells can recapitulate the data obtained from standard ChIP-seq (221, 222).

Combining ChIP-seq with FAIRE, HiC and genome-wide RNA expression studies (RNA-seq) will provide new insight into the epigenetic and chromatin binding events occurring during differentiation of NHU cells. This will require using similar approaches undertaken by Rickman *et al* and Lin *et al* when correlating HiC data with ChIP-seq and gene expression data. To infer the influence of

changes in regulatory regions identified by FAIRE on chromosome conformation, the FAIRE peaks unique to differentiating cells will have to be identified. Integrating all of this with an understanding of the effects of specific gene complexes, as done by Giannopoulo *et* al (199), will help to pinpoint which protein complexes target specific regions of the genome intrinsically important to urothelial cell identity.

5.2 Experimental Aims and Approach

5.2.1 Aims

The switch between proliferation and differentiation in cultured normal human urothelial (NHU) cells provides an experimental model for understanding what epigenetic mechanisms are involved in altering the transcriptional repertoire during *in vitro* differentiation.

It was hypothesised that the repressive histone methylation marks in control NHU cells at 24 h would mirror that in MRES regions of carcinoma samples, and that this methylation would change during differentiation.

It was the aim of the work described in this chapter to utilise next-generation sequencing technologies to discover what changes in histone modifications, gene expression, chromatin binding proteins and chromosomal conformation occur during the *in vitro* differentiation of normal human urothelial cells. This data was then to be used to infer changes in chromatin binding complexes at different genomic loci, and the influence of such complexes on histone modifications, gene expression and chromosomal interactions.

5.2.2 Experimental Approach

5.2.2.1 *TGAC Collaboration*

Collaboration was set up with The Genome Analysis Centre (TGAC, Norwich, UK) under round five of the BBSRC Capacity and Capability Challenge (CCC). The project outline was for TGAC to provide histone modification ChIP-seq, HiC, FAIRE and RNA-seq data from samples prepared from *in vitro* differentiating normal human urothelial cells. RNA from differentiation-induced and control NHU cells at two time points, 24 h and 144 h, were to be sequenced by RNA-seq from triplicate donor NHU cell lines. ChIP-seq, FAIRE and HiC were performed in differentiation-induced and control cells at 24 h and 144 h on a single donor line which was one of the samples submitted to RNA-seq (Table 5.2.1).

	NHU Donor 1				NHU Donor 2				NHU Donor 3			
Time-	24 h		144 h		24 h		144 h		24 h		144 h	
Status→	Ctrl	Diff	Ctrl	Diff	Ctrl	Diff	Ctrl	Diff	Ctrl	Diff	Ctrl	Diff
RNA-seq	X	X	X	X	X	X	X	X	X	X	X	X
FAIRE	X	X	X	X								
ChIP-seq	X	X	X	X								
HiC	X	X	X	X								

Table 5.2.1. Next generation sequencing schedule. Three NHU cell lines from independent donors were cultured for 24 h and 144 h under differentiation inducing (TZ&PD) or control (0.1% DMSO) conditions. Diff= differentiating, Ctrl = control. Material from one cell line was subjected to RNA-seq, FAIRE, HiC and ChIP-seq targeting two histone modifications. Two other cell lines were subjected to RNA-seq to allow confident identification of genes which change in expression over time during differentiation.

RNA-seq of three independent donor cell lines was performed to allow genes which are upregulated during differentiation across donors to be ascertained. These transcription patterns were then intended be used to provide a baseline of expression changes which occur around features including transcription factor binding, histone modifications and chromosome interaction profiles identified in FAIRE, ChIP-seq and HiC datasets respectively.

All cell culture and HiC, FAIRE and ChIP-seq DNA library construction was carried out at the University of York. DNA libraries were, by necessity, sequenced at TGAC. Quality control carried out by TGAC is included where possible. Bioinformatics assessment was carried out in full collaboration with TGAC.

5.2.2.2 *Cell Culture*

NHU cells from three donors (Y967, Y1192 and Y1214) were cultured in 10 cm dishes with KSFMc medium until 70% confluent. Cells were then treated with differentiation-inducing or control agents for 24 or 144 h as described in Materials and Methods 2.3.4.

To obtain samples for HiC, FAIRE and ChIP-seq, differentiated and control cells were fixed at 24 h and 144 h by drop-wise addition of 37% formaldehyde to growth medium to a final concentration of 1%. Formaldehyde was added to dishes whilst on an orbital shaker and left for 10 min, before being quenched by addition of a

1.25 M glycine solution to a final concentration of 125 mM. Parallel cultures were harvested in 2% (w/v) SDS for western blotting or Trizol™ (Life Sciences) for protein and RNA-seq respectively (Materials and Methods 2.4 and 2.7.5 respectively).

Successful induction of differentiation was demonstrated by western blot analysis of reduction in phosphorylated pERK1/2 knockdown at 24 h and upregulation of claudin 5 and FOXA1 relative to control at both time points. Cells from each donor were cultured and harvested in parallel to reduce introduction of sample variability.

5.2.2.3 *RNA-seq*

RNA was prepared for sequencing by adapter primer ligation and subjected to 25 bp single reads. Reads were mapped to the genome using *maq* (maq.sourceforge.net/). Reads were assigned to ensemble transcript ID (ENST) and differential expression calculated at TGAC using the DEseq method and "per-condition" dispersion values (223).

5.2.2.4 *ChIP-Seq*

Sonicated chromatin from 1 X 10 cm dish of formaldehyde fixed cells (4 x 10^6 cells) was diluted as required in RIPA buffer and subjected to ChIP using antibodies targeting either histone H3 trimethylated at lysine 4 (H3K4me3), histone H3 trimethylated at lysine 27 (H3K27me3), total histone H3 (positive control) or anti-rabbit IgG (negative control) (Materials and Methods 1.7.7.2).

Constitutively repressed or active genes have histone modifications H3K27me3 or H3K4me3, respectively, in regions proximal to their promoters in multiple human cell types (as reviewed in (224)). Thus, the ChIP-seq results would provide evidence for regions undergoing epigenetic changes related to transcription during differentiation.

All anti-histone modification antibodies were rabbit monoclonal ChIP grade antibodies (Cell Signaling) recommended to be used at a 1:50 dilution. Control anti-rabbit IgG (Santa Cruz) was used at 2 μg per control ChIP.

To confirm the success of ChIP using the Dynabeads system to active mark H3K4me3 and repressive mark H3K27me3, PCR was performed using 1 µL of 20 µL column eluate after isolation of DNA using a Qiagen PCR purification kit. PCR targets were regions 500 bp upstream from genes known to be constitutively silenced or expressed in human urothelium.

As ChIP to epigenetic marks which are relatively rare across the genome is unlikely to yield sufficient DNA for sequencing when using 1-2 X 10^6 cells starting material, the libraries were amplified using the SEQX kit as described in

5.2.2.5 *High Throughput Chromosome Conformation Capture*

Protocols for construction of HiC libraries published by Belton *et al* (70) were adapted for use with NHU cells (Materials and Methods 2.7.6). 25 X 10^6 fixed cells were homogenised by passing 10 times through a 21-gauge needle. Cells were then nuclease digested with HindIII enzyme, which leaves four-base 5' sticky end overhangs. The overhangs were filled in using endonuclease activity of Klenow polymerase fragment in the presence of adenine, guanine, thymine and biotinylated cytosine. The newly created DNA blunt ends were re-ligated under dilute conditions to promote ligation to DNA strands proximal in space to one another, such as those formaldehyde fixed in the same DNA-protein complex. Once ligation occurs, contiguous sequences are created from DNA strands that were interacting at the time of fixation, yet may have originated from distant genomic regions.

Successful digestion and ligation was demonstrated using PCR which would only be successful if distant HindIII sites had been brought together to form ligation junctions. The extent of biotinylation at ligated junctions directly affects the quality of HiC data, and was therefore estimated by digesting the PCR product with the NheI enzyme which specifically digests the newly formed sequence at fully filled in blunt-end ligation junctions.

HiC DNA libraries which demonstrated biotin incorporation were then treated with exonuclease to digest biotin from unligated fragments before shearing and streptavidin purification. Streptavidin-purified DNA was then adapted with sequencing primers and amplified. As a final quality control, amplified libraries

were subjected to digestion with NheI, as successful digestion is predictive of quality of HiC data (70). Libraries were subjected to 50 bp paired-end reads.

5.2.2.6 *Formaldehyde Assisted Isolation of Regulatory Elements (FAIRE)*

NHU cells were sonicated as for ChIP-seq and nucleosome-depleted regulatory elements were isolated from supernatants by phenol: chloroform extraction, as in Giresi *et al* (195).

DNA was size-selected to 250 bp using XP beads (Beckman Coulter). DNA was then subjected to a standard ChIP-seq workflow of end repair and adapter primer ligation before sequencing.

5.3 Results

5.3.1 Differentiation of Cell Lines

Protein samples harvested from the three cultured NHU cell lines showed that induction of differentiation using TZ and PD was associated with increases in differentiation markers claudin 5 and FOXA1 at both 24 h and 144 h relative to control cells (Figure 5.3.1).
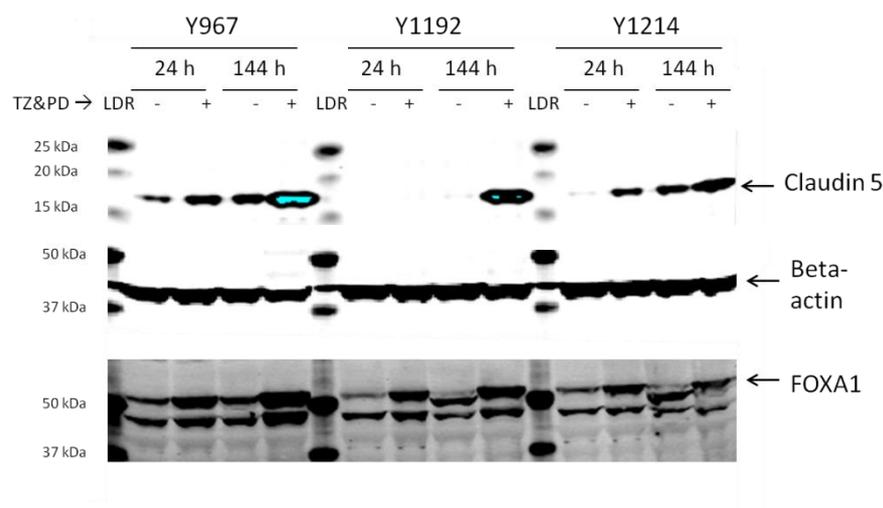
Figure 5.3.1. Upregulation of urothelial protein differentiation markers in NHU cell lines used for sequencing. Claudin 5 and FOXA1 expression was more abundant after treatment with differentiation-inducing agents (TZ&PD +) at both time points in all donor lines as compared to 0.1% DMSO control cells (TZ&PD -). pERK expression was not detected in differentiated cells at 24 h, but was variable at 144 h. Variable pERK expression did not affect upregulation of claudin 5 or FOXA1. Beta-actin shows comparable loading between lanes, blue colour of bands in Claudin 5 (e.g. 144 h TZ&PD +) indicates saturation of the detector. LDR = protein marker ladder.

5.3.2 RNA Quality Control - Bioanalyzer

RNA integrity is an important factor in obtaining reliable results from next-generation sequencing, and therefore was assessed before submitting samples for sequencing. Microfluidic separation of RNA in a Bioanalyzer RNA chip (Agilent) allows quantitation by measuring RNA fluorescence after laser excitation, and size-measurement by comparison to known size nucleotides. 18S and 28S ribosomal RNA are the most abundant RNA species in mammalian cells. For RNA to be good quality, peaks from both of these species should present and separated by a region of low noise. The algorithm contained within the Bioanalyzer software interprets the appearance and ratio of these peaks, and returns an RNA integrity number (RIN) of 0-10, with 10 being highest quality (68).

All RNA samples isolated from NHU cells were of very good quality, returning RIN values of 9.7-10 (Figure 5.3.2). RIN numbers are not predictive of quality of experimental results, but can be used as quality control.
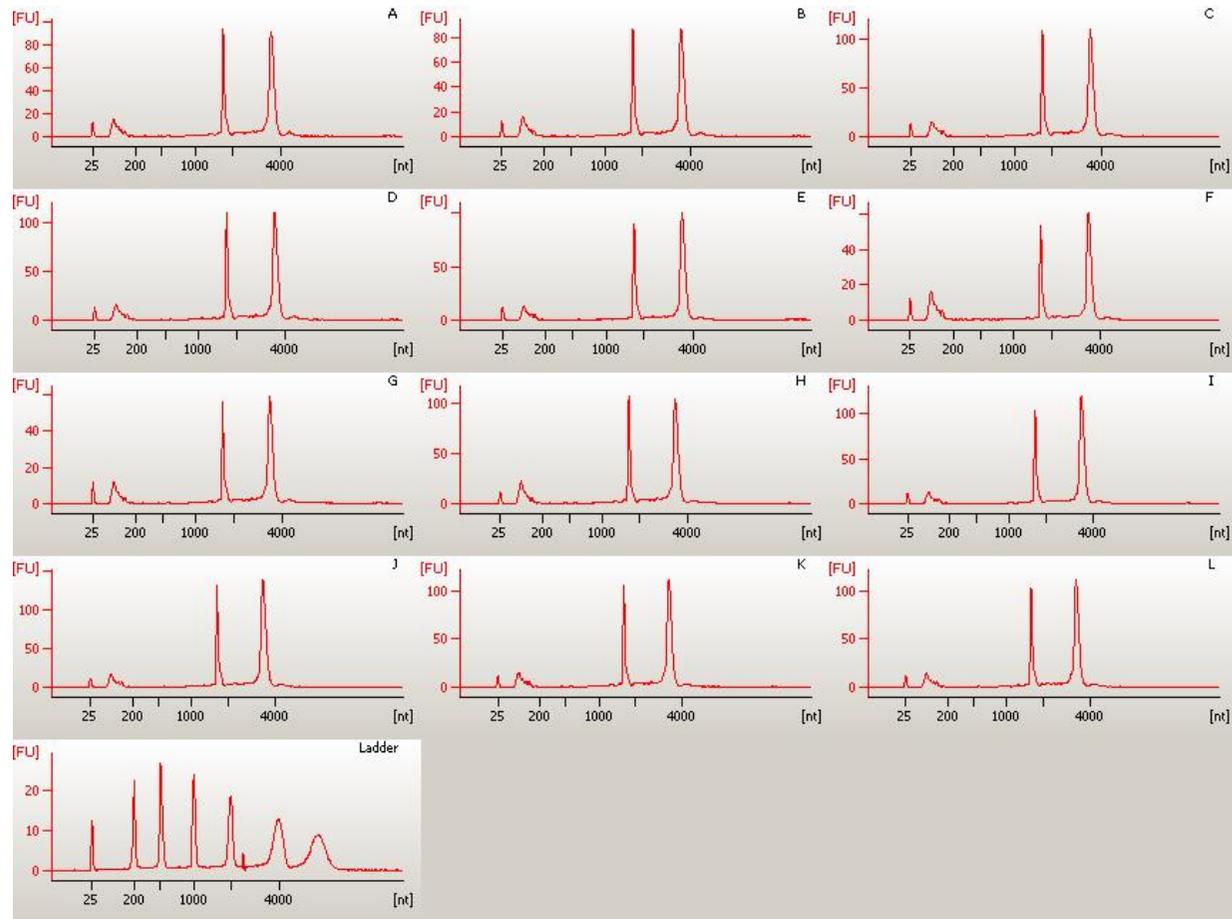
Figure 5.3.2. Bioanalyzer traces of RNA isolated from NHU cells for RNA-seq. Donor line Y967 (A-D), Y1192 (E-H) and Y1214 (I-L) treated with control or differentiation-inducing agents for 24 or 144 h respectively. Trace K shows the electropherogram of the standard size-ladder (increasing size L-R).

5.3.3 Optimisation of formaldehyde fixation

Formaldehyde fixation of NHU cells was tested by observing the appearance on western blots of cross-linked histone complexes after fixation for 0, 5, 10 or 30 min in 1% formaldehyde. Extra H3 reactive bands not present in non-fixed extracts were visible in cell extracts which had been fixed for 10 min (Figure 5.3.3). After 30 min, no further new bands appeared by comparison with the 10 minute treatment. Therefore, cross-linking was judged to be complete by 10 min and this time was used for all cross-linking experiments.
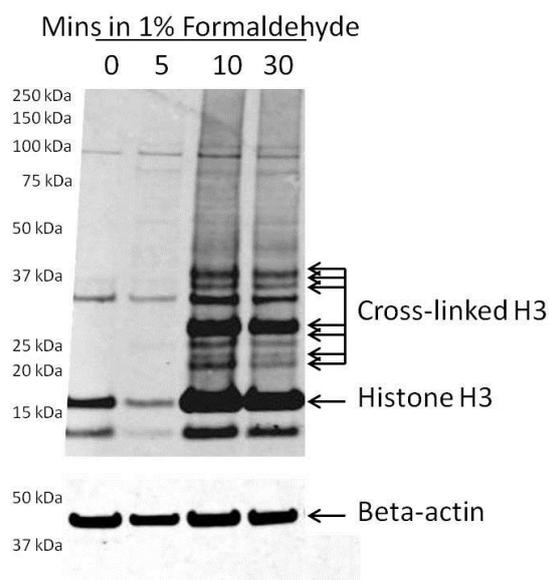
Figure 5.3.3. Formaldehyde fixation of NHU Cells. Appearance of additional histone H3 reactive bands relative to non-cross-linked samples was apparent after 10 min fixation in formaldehyde. After 30 min, no additional bands were seen. Cells were harvested in 2% SDS and 20 µg subjected to western blot as described in Materials and Methods 2.4, with the exception that samples were not heated to 70°C during preparation for SDS-PAGE in order to preserve cross-links.

5.3.4 Sonication of DNA for ChIP-Seq and FAIRE

Fragmentation of DNA is required allow efficient capture of DNA-protein complexes in ChIP. Sonication to mononucleosomal size (100-300 bp) is recommended for ChIP-seq targeting histone modifications (225). This increases the accuracy of mapping, by breaking target nucleosomes apart from chromatin from distal genomic regions with which they may have been interacting at the time of fixation.

To optimise the sonication process, NHU cells were subjected to sonication for increasing lengths of time in a Bioruptor (Diagenode) and samples taken to assess the DNA size profile (Figure 5.3.4). Sonication for 45 min showed near complete fragmentation to mononucleosomal size. 45 min sonication was thus performed on all subsequent samples, and size profiles of DNA checked visually on gels. If evidence of high molecular weight DNA was observed, samples were rejected (Figure 5.3.5).
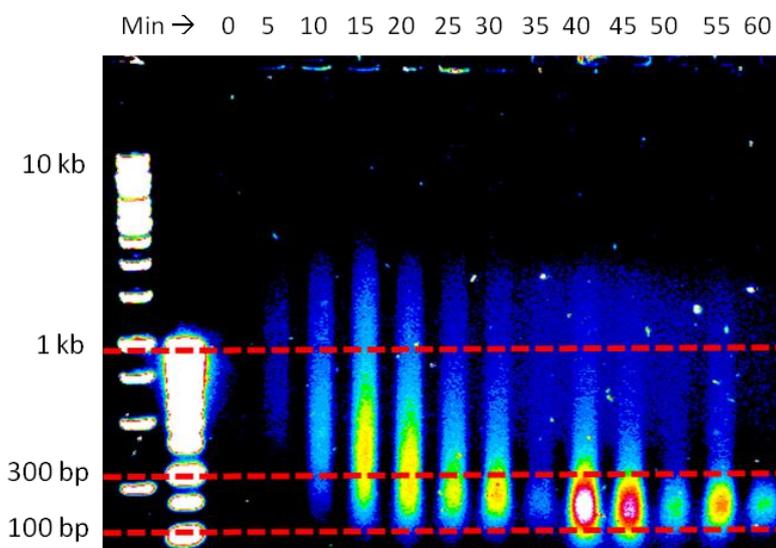
Figure 5.3.4. Optimisation of DNA sonication. For sonication, 1 X 10 cm dish of fixed NHU cells was prepared as in Materials and Methods 2.7.7.1. For optimisation, aliquots were taken after every 5 minutes sonication. DNA from aliquots was purified and electrophoretically separated in a 0.75% agarose gel. Pseudocoloured gel image shows DNA fragmented to 100-300 bp after 45 minutes sonication.
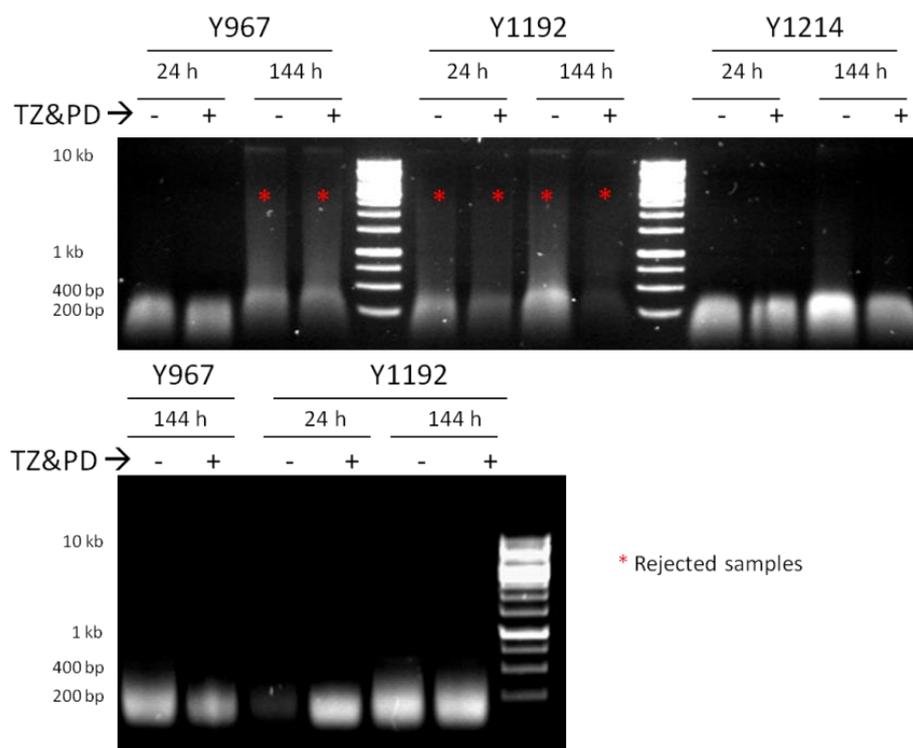
Figure 5.3.5. Sonication of NHU cells for ChIP-seq and FAIRE. Samples were sonicated for 45 min as in Figure 5.3.4 and aliquots taken after final round of sonication. If DNA did not show mononucleosomal size distribution of 100-300 bp, they were rejected (samples marked with *, upper image) and fresh samples prepared (lower image).

5.3.5 ChIP-seq Sample Preparation

ChIP-isolated DNA was subjected to PCR (ChIP-PCR) to demonstrate successful isolation of H3K27me3 and H3K4me3-associated regions. H3K27me should be associated with regions upstream of repressed genes, and H3K4me3 should be associated with regions around active transcription start sites (224).

Claudin 6 is not expressed in urothelial cells at any stage of *in vitro* differentiation (2) and GAPDH is constitutively expressed in all human cell types, although to varying degrees (72). Thus, for H3K27me ChIP-PCR in NHU cells, no PCR products should be observed with PCR targeting actively transcribed genes such as GAPDH, but products should be observed when targeting repressed genes such as claudin 6. The opposite would be expected for H3K4me3 ChIP-PCR. This was borne out in the results of the ChIP-PCR, which showed the expected results and was consistent for ChIP material from differentiated and control cells (Figure 5.3.6).
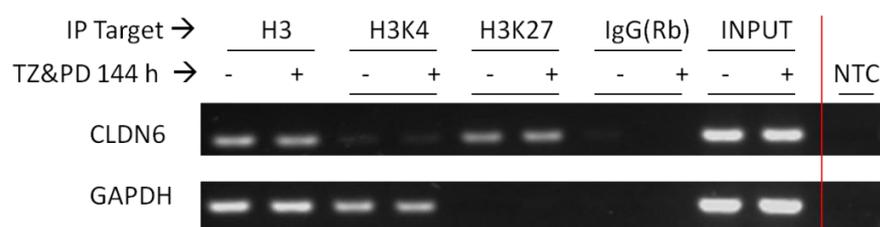
Figure 5.3.6. ChIP-PCR for H3K4me3 or H3K27me3. Histone H3 ChIP-PCR showed amplification of both CLDN6 and GAPDH, confirming that at least some of the DNA at these regions is associated with nucleosomes. Claudin 6 PCR resulted in relatively more product in H3K27me ChIP than H3K4me3 ChIP, as would be expected for a repressed gene. GAPDH PCR was successful in H3K4me3 ChIP and failed in H3K27me3, as would be expected from a ubiquitously expressed gene. PCR performed using eluates from anti-rabbit IgG were negative in all samples other than a trace amount of product in the CLDN6 144 h control ChIP. Input DNA from ChIP was used as a positive control and nuclease free H_2O as a no template control (NTC). Chromatin from 1 X 10 cm dish of 144 h differentiated and control samples from cell line Y1214 was sonicated, diluted in RIPA buffer into 4 X 600 μL aliquots and 1 X 50 μL aliquot as input sample. CLDN6 PCR was performed for 34 cycles and GAPDH PCR for 36 cycles.

ChIP was repeated for H3K4me3 and H3K27me3 for cell line Y1192 with 2×10^6 cells per IP and isolated DNA was subjected to amplification using the SEQX kit (Sigma) as per the manufacturer's instructions.

Two generic fragmented DNA libraries, donated by TGAC, were diluted to 100 pg and 1 ng total DNA and used as test subjects for amplification; these concentrations represent the extremes of the designated working range of DNA for amplification using the SEQX kit. 1 ng and 100 pg test libraries showed amplification curves which entered a linear phase at cycles 14 and 16 respectively (data not shown). NHU Chip-seq samples were then subjected to the SEQX workflow. Libraries amplified at similar time-points to the test libraries, indicating successful amplification (Figure 5.3.7).
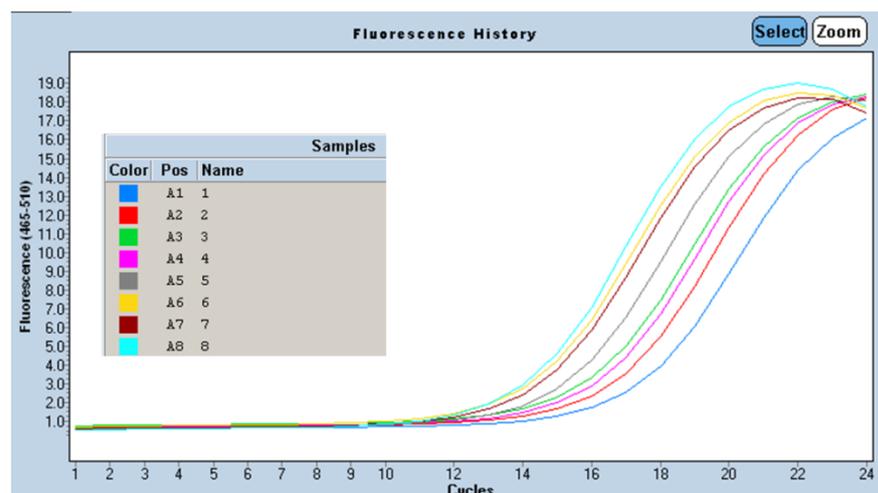
Figure 5.3.7. SEQX Amplification of ChIP samples. Screenshot of fluorescence (y-axis) against PCR cycles (x-axis) showing amplification of Y1192 ChIP samples. Samples are 24 h control, 24 h differentiated, 144 h control and 144 h differentiated ChIP samples of H3K4me3 (1-4) and H3k27me3 (5-8). Amplification was to be stopped 2-3 cycles into the plateau phase of PCR. Samples 1-4 (all H3K27me3) reached the end of plateau just as samples 5-8 were entering the plateau phase. As such, samples 1-4 were removed and submitted to the final extension step, and samples 5-8 were subjected to a further 3 rounds of amplification before extension in a separate machine. PCR was set up using components in the SEQX kit and amplification detected by fluorescence emission of SYBR Green, included at 1 in 75,000 dilution of SYBR Green in the PCR reaction. SYBR Green fluoresces when it intercalates with double-stranded DNA, thus the amount of fluorescence increases with the quantity of DNA.

DNA was purified using PCR purification kit (Qiagen), and then primers were enzymatically cleaved from the DNA using SEQX components before a second PCR purification. The size of libraries was checked by separation of 1 ng on a Bioanalyzer (Agilent) microfluidic electrophoresis instrument (Figure 5.3.8). Libraries were then subjected to ChIP-seq workflow as detailed in Materials and Methods 1.7.8.3.
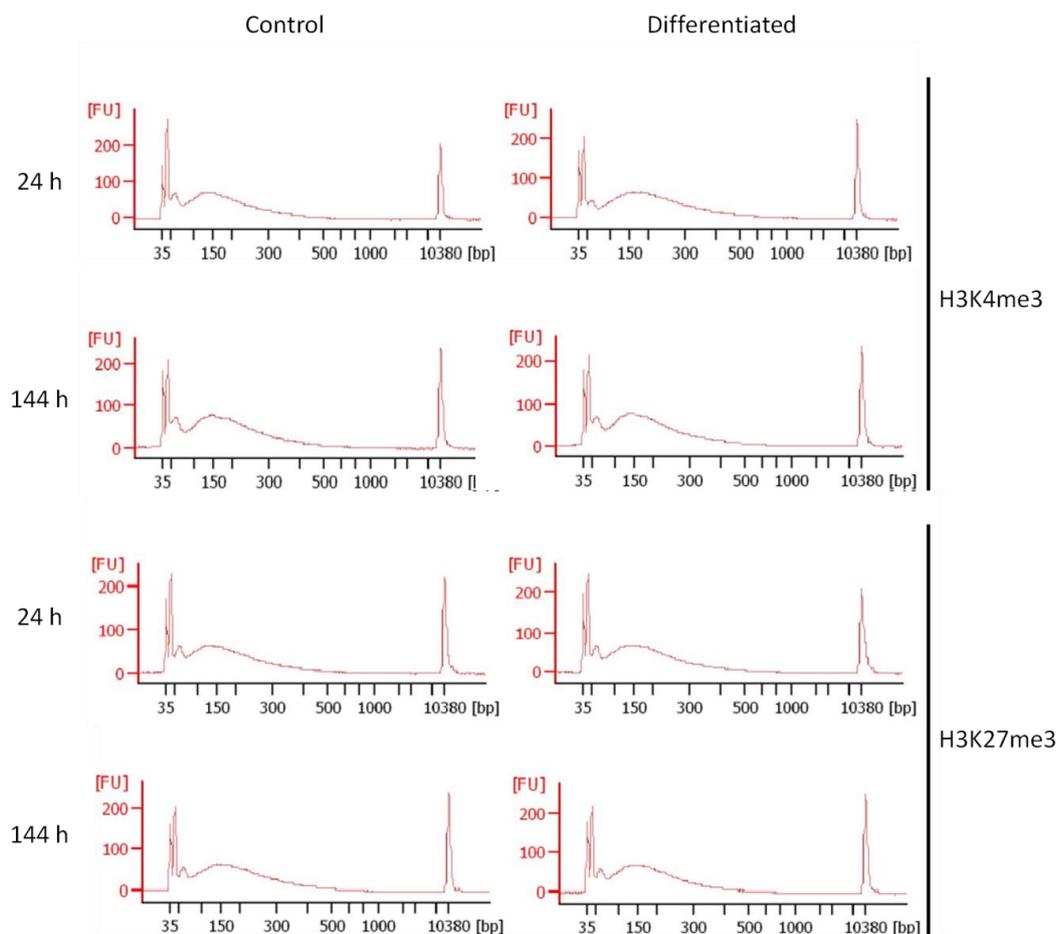
Figure 5.3.8. Bioanalyzer traces of amplified ChIP-seq libraries. All libraries show uniform size distribution around the expected 100-300 bp. This indicates that amplification resulted in no high molecular weight artefacts, and that cleaved primers were successfully cleaved from the amplicons. Primers give rise to the peaks just above the 35 bp size marker. Primers were removed from the samples in a subsequent XP bead cleanup step (data not shown).

5.3.6 ChIP-seq Sequencing

After ChIP-seq samples were sequenced,, the reads were mapped to the genome using maq and peaks where sequences were significantly enriched were called using model-based analysis for ChIP-seq (MACS) (76, 77, 226). The similarity of peak positions between samples was estimated using the Jaccard similarity index within the ChIPseeqer tool (227). H3K4me3 samples showed similarities to one another, as did H3K27me3 samples (Figure 5.3.9). Little similarity was seen between the two groups of samples. This is to be expected, as H3K4me3 and

H3K27me3 are rarely observed in the same positions, and patterns of individual marks across the genome would not be expected to change drastically.
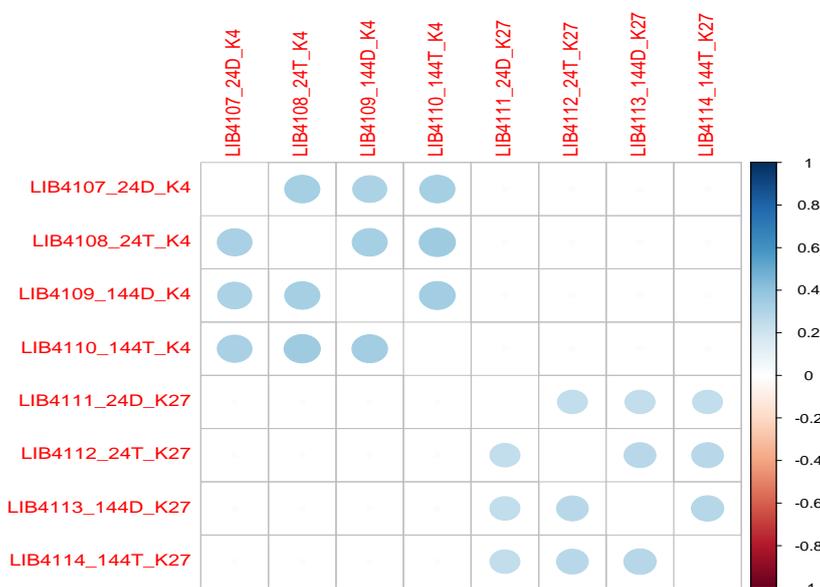
Figure 5.3.9. Jaccard similarity index for ChIP-seq. Cross-similarity of peak positions within ChIP-seq datasets shows internal consistency within ChIP datasets from each histone mark (darker blues show Jaccard scores approaching the highest similarity score of 1). All samples are from cell line Y1192. K4 = H3K4me3 ChIP library, K27 = H3K27me3 ChIP library. 24D /144D = 24/144 h DMSO control, 24T/144T = 24/144 h TZ&PD differentiated. Credit for undertaking Bioinformatics workflow for ChIPseeqer to Janet Higgins, TGAC.

5.3.7 FAIRE Sample Preparation

Material from 24 h and 144 h differentiated and control cells of NHU donor line Y1192 was sonicated as for ChIP-seq. Nucleosome-depleted DNA was isolated from sonication supernatants by mixing 1:1 with a 1:1 mix of phenol: chloroform before RNA digestion and clean-up by sodium acetate precipitation. The size distribution of isolated DNA was expected to be 100-300 bp, and this was demonstrated using a Bioanalyzer (Figure 5.3.10). Due to fragmentation to mononucleosomal size, small fragments were present at 40 bp; this was likely some inter-nucleosome DNA, which has an average length of 40 bp in human chromatin   (147).
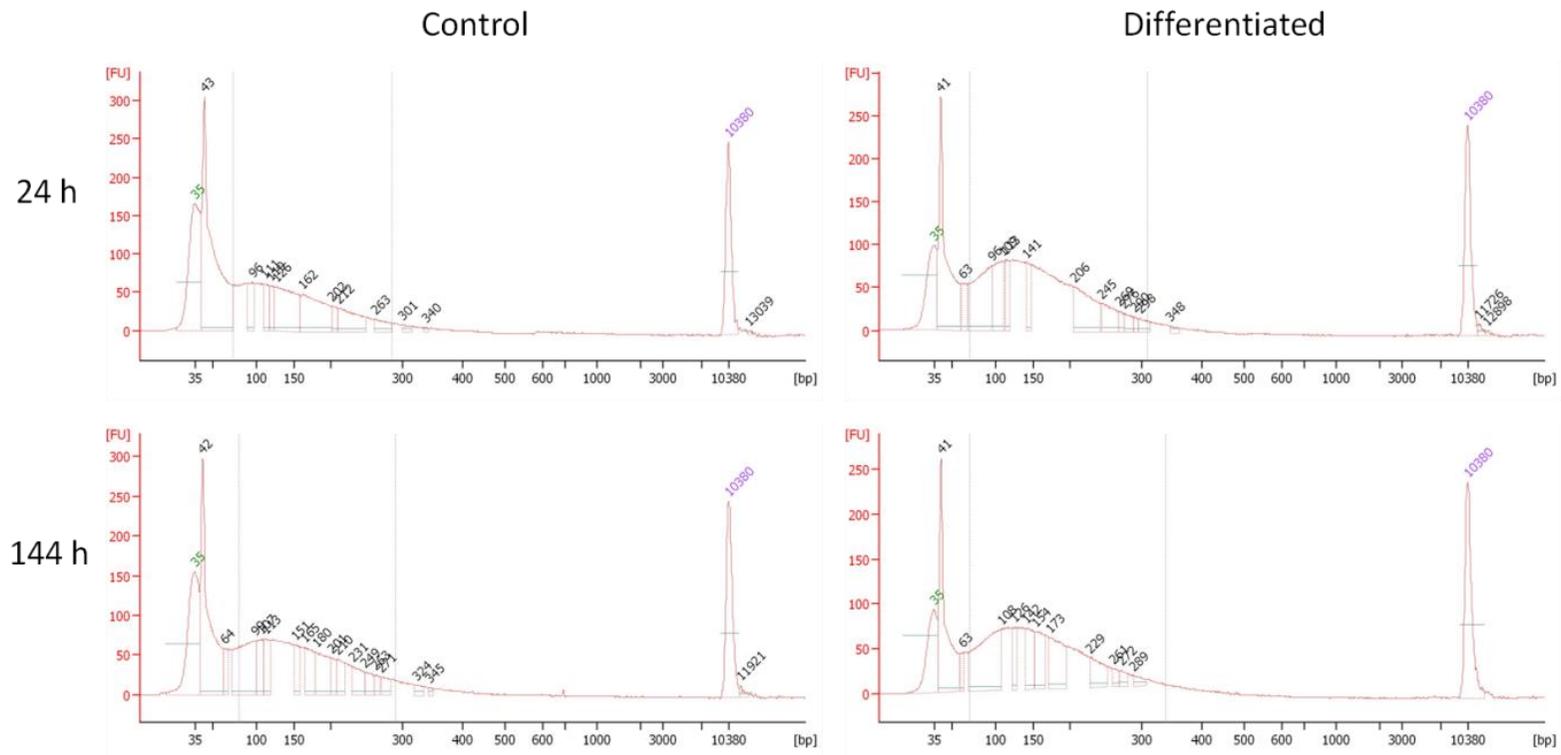
Figure 5.3.10. FAIRE DNA Bioanalyzer results post phenol: chloroform purification. Purified DNA showed expected distribution of fragment sizes from 75-300 bp (x-axis), but with additional peak at 40 bp. FU (y-axis) = fluorescence units.

Linker DNA was removed by size-selection, which is part of the standard FAIRE protocol (228). Size selection is performed by recovery of DNA from an agarose gel with an opening which allows recovery DNA eluting into the space to be captured as the fragments of the desired length pass through (EZ Gel Size Select, Life Technologies, Cat # G661002).

Length of size-selected DNA was measured using a LabChip GX system (PerkinElmer) (Figure 5.3.11). The library from 144 h differentiated cells did not show good recovery. Elution was continued for one minute extra and size of DNA recovered re-checked on a Bioanalyzer (Figure 5.3.12), which showed good recovery of target DNA. The samples were then prepared for sequencing in the same way as the ChIP-seq samples (Materials and Methods 1.7.8.3).
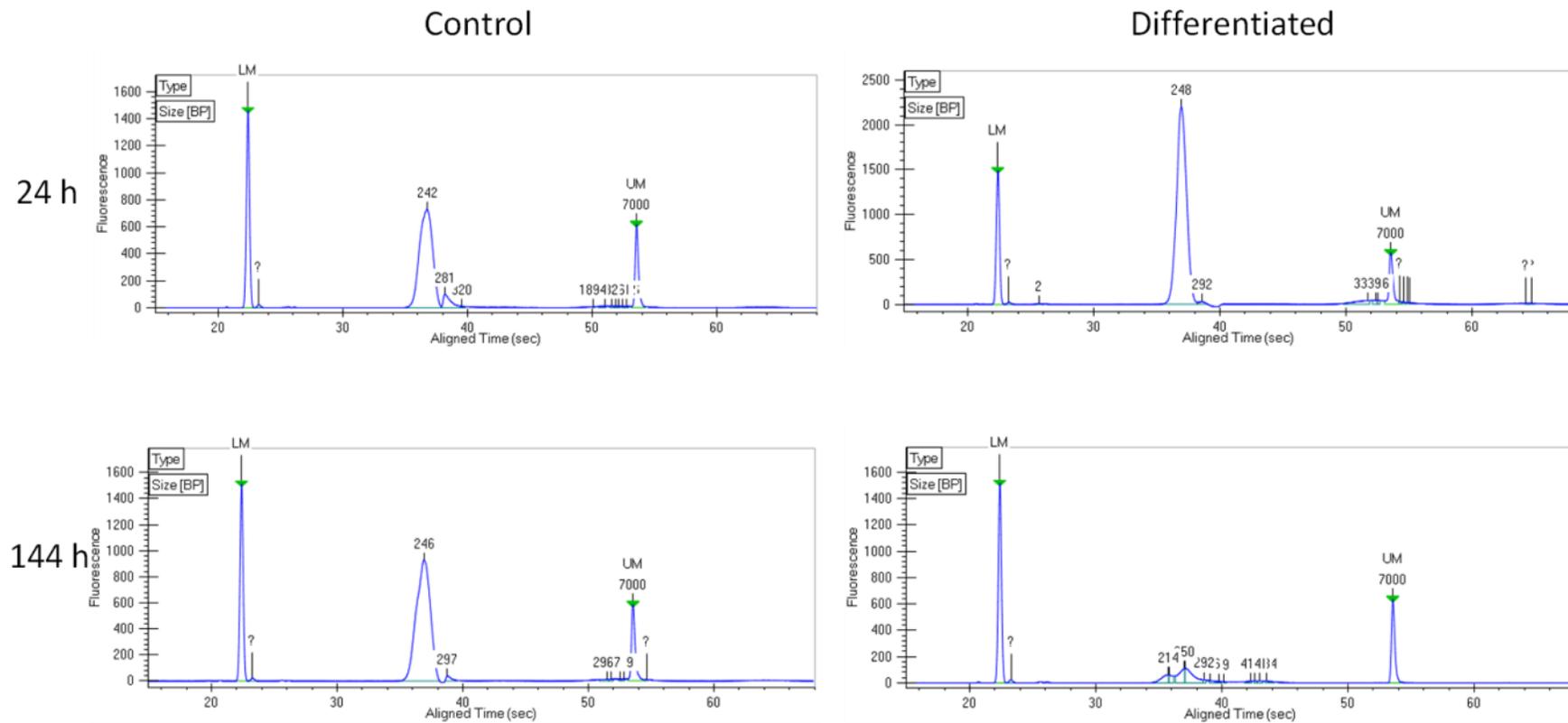
Figure 5.3.11. Size-selection of FAIRE samples. FAIRE DNA was electrophoretically separated using an EZ-Gel Size Select (Life Technologies) system. Once the ladder reached 250 bp, DNA was retrieved from the wells. DNA length of size-selected libraries was checked on a LabChip GX to ensure selection was successful. All libraries other than that from the 144 h differentiated sample showed good recovery at 250 bp; DNA from the 144 h differentiated sample was recovered after further elution (Figure 5.3.12).
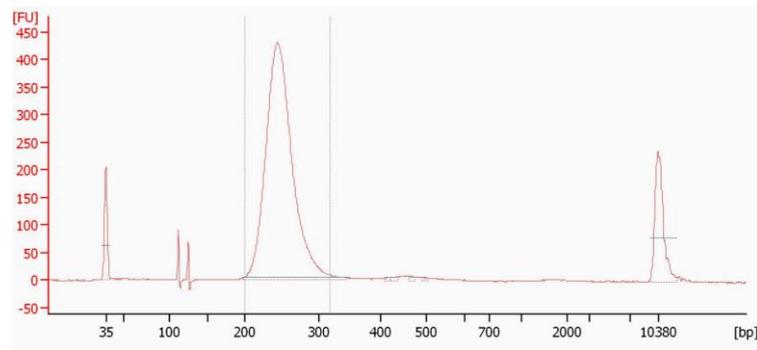
Figure 5.3.12. Bioanalyzer of 144 h differentiated FAIRE library after further elution on size-selection gel. Library was eluted for another 1 min and the DNA in the well was retrieved. The library was electrophoretically separated using a Bioanalyzer system, which showed a peak centred at around 250 bp (x-axis). FU (y-axis) = fluorescence units.

5.3.8 FAIRE Sequencing

FAIRE DNA libraries were submitted for sequencing and returned peaks 22-40 x 10^6 aligned peaks after removal of duplication of reads (Appendix 7.2.1), a data range which resembles previous FAIRE publications (40, 93, 196). Peaks within data were picked as for ChIP-seq for preliminary analysis of distribution of peaks. Distribution of peaks was cross-correlated using the Jaccard similarity index in ChIPseeqer to test samples for internal consistency. Only a minority of peaks in previous published datasets have been seen to change between cell types and / or treatments, so some correlation was expected, and was observed (Figure 5.3.13).
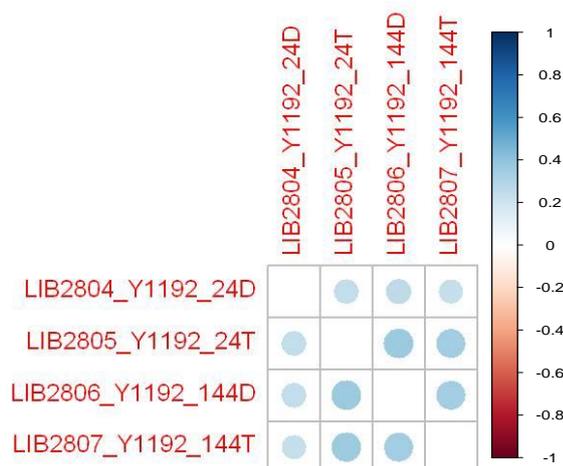
Figure 5.3.13. Jaccard cross-similarity of FAIRE-seq. Peak positions within FAIRE dataset from cell line Y1192 shows internal consistency within datasets (darker blues show Jaccard scores approaching the highest similarity score of 1). LIB numbers are internal library annotations from TGAC. 24D /144D = 24/144 h DMSO control, 24T/144T = 24/144 h TZ&PD differentiated. 24D is the most different from all of the other libraries. Credit for undertaking Bioinformatics workflow for ChIPseeqer to Janet Higgins, TGAC.

As discussed previously, many FAIRE peaks are common across cell types. Genes expressed under phenotype-specific conditions would be expected to have unique FAIRE peaks in and around their coding region (40). The locations of FAIRE peaks in NHU cell data at specified areas of the genome were visualised using the University of California, Santa Cruz (UCSC) genome browser (genome.ucsc.edu). The UCSC browser interrogates peak files for density of reads at user designated regions, and allows comparisons to be made with public datasets. NHU data was compared with a FAIRE dataset from an immortalised urothelial cell line and three other human cell types (lymphoblastoid cell line GM12878, human endothelial stem cells and K562 leukaemia cell line) included in the ENCODE project (228). The coding region of UPK2 was chosen for inspection, as it is expressed most in differentiated cells at the 144 h time-point. Correlation was observed at one site between all NHU FAIRE samples and public datasets (Figure 5.3.14). Flanking this site were two regions of peak density which were unique to NHU cells differentiated for 144 h.
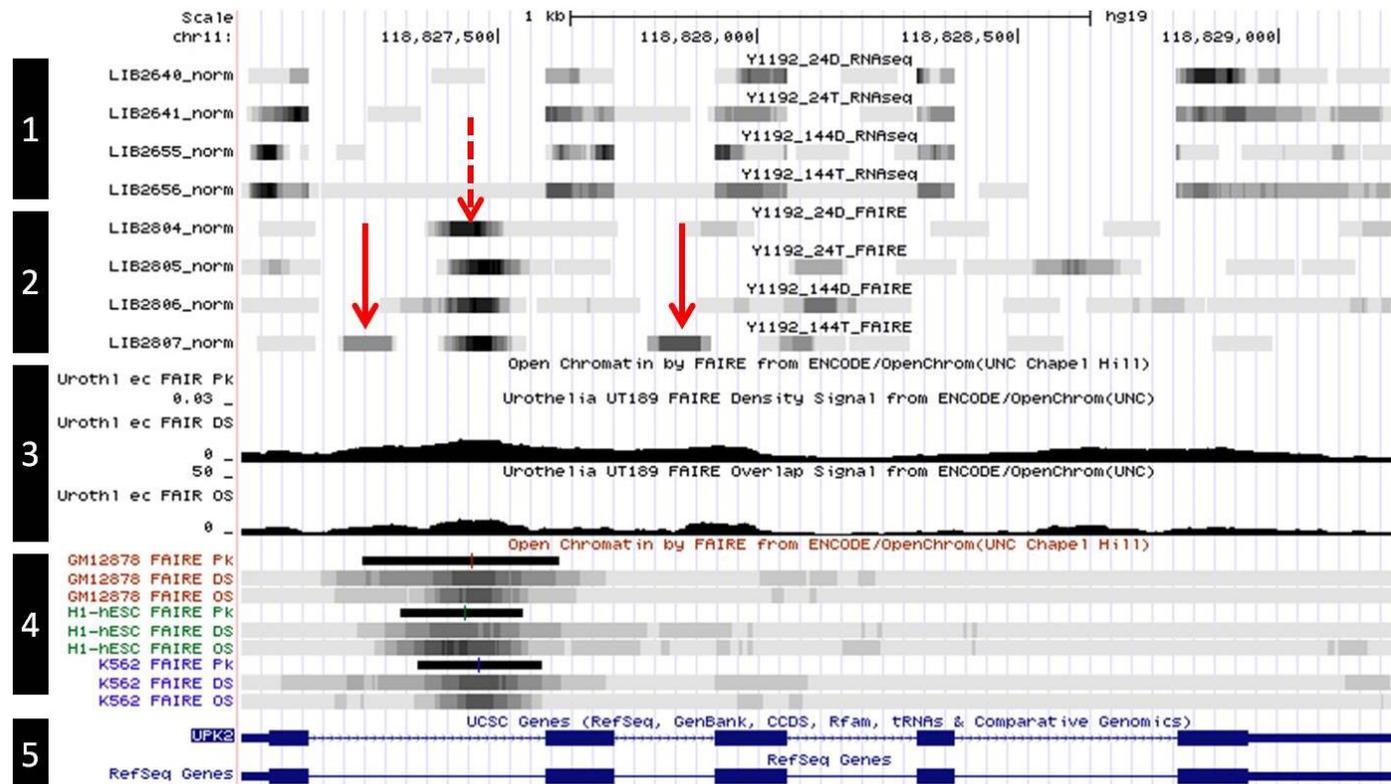
Figure 5.3.14. Unique FAIRE signal in differentiated cells. FAIRE peak density and RNA-seq abundance tracks from NHU cells displayed in the UCSC genome browser. Left to right displays the region encompassing the coding region for the differentiation-associated gene UPK2 centred around position 118,828,000 on chromosome 11. Key: 1= RNA-seq, 2= NHU FAIRE, 3 = Pre-existing FAIRE tracks in the UCSC browser from immortalised urothelial cells, 4 = FAIRE tracks from human GM12878, H1-hESC and K562 cells, 5 = positions of known coding exons as annotated in UCSC and RefSeq. NHU FAIRE tracks show some peaks common to all samples (dashed red arrows), and some intronic peaks unique to differentiation-induced samples at 144 h (solid red arrows).

This area (chromosome 11: 118,827,450 – 118,828,000) was analysed for the presence of all factor motifs using Consite (http://asp.ii.uib.no:8090/cgi-bin/CONSITE/consite). This revealed that the site highlighted in Figure 5.3.14 common to all cell types was enriched in myc proto-oncogene protein (MYC), upstream stimulatory factor 1 (USF) and aryl hydrocarbon receptor nuclear translocator (ARNT) binding sites, whereas the two peaks present in differentiated urothelial cells both contained (Zinc finger protein SNAI1) SNAI1 binding sites (data not shown).

5.3.9 HiC Library Preparation

Fixed cells were suspended in lysis buffer, homogenised and DNA digested with HindIII as in the HiC protocol (Materials and Methods 2.7). An aliquot of the resulting DNA library was taken, cross-links reversed and protein digested by heating at 65°C overnight in the presence of proteinase K. DNA was purified by phenol: chloroform extraction and sodium acetate precipitation. The size of the libraries was checked to determine size (expected to be 10 kb) by electrophoresis (Figure 5.3.15). The size of DNA fragments was approximately 10 kb with minor streaking, similar to that seen in literature reports for HiC samples (229).
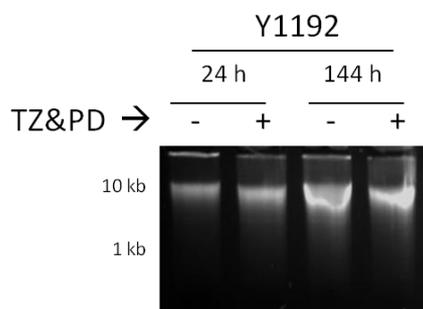
Figure 5.3.15. HiC libraries after HindIII digestion. Fixed DNA from cell line Y1192 subjected to control (TZ&PD -) or differentiation-inducing (TZ&PD +) were homogenised and digested with HindIII (Materials and Methods 2.7.8.1). The size-range of the libraries was determined by loading 500 ng of DNA onto a 0.75% agarose gel. The majority of fragments were 10 kb, with a similar distribution of non-fragmented and smaller fragments to that seen in published HiC methods papers (229).

Sticky ends left by HindIII digestion of libraries were filled in, in the presence of biotinylated cytosine. Libraries were then subjected to ligation using T4 ligase under dilute conditions. Cross-links were reversed by heating at 65°C overnight

and protein digested by incubation with proteinase K. DNA was phenol: chloroform purified and phenol carry over removed using Amicon 30 kDa ultra centrifugal filters (Millipore) before RNase treatment.

The success of the ligation was assessed by generation of PCR using primers which should only generate amplicons if HindIII sites were ligated together (Figure 5.3.16).
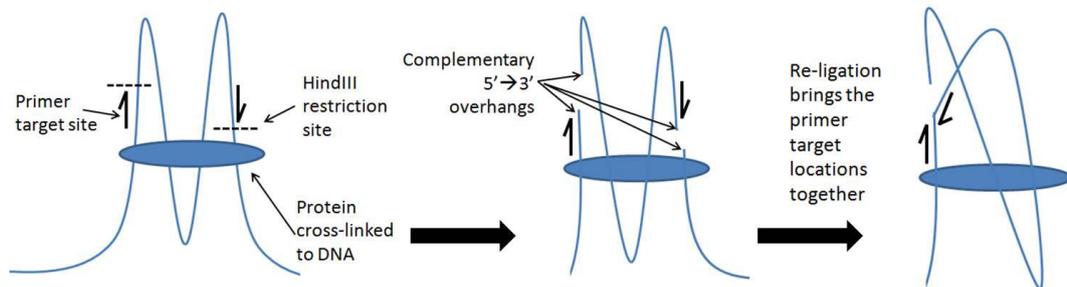
GGTGCAGCTGAGCTAGGCAGCA**GCAAGCATTCCTGGGGTGGCA**TAGTGGGGTGGT GAATACCATGTACA**AAGCTT**GTGCCCAGACTGTGGGTGGCAGTGCCCCACATGGCCGCT TCTCCTGGAAGGGCTTCGTATGACTGGGGGTGTTGGGCAGCCCTGGAGCCTTCAGTTGC AGCCATGCCTTAAGCCAGGCCAGCCTGGCAGGGAAGCTCAAGGGAGATAAAATTCAACC TCTTGGGCCC**TCCTGGGGGTAAGGAGATGCTGCA**TTCGCCCTCTTAATGGGGAGGT GGCCTAGGGCTGCTCACATATTCTGGAGGAGCCTCCCCTCCTCATGCCTTCTTGCCTCTT GTCTCTTAGATTTGGTCGTATTGGGCGCCTGGTCACCAGGGCTGCTTTTAACTCTGGTA AAGTGGATATTGTTGCCATCAATGACCCCTTCATTGACCTCAACTACATGGTGAGTGCTA CATGGTGAGCCCCAAAGCTGGTGTGGGAGGAGCCACCTGGCTGATGGGCAGCCCCTTC ATACCCTCACGTATTCCCCCAGGTTTACATGTTCCAATATGATTCCACCCATGGCAAATT CCATGGCACCGTCAAGGCTGAGAACGGG**AAGCTT**GTCATCAATGGAAATCCCATCACCA TCTTCCAGGAGTGA

Figure 5.3.16. Design of PCR primers for HiC ligation quality control. Primers locations (green text) were designed targeting the forward strand at sites upstream of two potential HindIII digestion (red text) sites located close to one another near the GAPDH gene. As both primers target the forward strand, no PCR product should result from these primers using human genomic DNA. Human genomic DNA correctly manipulated by HiC potentially results in ligation of the two HindIII target sites, allowing PCR reaction to proceed.

The success of fill-in and biotinylation of ligation junctions was assessed by digestion of the PCR products, which should respond to NheI and not HindIII

when the HindIII sticky ends were filled in and ligated (Figure 5.3.17). 50-60% of PCR products were digested with NheI, as achieved in the HiC method paper (70), indicating that the libraries should be of sufficient quality for sequencing.

	24 h				144 h			
	Control		Differentiated		Control		Differentiated	
	HindIII	NheI	HindIII	NheI	HindIII	NheI	HindIII	NheI
Uncut (%)	55.2	45.3	57.3	47.1	65.7	37.1	60.1	40.8
Cut (%)	44.8	54.7	42.7	52.9	34.3	62.9	39.9	59.2
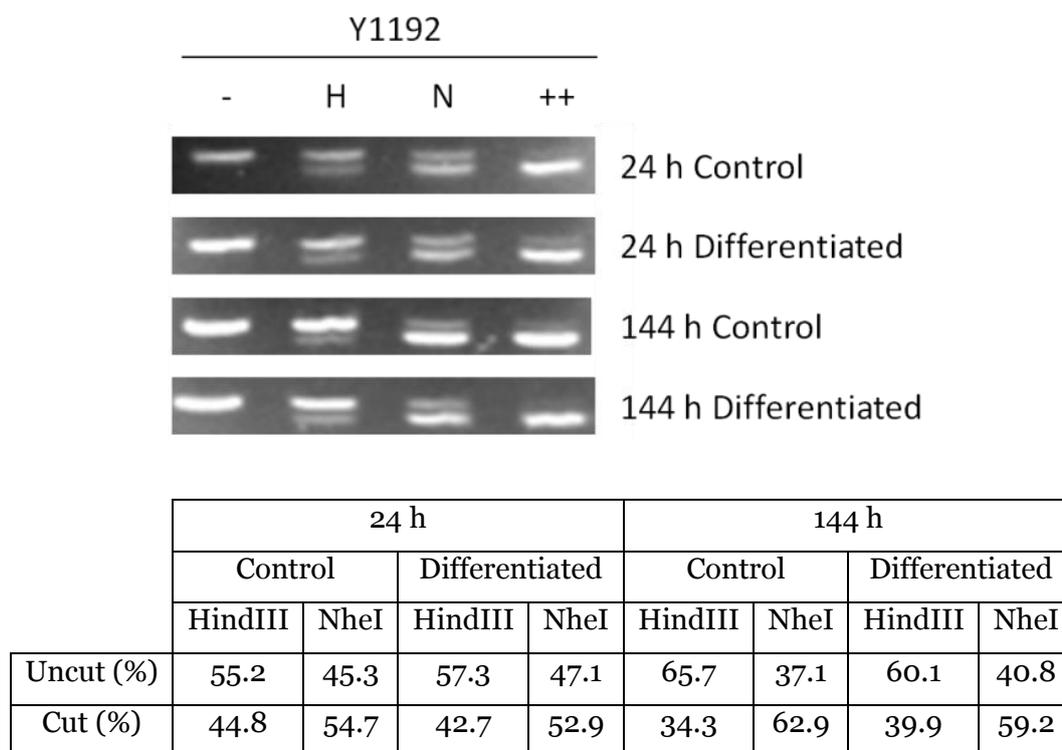
Figure 5.3.17 Digestion of HiC library PCR products. PCR products spanning two potential HindIII cut sites (Figure 5.3.16) were amplified from HiC libraries as described in Materials and Methods 2.7.6.4. PCR products were digested using HindIII (H), NheI (N) or both enzymes (++). Directly comparing the mean intensity of the uncut to cut portion of bands, NheI digestion resulted in 50-60% cutting in all products, as in the HiC method paper (71). All PCR products in this assay were shortened by 48 bp after nuclease treatment, rather than cleaved in two as in the original method. These calculations of cut percentages are therefore likely to be slight underestimates as 11% of the nucleotides from the cut band are not contributing to its fluorescence.

Libraries were exonuclease treated to remove biotin from unligated HindIII cuts sites using T4 DNA polymerase, preventing them from being captured in downstream affinity purification. Libraries were subsequently sonicated to 100 - 300 bp on a Covaris S2 (Covaris) instrument, and size checked by gel electrophoresis (Figure 5.3.18).
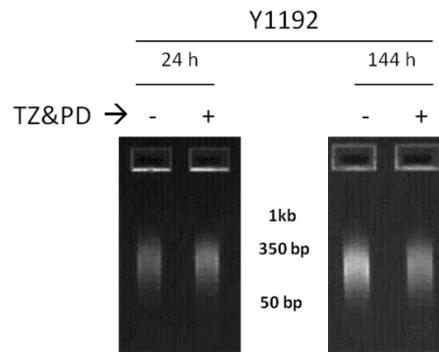
Figure 5.3.18. Sonication of HiC Libraries. Y1192 HiC libraries were fragmented to 100-300 bp using a Covaris S2 instrument. DNA was separated by gel electrophoresis. Most visible fluorescence from DNA was observed with mobilities within the range of those of the 50 bp and 350 bp markers, indicating successful fragmentation.

After determining the size of libraries, DNA <100 bp and >300 bp was removed by XP bead size-selection (Materials and Methods 2.7.6.7). The size of the purified libraries was determined to be centred at 200 bp on the Bioanalyzer (Figure 5.3.19).
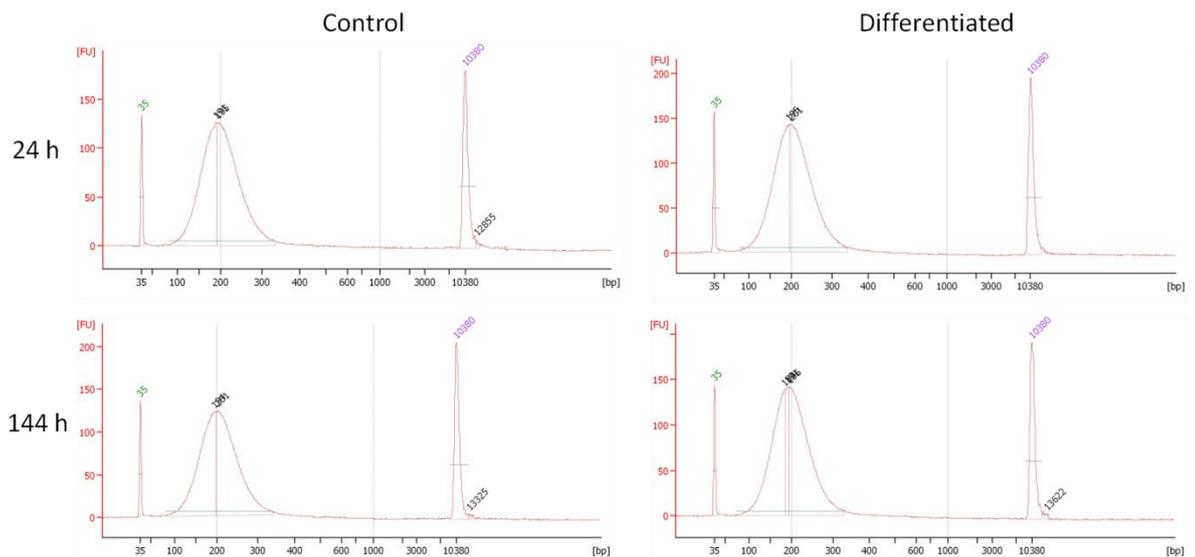
Figure 5.3.19. Bioanalyzer analysis of fragmented HiC library. Size distribution of libraries after sonication and XP bead purification. All libraries had size profiles centred at around 200 bp as expected. FU (y-axis) = fluorescence units.

Libraries were end-repaired and A-tailed to fix DNA ends broken by sonication and allow adapters for sequencing to be ligated. Next, streptavidin purification was performed to isolate DNA containing target ligation junctions that are labelled with biotin-C. Purified DNA was ligated to sequencing primers and an aliquot amplified for 12 or 18 cycles using a thermal cycler (Materials and Methods 2.7.6.10). Over-amplification of libraries results in PCR artefacts which cause biases in analysis of DNA libraries. To avoid generating these and maintain enough amplification to obtain sufficient material to sequence, amplification was tested for 12 and 18 cycles on 144 h control and differentiated libraries. When separated on a Bioanalyzer, 12 cycles of amplification showed the same library peak shape as before streptavidin purification, whereas 18 cycles showed the generation of high molecular weight material (Figure 5.3.20).
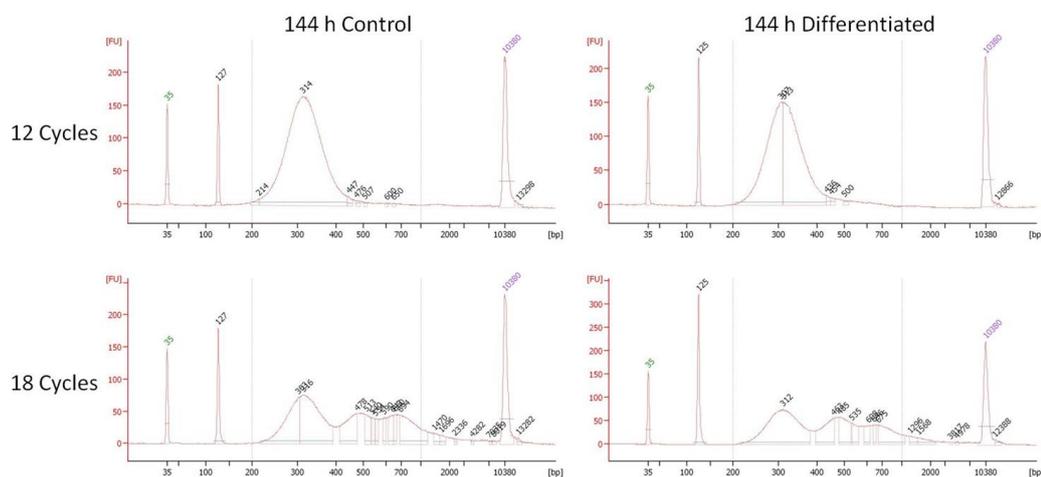
Figure 5.3.20. Bioanalyzer plot of streptavidin purified HiC libraries amplified for 12 and 18 cycles. Ideal peaks around (314 bp) were observed for samples after 12 cycles. Peak observed at 125 bp was dimers of sequencing adapters. FU (y-axis) = fluorescence units.

As a quality control to test if the NheI sequence from ligated sequences was present in amplified libraries, samples subjected to 12 rounds of amplification were incubated with NheI enzyme. Both libraries showed disruption of peak shape, with the production of smaller length products indicative of cleavage of the libraries (Figure 5.3.21).
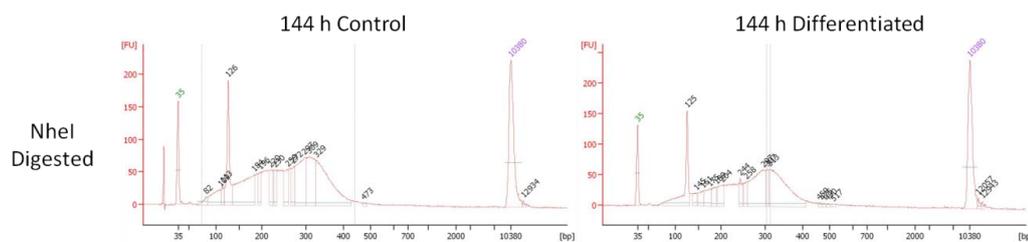
Figure 5.3.21. Digestion of amplified HiC libraries. Biotinylated sequences at ligation junctions in HiC libraries should contain NheI recognition sites. Digestion of aliquots of amplified libraries was used by Belton *et al* (70) as a measure of library quality. After incubation with NheI, both libraries showed reduction in peak height relative to the sequencing adapter dimer, and significant production of smaller length products as compared to those seen in 1.2.17). FU (y-axis) = fluorescence units.

After demonstrating digestion of two of the amplifications, fresh aliquots of all four HiC libraries were amplified for 12 cycles, and sequencing adapters removed by cleanup with XP beads before checking size-distribution on the Bioanalyzer (Figure 5.3.22). All libraries had expected peak shape, with no high molecular weight artefacts. Libraries were then submitted to the sequencing workflow.
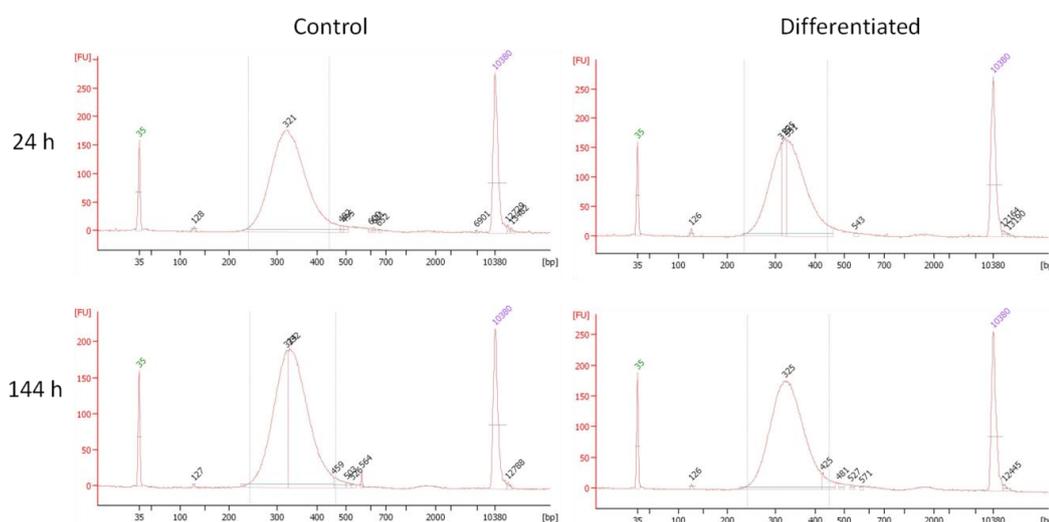
Figure 5.3.22. Amplified HiC libraries for cell line Y1192. Libraries all showed expected peak shapes and low abundance of high molecular weight DNA and sequencing adapters, thus passing quality control criteria to continue with sequencing. Libraries amplified for 12 cycles using 1.5 µL of 20 µL template DNA, and then the low molecular weight DNA was removed using an XP bead clean-up. FU (y-axis) = fluorescence units.

5.3.10   HiC Sequencing

HiC libraries were submitted for sequencing and resulting sequences mapped to the genome, and displayed as dot-plots (Figure 5.3.23). Plots bore sufficient qualitative resemblance to other HiC maps to be confident that further processing would be fruitful.
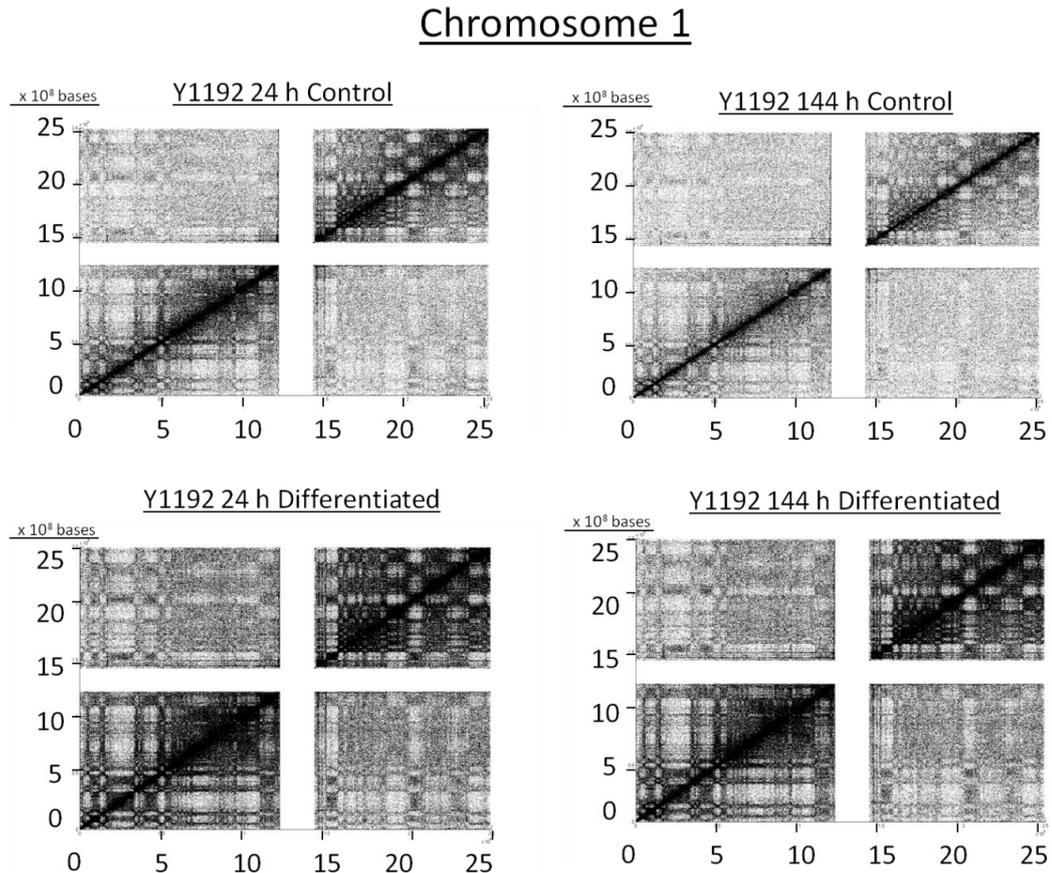
Figure 5.3.23. Non-normalised chromosome 1 interaction dot plots for HiC dataset from cell line Y1192. Individual dots in the plot represent a sequence read which arose from two separate sites. For example, a dot at position 2 on the x-axis and 10 on the y-axis is from a sequence which was a ligation event between two interacting DNA strands at 2 x 10⁸ and 10 x 10⁸. X-axis and Y-axis are positions on the chromosome plotted in 50 MB (5 x 10⁸ bases) increments. The plaid pattern is reminiscent of patterns seen in other HiC publications, and looks similar in each of the samples apart from the difference in intensity at some sites which is brought about in this representation because of a lack of normalisation of the differences in overall read numbers.

5.3.11 RNA-seq Results

RNA isolated from three NHU cells lines under differentiation or control conditions at 24 h or 144 h was subjected to massively parallel sequencing and sequences mapped to the Ensembl GRCh37 and UCSC hg19 annotations of the human genome using tophat2 (Materials and Methods 2.7.9.1). It was the intention of this experiment to use RNA from different donors to assess which transcripts were consistently altered during urothelial differentiation.

In order to normalise for overall transcript abundance and allow for comparisons of numbers of sequences arising from particular genes, the "per-condition" DE-seq method (223) was used to generate Benjamini-Hochberg adjusted p-values (p-adj) for differentially expressed ensembl (www.ensembl.org) transcripts between libraries from all three NHU donors at 24 and 144 h of control or differentiation treatment.

Using this approach, 1038 and 888 ensemble transcripts (ENST) were identified as being consistently differentially expressed ≥2-fold with adjusted p-values of p≤0.05 across donors between differentiation-induced and control cells at 24 h and 144 h respectively (Supplementary Materials 3.1.1.1).

577 transcripts were upregulated and 461 downregulated in differentiation-induced cells relative to controls at 24 h. 572 transcripts were upregulated and 315 downregulated in differentiation-induced cells relative to control at 144 h.

Transcripts upregulated ≥2-fold at 24 h with p-adj ≤0.05 included the previously identified urothelial differentiation markers, KRT13 (49), claudin 4 (2)and IRF1 (34) (Figure 5.3.21). Transcripts from UPK2, the terminal differentiation marker of human urothelial cells (25, 230), was detected as upregulated ≥2-fold with p-adj ≤0.05 at 144 h, as again were KRT13 and claudin 4. MAPK1 expression, which is not known to be affected by induction of differentiation, showed no significant alteration in mean expression levels throughout the experiment.

KRT14, which is downregulated upon differentiation, was reduced in abundance ≥2 fold at 144 h, but did not attain a p-adj value ≤0.05. Similar to this, expression of some PPARγ transcripts was increased as expected at 144 h, although not to a statistically significant extent (p-adj ≥0.05). As with previous observations, small amounts of PPARG2 transcript (ENST00000287820) were detected in cells after 144 h, although this time only in differentiated cells.

The observation that genes with high expression had consistent fold-changes, but have p-adj values ≥0.05 was suspected to be due to the variation between donor samples. The raw normalised counts for transcripts in each library were inspected (Supplementary Materials 3.1.1.2) and the read counts for specific transcripts from genes with known activity in urothelial differentiation were interrogated (Figure 5.3.24). The cell line Y1214 had much higher expression of the squamous cell marker KRT14 and although all the differentiation markers AQP3, FABP4, KRT13, CLDN3, CLDN4 and IRF1 upregulated in the expected manner, the number of transcripts from each was much lower than from the other two cell lines. This variance affected the p-values due to the increase in the standard deviation. Despite such issues with variance, one advantage to pooling data from donors is that genes which change expression levels robustly across donors will be highlighted, as exemplified by KRT13, UPK2 and CLDN4.
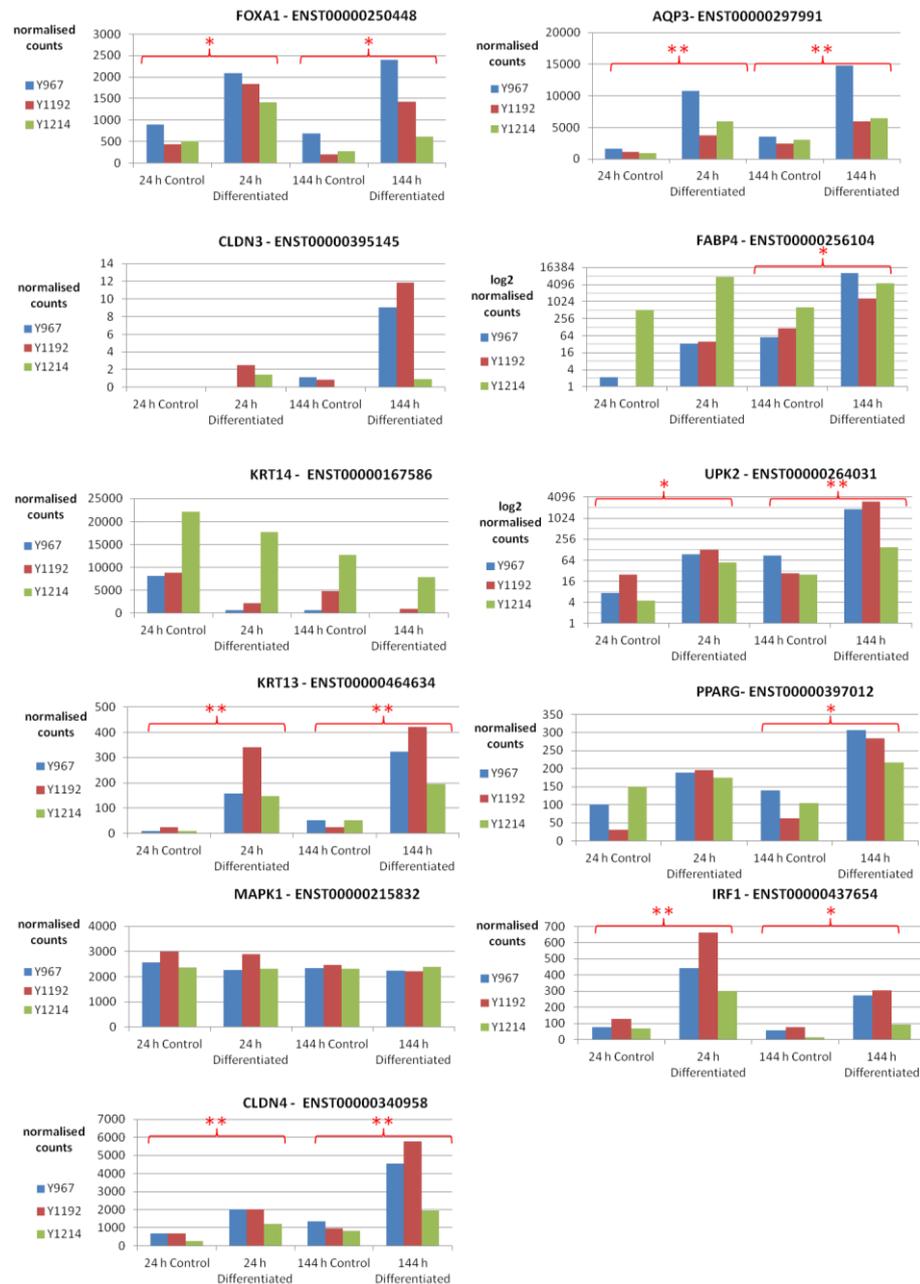
Figure 5.3.24. RNA-seq counts transcripts from differentiation-associated genes. Genes marked with * or ** have mean 2-fold change between control and differentiated extracts at each time-point with p≤0.05 or Bonferroni-adjusted p-value p≤0.05 respectively, as calculated using the DEseq method (223). KRT14 was expected to be downregulated with differentiation, and MAPK1 was not expected to change. All other genes were expected to upregulate with differentiation. All genes changed in the expected direction in all cell lines, but not all were statistically significant (p ≤ 0.05) when averaged across the three cell lines. The reason for this appeared to be the Y1214 cell line, which had consistently weaker induction of differentiation marker genes compared to the other two cell lines. All counts were normalised to the total number of reads for each line.

Analysis of genes by assessing if groups participate in known pathways, have similar functions, or have similar transcription factor binding sites can help to extract information about the control and function of groups of genes in large lists. To this end, transcription factor binding site enrichment analysis was performed on the promoters of genes up and down-regulated at each time-point. Promoter analysis of the -950 to +50 region of upregulated genes using the PSCAN tool (http://159.149.160.51/pscan/) did not reveal any of the upregulated DNA binding proteins to have enriched binding sites around the proximal promoter region of upregulated genes. This was consistent with the theory that the genes upregulated at early time-points are involved in upregulation of genes at later time points. PPARγ:RXR was 13th in the list of transcription factors with binding sites in genes upregulated at 24 h, and moved up to 2nd at 144 h (Table 5.3.2). PPARγ:RXRα was the only transcription factor motif unique to upregulated genes across both time points, consistent with the likelihood that it has a role to play in differentiation. The transcription factors SP1, KLF4, INSM1 binding sites were common across all lists.

PSCAN Promoter Analysis of Genes Changing ≥2-fold							
TF Associated with Genes Down at 24 h		TF Associated with Genes Up at 24h		TF Associated with Genes Down at 144h		TF Associated with Genes Up at 144h	
Gene Name	P-value	Gene Name	P-value	Gene Name	P-value	Gene Name	P-value
SP1	3.67E-16	Klf4	4.25E-08	TFAP2A	5.16E-07	Klf4	0.00051
TFAP2A	2.58E-11	SP1	2.48E-07	NFKB1	2.07E-05	PPARG::RXR	0.00104
Egr1	6.62E-10	CTCF	1.47E-06	Tcfcp2l1	0.00024	TFAP2A	0.00150
Klf4	5.39E-08	TFAP2A	1.86E-05	Klf4	0.00029	SP1	0.002211
NFKB1	2.85E-07	MZF1_5-13	6.62E-05	GABPA	0.00035	Zfp423	0.00277
GABPA	1.11E-06	INSM1	0.000117	PLAG1	0.00038	FEV	0.00438
HIF1A::ARN	1.88E-06	Pax5	0.000161	HIF1A::ARN	0.001171	PLAG1	0.00581
Pax5	6.63E-05	PLAG1	0.000173	Egr1	0.001223	INSM1	0.00687
Zfx	6.92E-05	Tcfcp2l1	0.00029	Zfx	0.001953	Tcfcp2l1	0.00735
E2F1	0.00028	Zfx	0.000331	E2F1	0.004215	ESR1	0.00789
Arnt::Ahr	0.00037	Egr1	0.00050	INSM1	0.00479	RREB1	0.00831
MZF1_5-13	0.00064	znf143	0.00060	SP1	0.006011		
MIZF	0.00069	PPARG::RXR	0.00104	Pax5	0.00627		
INSM1	0.00079	NHLH1	0.001616	NF-kappaB	0.007103		
EBF1	0.001223	NF-kappaB	0.001935	Arnt::Ahr	0.00805		
ELK4	0.001522	MZF1_1-4	0.00436	MZF1_1-4	0.00964		
NF-kappaB	0.001741	Zfp423	0.00629				
ELK1	0.001786	HIF1A::ARNT	0.009641				
Zfp423	0.00205	Arnt::Ahr	0.009821				
CTCF	0.00434						
Mycn	0.00470						
NHLH1	0.00499						
Arnt	0.007198						
RELA	0.007417						
Myc	0.00805						
RREB1	0.008152						

Table 5.3.1. PSCAN promoter analysis of -950 to +50 region of differentially regulated genes using JASPAR database motifs. Transcripts upregulated or downregulated ≥2-fold with an adjusted p-value of ≤0.05 were submitted, and all factors with p≤0.01 are shown. Red highlighted rows show factors common between up and downregulated genes at 24 h, and green shows common factors between up and downregulated genes at 144 h. PPARG:RXRA was uniquely present in upregulated at both 24 h and 144 h. NFKB1 and GABPA were uniquely present in promoters of downregulated genes at both time points.

The Gene Ontology (www.geneontology.org) aims to assign attributes to genes and proteins in model organisms, including humans allowing rapid comparison of large gene lists for common functions. Genes upregulated or downregulated across all three donor lines by ≥2-fold at 24 h and 144 h were split into separate lists and submitted to the GSEA tool (http://www.broadinstitute.org/gsea). The tool was used to match the gene lists for enrichment of gene function using the Gene Ontology "molecular function" tab.

The top three functions enriched in genes upregulated after 24 h differentiation were: 23 genes associated with "oxidoreductase activity" (GO:0016491), 25 genes associated with "DNA binding" (GO:0003677) and 5 genes associated with "aldo-keto reductase activity" (GO:0004033) (full lists for all gene ontology in Supplementary Materials 3.1.2.5). Genes associated with oxidoreductase activity included cytochrome P450 proteins, which are associated with the oxidative metabolism of numerous compounds, including xenobiotics and endogenous lipid. The upregulated genes associated with aldo-keto reductase activity are involved in lipid processing. Two of these genes process steroid and hormone signalling molecules, and are specifically involved in inactivation of progesterone (AKR1C1) or conversion of androstenedione to testosterone (AKR1C3). This upregulation of lipid metabolic enzymes indicates that PPARγ shares some common target genes with those upregulated during adipogenesis. Indeed AKR1C3 is known to be part of the prostaglandin processing pathway in adipocytes (231). At 24 h the 25 upregulated genes coding DNA binding proteins included the proteins with transcription factor activity, including: histone deacetylase 1 (HDAC1), transcriptional-regulating factor 1 (TRERF1), hypoxia-inducible factor 1-alpha (HIF1A), transcription elongation factor SPT4 (SUPT4H1), TAR DNA-binding protein 43 (TARDBP), signal transducer and activator of transcription 1-alpha/beta (STAT1), FOS-like antigen 2 (FOSL2), trans-acting T-cell-specific transcription factor GATA-3 (GATA3), nuclear factor erythroid 2-related factor 2 (NFE2L2), putative homeodomain transcription factor 1 (PHTF1), tripartite motif-containing protein 29 (TRIM29) and TSC22 domain family protein 3 (TSC22D3). None of which has any previously described role in differentiation of human urothelial cells, although FOSL2 is necessary for adipogenesis (232).

In genes upregulated at 144 h, the three most enriched molecular functions were: "actin binding" (GO:0003779), "hydrolase activity acting on ester bonds (GO:

0016788)", and "cytoskeletal binding" (GO:0008092). This enrichment for cytoskeletal proteins suggests that at 144h, the genes associated with the structural changes dominate the differences between the differentiated and control cells.

When subjecting the same lists to "canonical pathway" analysis in GSEA, which searches for members of known pathways in gene lists, 37 and 32 genes at 24 h and 144 h respectively were involved in immune system function. A sub-set of immune-related genes were associated with interferon signalling and / or antigen processing and presentation. This could represent the establishment of the urothelial programme for innate immune signalling, a pathway which could be influenced by the transcription factor IRF1 which has previously been identified as being associated with urothelial differentiation (34).

Genes upregulated at both early and late time-points in differentiation may be key to maintenance of cell identity. Analysis of transcripts which were upregulated at both time-points across all three donors using the DE*seq* results revealed 50 transcripts common to both comparisons (Table 5.3.2). Some of these were from known differentiation-associated genes such as KRT13, AQP3, IRF1 and CLDN4. The remaining genes contained several transcription factors such as ETS-related transcription factor (ELF3), retinoic acid receptor responder protein 1 (RARRES1), T-box transcription factor TBX3 (TBX3), and protein L-Myc-1 (MYCL1). GO analysis of the transcripts using GSEA revealed that centrobin (CNTROB), 15-hydroxyprostaglandin dehydrogenase [NAD(+)] (HPGD), tumour protein D52-like 1 (TPD52L1), promyelocytic leukemia (PML), and TBX3 are all associated with cell cycle control, likely for maintenance of quiescence after differentiation.

Gene	Transcript	Ctrl Count	Diff Count	Fold	Ctrl Count	Diff Count	Fold
KRT13	ENST00000246635	713.75	7368.	10.32	1127.1	10629	9.43
AQP3	ENST00000297991	1183.6	6805.	5.75	3003.	9073.	3.02
ELF3	ENST00000367283	236.2	3830.	16.21	944.9	5252.	5.56
TBX3	ENST00000349155	1238.	3407.1	2.75	1701.1	4334.	2.55
IGFBP3	ENST00000275521	110.95	1747.6	15.75	808.2	4880.	6.04
CLDN4	ENST00000340958	545.8	1754.7	3.21	1040.	4099.	3.94
MYCL	ENST00000397332	281.3	1855.5	6.60	679.4	2435.	3.59
SPTSSB	ENST00000359175	40.47	296.4	7.32	163.39	2358.	14.43
ACER2	ENST00000340967	178.11	557.0	3.13	226.8	1972.3	8.70
PPP1R12B	ENST00000391959	146.61	971.15	6.62	291.15	1439.	4.94
PSMB10	ENST00000358514	275.9	1039.	3.77	317.33	1250.	3.94
TBX3	ENST00000257566	434.9	1164.6	2.68	327.8	854.3	2.61
HPGD	ENST00000422112	16.18	114.11	7.05	55.49	1853.	33.41
TMPRSS2	ENST00000332149	43.29	294.3	6.80	74.17	1661.4	22.40
RARRES1	ENST00000237696	24.44	194.6	7.96	88.36	1691.9	19.15
HMGCS2	ENST00000544913	11.62	119.80	10.31	94.92	1481.0	15.60
FAM174B	ENST00000327355	60.52	402.2	6.65	125.6	1110.3	8.83
ACSL5	ENST00000354273	124.9	362.0	2.90	129.2	1146.9	8.87
ACSL5	ENST00000356116	124.9	362.0	2.90	129.2	1146.9	8.87
TMEM184	ENST00000449955	84.98	348.4	4.10	320.0	1144.2	3.57
HID1	ENST00000425042	60.31	412.0	6.83	94.69	870.8	9.20
PTGR1	ENST00000238248	190.9	584.0	3.06	138.35	683.2	4.94
KRT13	ENST00000468313	39.49	491.6	12.45	106.73	749.18	7.02
BCAS1	ENST00000395961	8.28	115.96	14.01	38.83	1034.	26.65
PLEKHA7	ENST00000448080	78.33	380.2	4.85	198.9	735.79	3.70
GPR160	ENST00000355897	23.90	93.31	3.90	43.19	738.7	17.11
CCDC64B	ENST00000572240	42.62	180.6	4.24	169.8	615.39	3.62
PRSS27	ENST00000302641	27.11	115.92	4.28	153.17	659.4	4.31
PSCA	ENST00000513264	14.98	86.04	5.74	62.34	674.2	10.82
KRT13	ENST00000464634	14.82	215.44	14.54	42.91	313.47	7.31
MUC20	ENST00000320736	13.82	161.04	11.65	63.18	326.7	5.17
RNF213	ENST00000560083	25.52	302.1	11.84	0.00	147.76	Inf
PSCA	ENST00000505305	7.07	73.15	10.34	30.10	371.44	12.34
IRF1	ENST00000472045	29.56	228.3	7.72	36.06	160.3	4.45
CYP4F12	ENST00000550627	5.56	55.66	10.01	60.25	331.3	5.50
APOL3	ENST00000397293	0.00	119.72	Inf	0.00	206.2	Inf
RASSF2	ENST00000379400	0.00	28.01	Inf	0.00	255.0	Inf
COQ2	ENST00000311469	0.38	70.32	184.5	0.00	153.4	Inf
PML	ENST00000268059	0.37	70.32	189.3	0.55	123.41	225.0
NBEAL2	ENST00000292309	0.00	86.45	Inf	0.00	62.65	Inf
TMCO4	ENST00000375122	0.00	43.98	Inf	0.19	103.6	532.7
ENPP4	ENST00000321037	0.00	42.87	Inf	0.00	84.66	Inf
HPGD	ENST00000506910	0.00	27.61	Inf	2.73	90.51	33.12
NFKBIB	ENST00000509705	0.00	88.87	Inf	0.00	27.35	Inf
PBXIP1	ENST00000368465	0.00	25.64	Inf	0.00	76.81	Inf
MUC1	ENST00000368395	0.00	40.95	Inf	0.29	58.38	199.19
ADAM15	ENST00000526491	0.00	75.82	Inf	0.00	14.66	Inf
CNTROB	ENST00000380255	0.00	49.28	Inf	0.00	29.45	Inf
TPD52L1	ENST00000392482	0.00	50.22	Inf	0.00	24.69	Inf
FAM193B	ENST00000506955	0.00	33.50	Inf	0.00	40.34	Inf

Table 5.3.2. Transcripts upregulated in differentiated samples at both 24 and 144 h as calculated using the DE*seq* method. Transcripts ranked by average combined expression in differentiation-induced samples. Some genes are represented by multiple transcripts.

5.4 Discussion

5.4.1 RNA-seq

NHU donor cell lines showed upregulation of the differentiation markers FOXA1 and claudin 5 at the protein level, and RNA-seq results confirmed the upregulation of known differentiation-associated genes. One cell line (Y1214) responded with less robust upregulation of differentiation marker genes relative to the other two cell lines, and maintained some expression of undifferentiated makers such as KRT14. Although the transcripts from these urothelial differentiation marker genes all changed expression as expected, this variation within the donors affected the significance of p-values for some genes. Of the transcripts from differentiation-associated genes discussed in the chapter above, those which did not fall below p(adj)≤0.05, such as CLDN3, often had low transcript count numbers, which is known to affect the power of statistical analysis of sequencing data due to high variability of reproducibility in sampling of low abundance transcripts, which occurs even within technical replicates(223).

This biological variation within groups with lower numbers of replicates is known to cause statistical tests such as the Poisson distribution to make false rejections of differential expression. DE*seq* extends the efforts of previous methods such as *edgeR*, which were specifically designed to circumvent the problems of calculating differential expression between biological replicates which have high natural variation and often have smaller numbers of replicates, thus preventing true mean and variance from being calculated. However, the authors of DE*seq* did note that transcripts with low counts will require very high fold changes to be considered significant. Thus, although one donor line caused some issues DE*seq was* successful at calling differential expression (p<0.05) in the genes analysed in all but the most extreme cases where expression was low (<100 counts) or fold-change for one sample was significantly lower. The adjustment of the p-values by the Benjamini-Hochberg (233) method to reduce the false discovery rate (FDR) did however remove some of the known differentiation genes which attained a p≤0.05 and had an appreciable fold-change , including FOXA1 and PPARγ .

One drawback from assigning RNA-seq data to individual transcripts is that there may be multiple transcripts for genes, and all of them may not change their expression levels. If only one of the transcripts changes significantly, then it may be the case that the gene is being transcribed from a different promoter or that a

new transcript or splice variant is being expressed. It may also be that the transcript which alters is much less abundant than the other transcripts, thus highlighting a change which may have little overall effect on the relative levels of the gene. In this case the analysis may benefit from the data being re-mapped so that all reads are assigned to a single gene rather than individual transcripts. Having individual values for each gene would provide a clearer picture of genes which change their overall expression, and alternative differential expression approaches such as CuffLinks (234) could be used in parallel to ascertain which genes change their promoter usage or change their isoform expression pattern.

The presence of lipid and hormone processing pathways in the analysis presented here suggests that urothelial cells either actively produce or metabolise hormones. AKR1C1 inactivates progesterone, which is present in the urine in both males and females. As progesterone is a known mitotic stimulant (235), urothelial cells *in vivo* which encounter progesterone in the urine may need to counter this stimulus to maintain quiescence.

The association of innate immune-system genes with induction of differentiation in the RNA-seq likely reflects the role the differentiated urothelium has in repelling infections, as do other epithelia. The innate immune response at epithelial surfaces is mediated by secreted immunoglobulin A (IgA). Human bladder urothelium has been reported to have expression of secretory IgA at the luminal surface (236), and IgA is detectable in the urine of healthy donors (237). IgA is transported across the epithelia by the polymeric immunoglobulin receptor (PIGR), which was not associated with the gene ontology analysis but had significantly upregulated transcripts in urothelium differentiated at 144 h. PIGR contributes to mucosal epithelial innate immunity by transporting immunoglobulins produced in the lamina propria across the cell layers for secretion into the lumen (238). PIGR expression has been reported in human urothelium, but only at the superficial surface(239). If PIGR functions in urothelium as it does in other epithelia, it might be expected to be present in all the cell layers. As this is not the case in normal quiescent tissue, the route of IgA transport is unclear.

Of the transcription factor genes upregulated at 24 h, GATA3, STAT1, HIF1A and NFE2L2 have sequence-specific binding motifs in the database used by PSCAN to assess for enrichment of potential transcription factor binding sites in the

promoters of chosen genes. Only HIF1A, as part of the DNA-binding dimer HIF1A:ARNT, had enriched binding sites present in promoters of differentially regulated genes. HIF1A:ARNT binding sites were enriched in both up and downregulated genes at 24 h, and downregulated genes at 144 h, suggesting that if HIF1A has a role to play it may be inhibitory to some cellular processes.

The PSCAN promoter analysis tool has a maximum target range of 1000 bp upstream from designated transcription start sites, and also only returns significance values based on the enrichment over the entire list. Whilst this is useful for overall enrichment, it can overlook more subtle aspects of gene regulation such as if small subsets of genes within the list have promoter regions enriched for specific factors not present in others. Transcription factor analysis tools such as the promoter analysis and interaction network (PAINT) (240) can scan up to 5000 bp and assign the most enriched binding within sub-sets of scanned regions, excluding genes for which no enrichment of known factors is seen. Using PAINT to analyse the 2000 bp region upstream of genes upregulated at 24 and 144 h returned a smaller subset of factors enriched in a sub-set of genes with enriched factors which bore little resemblance to the factors highlighted by PSCAN (Supplementary Materials 3.1.2.1 and 3.1.2.2). Rather than PPARγ:RXRα being common to genes upregulated at both time points, ETS1 and ELK1 were common factors present only in a sub-set of genes. The user-directed nature of analysing transcription factor binding site enrichment and the differences between such algorithms make the data obtained from this type of analysis variable, and is one reason why the data on actual binding occurrences obtained from approaches such as FAIRE-seq are favoured.

5.4.2 ChIP-Seq, FAIRE and HiC

Histone modifications subjected to ChIP showed the expected pattern at promoters of expressed and non-expressed genes. Results were not fully processed at the time of writing, but initial mapping of peaks showed there was similarity between the peaks from each target histone modification. This provides evidence that the samples are internally consistent, and that the DNA captured by ChIP to each histone modification was different to the other. Clustering analysis of peaks was complete at the time of writing for one sample (24 h control cells). The peaks in the H3K4me3 sequencing clustered around promoters of highly expressed

genes, and peaks for H3K27me3 clustered around non-expressed and low-level expression genes (Appendix Figure 7.2.2.1).This sequencing approach should therefore yield novel insights into the distribution of the two histone marks in *in vitro* differentiating NHU cells.

FAIRE samples also showed internal consistency in the sequences present in the DNA. This consistency was expected as FAIRE peaks have been found to be similar between cells and tissues. Within cell types FAIRE peaks which change are generally observed around genes which change expression. Such peaks were observed in the UCSC density tracks of FAIRE data from NHU cells.

The preliminary results from the HiC sequencing data have produced some chromosome dot plots which look remarkably similar to each other, as do those in previous publications (71, 241). This is indicative that the samples will yield good quality data. At the time of writing data analysis was ongoing.

5.5 Conclusions and Future Work

At the time of writing, work on data processing all aspects of these sequencing projects was ongoing. Results thus far are encouraging that the data will be of good quality and provide a rich resource to further the understanding of the biology of urothelial differentiation.

The RNA-seq data was the most complete of these datasets at the time of writing, and one analysis was performed which highlighted known differentiation-associated genes, and revealed genes from other aspects of urothelial cell biology that are intertwined with the development of differentiation. Understanding if each of these parts of the differentiated urothelial phenotype (lipid processing, innate immune system, cytoskeletal components) is under the influence of the same or different transcription factor networks will aid understanding of the differentiated urothelial phenotype. It is possible to perform such transcription factor network analyses using predictions of transcription factor binding sites, but it is much more beneficial to have additional information about actual protein-DNA binding events which occur around upregulated genes, such as will be provided by the FAIRE data once data processing is completed.

Single time-point ChIP-seq maps have been made of histone modifications in primary cultured human urothelial cells(219), but the cells were undifferentiated. If the results from the ChIP-seq performed are of good quality after final processing, then these would be the first the first histone maps to identify active and repressive chromatin marks unique to differentiated urothelial cells. The use of H3K4me3 and H3K27me3 as ChIP targets will allow genes which are repressed, expressed and rare genes which are poised for expression (promoters contain both marks) to be elucidated (242).

The expanding number of publications in the fields related of chromosome organisation has revealed that there is a similar organisation to the human genome across multiple human cell types at the 0.1-1 MB scale. Regions of co-regulated genomic features are arranged in topologically-associated domains (TAD), which are separated by insulating genomic and protein factors that contain the spread of histone and DNA-associated modifications (243). This allows regions of the genome to be co-regulated by a change in factors which bind the local chromatin environment, but for this regulation to be contained within the target region. Although the majority of the architecture will be similar across all

the time-points tested, other groups have identified specific regions associated with a change in cell phenotype (207, 209).

No other publications could be found in the literature where HiC and FAIRE datasets had been combined. This dataset therefore has the potential to provide novel insight into the relationship between chromatin composition and architecture.

Many possibilities exist for analysis and validation of this dataset, but some pertinent experiments which could be done are:

- FAIRE data can be used to build up a genome-wide picture of motifs with bound protein using the approach of Giannopoulo *et al* (199)

- This chromatin composition identified by FAIRE can be correlated with ChIP-seq maps and RNA-seq to determine which chromatin complexes are targeting which genes by assessing which complexes affect the transcriptional status of the surrounding genes within the transcriptional domains identified by HiC.

- Any motifs which are found in FAIRE data which correspond to unknown factors could be further investigated by targeting the motif for immunoprecipitation with biotinylated bait DNA. The captured DNA-protein complexes could be subjected to proteomic analysis, and the likely binding protein(s) elucidated by comparing protein identifications to those from parallel non-related immunoprecipitations (232).

- Chromatin-binding sites in the FAIRE data could be validated by ChIP-PCR to one or more factors, such as PPARγ.

- Genes marked with the "poised for transcription" histone marks (both H3K4me3 and H3K27me3) which are not actively transcribed in differentiated cells could be those which are important in reversing the differentiated phenotype. Expression of these genes and / or their protein products could be tested under a situation where differentiated cells are stimulated to divide, such as wound-healing.

6 Thesis Overview and Conclusions

PPARγ expression was very complex in NHU cells, but overall localisation of was relatively unaffected by differentiation. Several PPARγ isoforms were extraction resistant in NHU cells independent of the induction of differentiation, and in the presence of ERK1/2 signalling at early time points (6 h) as evidenced in this study by western blot.

The isoform which was nuclear under these conditions was not the 50 kDa PPARγ isoform investigated previously by Varley *et al* (26), but a larger 58 kDa isoform which has been detected in other studies of NHU cells (122) but not investigated as it was not recognised as the 58 kDa isoform.

PPARγ2 siRNA did not reduce abundance of PPARγ isoforms, but was more effective than PPARγ1/2 siRNA at preventing the upregulation of PPARγ isoforms, CK13 and CLDN4. As the PPARγ2 siRNA was more efficient, and PPARγ2 mRNA was much lower in abundance as compared to PPARγ1 mRNA as detected by both PCR and RNA-seq, this could indicate that whichever isoform is PPARγ2 is only transcriptionally active between the time when it is translated and when it is sequestered or broken down.

The PPARγ at 58 kDa was reactive with an anti-PPARγ2 antibody and was constitutively present in NHU cells, but its near total sequestration in the extraction-resistant fraction indicates that its transcriptional potential is tightly controlled. This resistance to extraction suggests that PPARγ isoforms may be permanently associated with the nucleoskeleton due to either endogenous ligands initiating binding to the DNA but not initiating transcription, or via association with proteins such as SAFB1 which were detected in by mass spectrometry in NHU extracts.

As NHU cells must form a heterogeneous population of cells when they re-differentiate, it may be necessary for some to actively inhibit the differentiation process to maintain their basal or intermediate identity. This heterogeneity, as observed by uneven AQP3 distribution in fully differentiated cultures may underlie the complexities of PPARγ isoform expression, with some cells possessing the full-length isoforms and others having smaller isoforms.

Such complex splicing is controlled at the post-translational level by the splicing machinery and heterogeneous ribonucleoproteins. Ribonucleoproteins and SMARC family proteins were as identified as differentially regulated between in control and differentiating NHU cells in both transcriptomics and mass spectrometric data. Further investigation of a selection of those observed to be changing such as SMARCC2, SMARCB1, SMARCE1, SMARCA2, HNRNPK, HNRNPAB, HNRNPC and DDX5 would be required to reveal of these mechanisms have a pivotal role to play in urothelial differentiation. If any of these proteins could be identified by mass spectrometry as preferentially co-precipitating with PPARγ from nuclear co-IP extracts of differentiated or control cells, then they may warrant further investigation as potential co-activators or repressors.

Mass spectrometric datasets identified a rich sub-nuclear proteome. Laminins, lamins and fibronectins dominated DNase extracts in NHU cells, and were more abundant in extracts from non-differentiated cells than differentiated ones. The laminins are well known to influence cell identity. As they are more prone to extraction after nucleases, this suggests that their structure is intertwined with that of the DNA. In this way they may be serving to maintain urothelial cells in a non-differentiated phenotype. Suppression of expression of one or more of the subunits of laminin 5 would be able to answer this question.

Post-translational modification of PPARγ is a well known to be involved in transrepression of inflammation-associated immune system genes whose expression is promoted by NF-κB (119, 128-132). As many genes associated with the innate immune system were observed to have increased expression in differentiating NHU cells, if patients with interstitial cystitis have altered levels of this potentially modified PPARγ, they may be less able to recover from inflammation. This potentially modified version of PPARγ has been observed to be absent in a range of tumour cell lines (Appendix Figure 1.2.2). In addition, NF-κB sequence-specific motifs were found to be enriched in the promoters of some genes with altered expression levels in differentiated NHU cells.

The next-generation sequencing datasets created in this study hold the potential to answer many potential questions. Use of single values for genes will more clearly reveal the genes that change their overall abundance of transcript, rather than just single transcripts. Combining approaches such as CuffDiff to isolate

differential splice variants of genes with FAIRE data could help to precisely identify the chromatin-binding factors which influence single genes, providing that the binding motif of the protein is known, or can be identified captured using affinity bait DNA.

Building up a genome-wide picture of the binding motifs present in the FAIRE holds great promise for isolating regulatory elements unique to differentiating urothelial cells. The single region which was investigated to UPK2 yielded peaks unique to differentiated cells enriched in SNAI1, which is known to influence gene expression in several other differentiation systems (176, 244).

Both the proteomic and genomic datasets in this thesis contain computational inferences of the presence and differential abundance of proteins and genes. These data were used to generate statistical estimates about differences in the expression of RNA, or protein abundance in sub-cellular fractions, in control and differentiation-induced urothelium. Since the observations are based on statistical methods, they are subject to false-positive and false-negative errors inherent to such calculations. As such, the observations are in need of further validation in the context of both the estimated abundance of proteins and genes, as well as in relation to any potential biological role in urothelial differentiation of differentially abundant candidates.

The proteomic identifications in this study are based on the analysis of MS/MS data generated from samples taken from a single donor line, and as such can only be considered discovery-stage data. To validate observations in the study, replicate extracts could be prepared and candidate differentially present proteins could be quantitated using measured peptide intensity after selected reaction monitoring of proteotypic peptides, with the aim of increasing the number of peptides in the target proteins with quantitation. A sub-set of the identified proteins could then optionally be further validated by western blot if suitable antibodies are available. Relative quantitation of any candidate differentially present proteins using either of the aforementioned approaches should be normalised to core histone variants, which should be present in similar amounts in all extracts.

To ascertain whether these proteins confirmed as having differential abundance between the chosen extracts are reproducibly differentially present across donor lines will require repetition of the experiments. Replicate numbers required for

any study depend on the variation inherent in the biological and technical systems being interrogated. Previous studies using 2D gel-based studies aiming to identify required numbers of replicates needed to demonstrate differential expression across donors have shown that based on the observed biological variation in measured intensity, in order to achieve 80% statistical power at a two-fold change in protein abundance required a minimum of 7 replicates, increasing to 10 replicates to show a 1.5-fold change (15). Modelling of the effects of total combined biological and variation on the number of replicates needed to demonstrate a chosen effect size for a particular protein have shown that combined variation needs to be as low as 20% to allow four biological replicates to achieve a power of 80% and a p-value ≤0.05 for2-fold changes in relative protein abundance (29). In order to make a power calculation to estimate the number of replicate donors needed to obtain enough measurements to control the false discovery rate of differential abundance, the variation between intensity measurements in technical replicates in the current study could be used as a guide as to the extent of technical likely variation, but in order to accurately estimate biological variation further pilot studies using new donor samples are needed (11).

At the time of writing, the RNA-seq data represents the most complete genomics dataset produced in this study. The statistical assessment of the RNA-seq showed that many transcripts from known differentiation-associated genes were robustly differentially regulated across the donor lines. Further validation of a selection the known differentially-expressed genes in the RNA-seq using quantitative PCR on cDNA libraries produced from the same RNA samples used for sequencing would validate the statistical assessment of the RNA-seq, and justify further validation of potentially novel differentiation-associated transcripts.

RNA-seq data can be used to infer modulation of signalling pathway activity via expression or repression of key members of the pathways themselves, or their known targets. This can be performed using software such as Ingenuity (http://www.ingenuity.com), or through gene ontology analysis tools such as GSEA by searching for members of canonical pathways in the Kyoto encyclopaedia of genes and genomes (KEGG). Alternately, known interactions between differentially regulated genes could be sought through database searching engines such as STRING (http://string-db.org). If any interaction networks or canonical pathways are found which are potentially associated with the development of the differentiated phenotype, then the role of these candidate pathways in

differentiation could be elucidated by targeting important proteins within the networks with pharmacological modulation, or by upregulating their expression using viral transfection, or downregulating its expression with siRNA or shRNA.

7 Appendix

7.1 PPARγ in Differentiation of Human Urothelium in vitro

7.1.1 PPARγ in Previous Studies of NHU Cells

Previous observations of PPARγ in undifferentiated NHU cells, NHU cells induced to differentiate for 24 h with TZ and PD, and differentiation-induced 3T3-L1 pre-adipocytes (Figure 7.1.1.1).
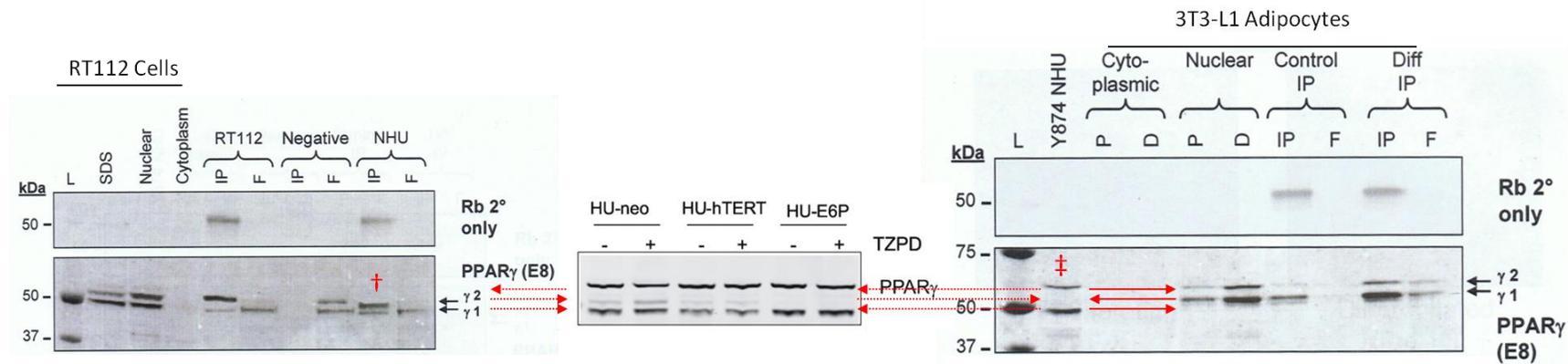
Figure 7.1.1.1. PPARγ in NHU cells and pre-adipocytes. ‡ = whole cell lysate of undifferentiated NHU cells, * = whole cell lysate of NHU cells treated for 24 h with TZ and PD, † = nuclear co-IP extracts of NHU cells treated for 24 h with TZ and PD. Left blot: Georgopoulos *et al* showed the upregulation of the presumed 52 kDa PPARγ isoform in NHU (HU-neo) cells, a response which was absent in immortalised NHU cells (HU-h TERT and HU-E6P). Centre blot: Fleming (122) compared the molecular weights of PPARγ present in whole cell lysates (Y874 NHU) of undifferentiated NHU cells with those present in nuclear co-IP extracts and subsequent immunoprecipitations from proliferating (P) and differentiating. The largest isoform of PPARγ in NHU cells appears equivalent to the PPARγ from 3T3-L1 cells. The smallest isoform of PPARγ in NHU lysates in this blot appears smaller than the PPARγ1 in 3T3-L1 cells. Right blot: Fleming (122) also compared PPARγ from whole cell lysates of differentiated NHU cells (SDS) with nuclear extracts of NHU (nuclear) and immunoprecipitations from nuclear extracts of NHU and RT112 cells. The two closely migrating PPARγ isoforms resemble the ones observed in nuclear extracts from NHU cells the current study and those observed by Georgopoulos *et al* (left blot). The inconsistent presence of the largest PPARγ isoforms in the two whole cell lysates known to be prepared using SDS (centre and right blots above) correlates with the variable solubility of the same band observed in the current study. This suggests that the 50 kDa isoform of PPARγ is a distinct isoform not present in 3T3-L1 cells, the 52 kDa isoform is equivalent in molecular weight to the PPARγ1 in 3T3-L1 cells, and the largest observed isoform in NHU cells is equivalent to the PPARγ2 in NHU cells.

7.1.2   Nuclear co-IP Extractions and Differentiation of NHU Cells

PPARγ (E8)
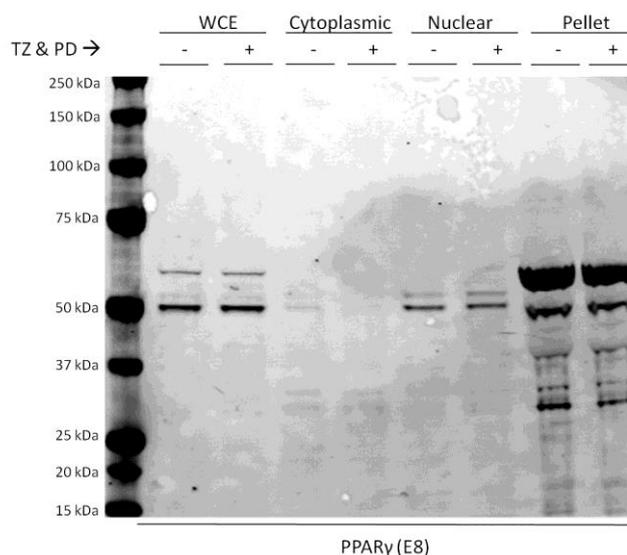
Figure 7.1.2.1. Active Motif Nuclear co-IP Extractions cell line Y967. NHU Cell line Y967 was treated with 0.1% DMSO vehicle control (-) or induced to differentiate with 1 μM troglitazone and 1 μM PD153035 (+) for 24 h before subjecting to fractionation using the Active Motif Nuclear Co-IP kit. 20 μg of whole cell extracts (WCE) cultured in parallel, 20 μL hypotonic lysis fraction (cytoplasmic), 20 μL nuclease-extracted (nuclear) and 20 μL non-solubilised (pellet) fractions were separated by western blot and membranes labelled for PPARγ. PPARγ distribution was similar for both control and differentiated cells.

A

Claudin 5

B
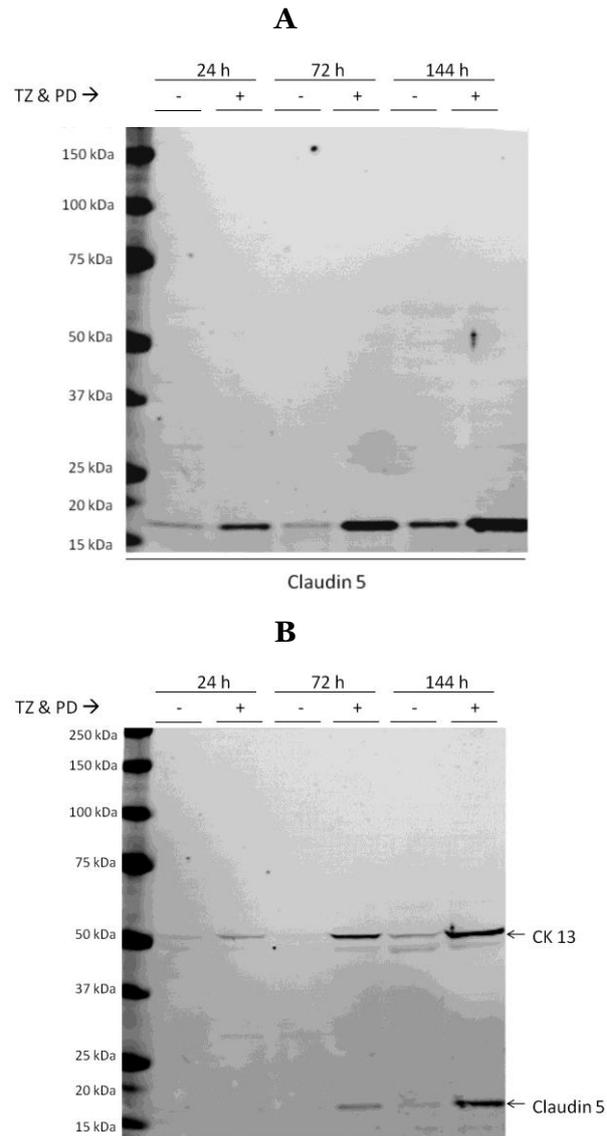
Figure 7.1.2.2 A, B. Induction of Expression of Urothelial Differentiation Markers Claudin 5 (A) and Cytokeratin 13 (CK13) (B) in NHU Cell Line Y967. Cells were treated with 0.1% DMSO vehicle control (-) or induced to differentiate with 1 μM TZ and 1 μM PD (+) for 24, 72 and 144 h. WCE were taken and 20 μg separated by western blot and membranes labelled for Claudin 5 and then CK13. Both CK13 and Claudin 5 labelling became more intense in cells treated with TZ and PD over the time course.
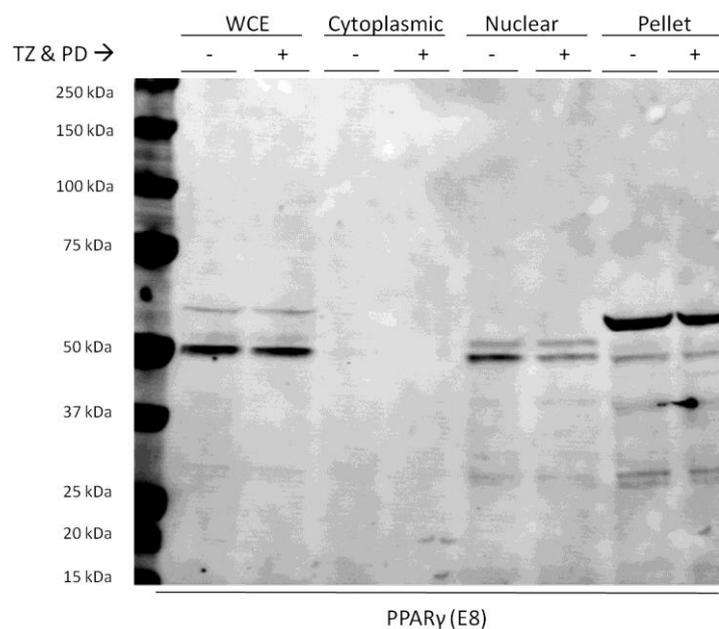
Figure 7.1.2.3. Active Motif Nuclear Extractions NHU Cell line Y1077. Cells were treated with 0.1% DMSO vehicle control (-) or induced to differentiate with 1 μM troglitazone and 1 μM PD153035 (+) for 24 h before subjecting to fractionation using the Active Motif Nuclear Co-IP kit. 20 μg of whole cell extracts (WCE) cultured in parallel, 20 μL hypotonic lysis fraction (cytoplasmic), 20 μL nuclease-extracted (nuclear) and 20 μL non-solubilised (pellet) fractions were separated by western blot and membranes labelled for PPARγ. PPARγ distribution was similar for both control and differentiated cells.

A

Claudin 5

B
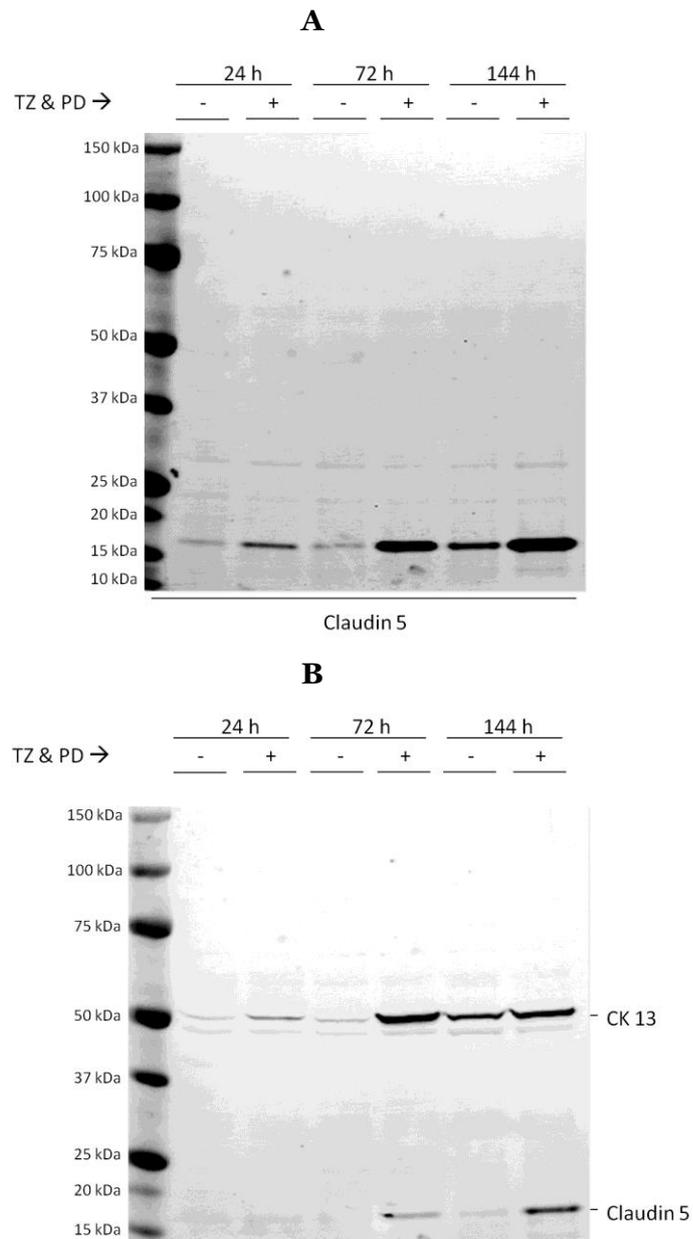
CK 13

Claudin 5

Figure 7.1.2.4 A, B. Induction of differentiation in Y1077. Expression of Urothelial Differentiation Markers Claudin 5 (A) and Cytokeratin 13 (CK13) (B) in NHU Cell Line Y1077. Cells were treated with 0.1% DMSO vehicle control (-) or induced to differentiate with 1 µM TZ and 1 µM PD (+) for 24, 72 and 144 h. WCE were taken and 20 µg separated by western blot and membranes labelled for Claudin 5 and then CK13. Both CK13 and Claudin 5 labelling became more intense in cells treated with TZ and PD over the time course.

7.1.3   Comparison of Millipore and Pierce IP Systems
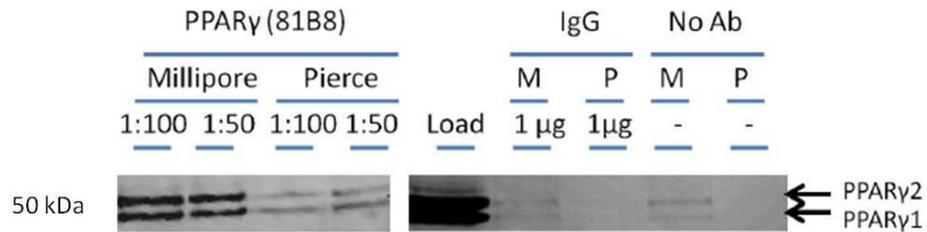
Figure 7.1.3.1 Immunoblot showing capture of PPARγ (50 and 52 kDa) using two dilutions of antibody and negative controls, with comparison of Pierce and Millipore agarose supports in IP from Nuclear co-IP chromatin extracts. Comparison of 20 µg chromatin extract loading control (load) to eluates from agarose with varying concentrations or amounts of (1 µg / 1:50) of reversibly bound (Millipore) or covalently conjugated (Pierce) antibodies. Millipore and Pierce agarose gels both retain target proteins. Millipore agarose gave low-level signal for PPARγ in the IgG and antibody-free controls, suggesting high background levels of non-specific retention.
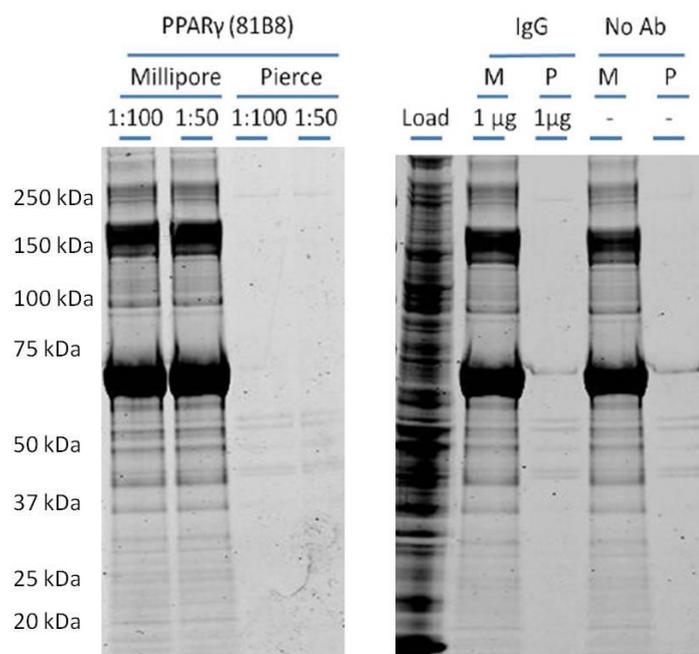
Figure 7.1.3.2 Coomassie stained SDS-PAGE gel of eluates from antibody and control immunoprecipitations performed with antibodies bound to Pierce or Millipore kits. 20 µL (of 75 µL) eluates from Millipore and Pierce comparison IP. Pierce columns show much lower retention of proteins than the Millipore. The pattern of bands in the lane containing eluate from the antibody-free (No Ab) incubation is similar to the staining in other lanes, suggesting the majority of bands visible are from non-specifically retained proteins.

7.1.4 PPARγ in Sequential Salt CSK Extractions
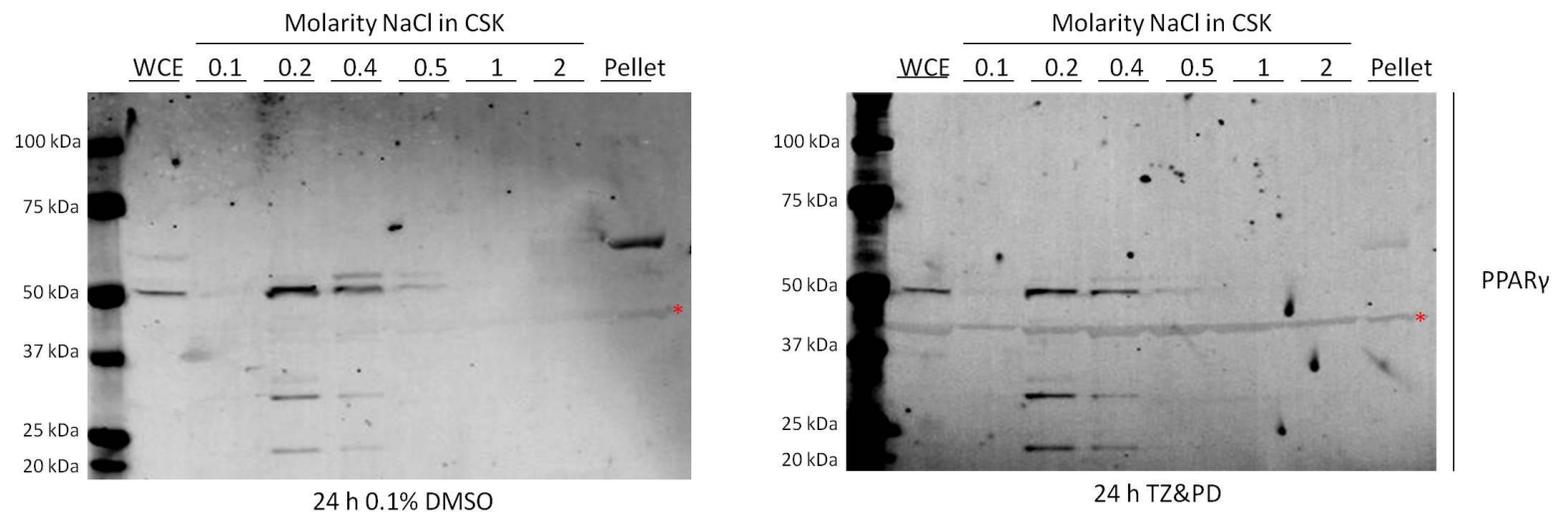
Figure 7.1.4.1. PPARγ localisation in sequential salt extracts at 24 hours after induction of differentiation. Control (0.1% DMSO, left blot) and differentiation-induced (TZ&PD, right blot) cells were sequentially extracted in CSK buffer with 0.1% Triton-X100 and NaCl concentrations ranging from 0.1 to 2 M. Extraction profiles of PPARγ-reactive bands were not affected by induction of differentiation. Red asterisk denotes Beta-actin labelling which was present as a result of cross-contamination of PPARγ antibody aliquots for western blots. Cell line Y1152, passage 5.
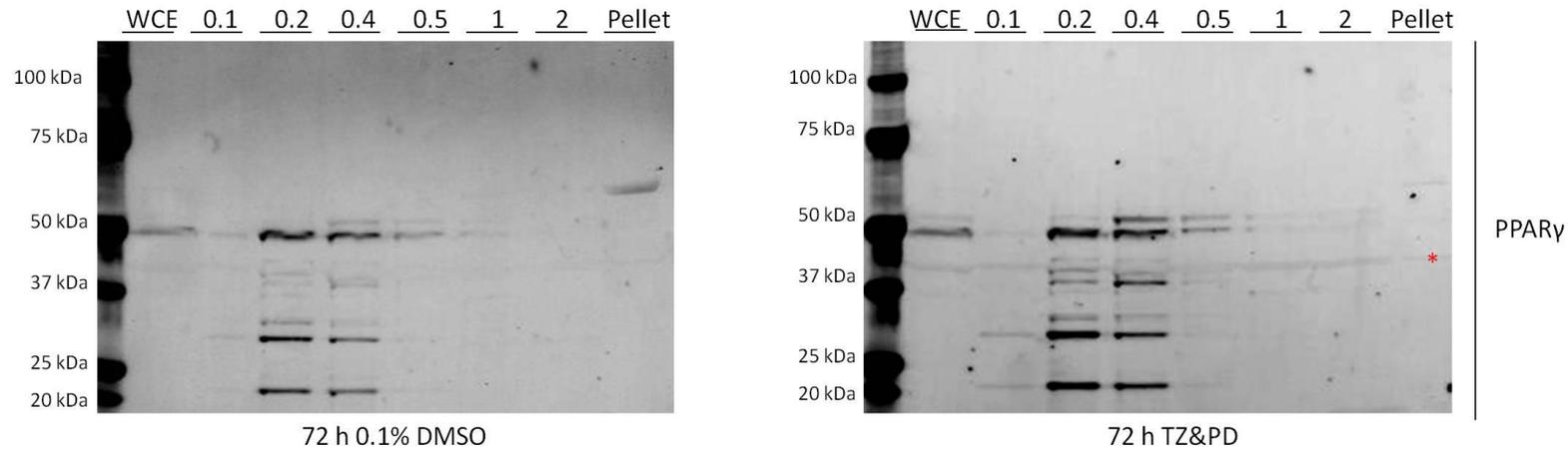
Figure 7.1.4.2. PPARγ localisation in sequential salt extracts at 72 hours after induction of differentiation. Control (0.1% DMSO, left blot) and differentiation-induced (TZ&PD, right blot) cells were sequentially extracted in CSK buffer with 0.1% Triton-X100 and NaCl concentrations ranging from 0.1 to 2 M. Extraction profiles of PPARγ-reactive bands were not affected by induction of differentiation. Red asterisk denotes Beta-actin labelling which was present as a result of cross-contamination of PPARγ antibody aliquots for western blots. Cell line Y1152, passage 5.

7.1.5 Expression of PPARγ in Bladder Tumour Cell Lines
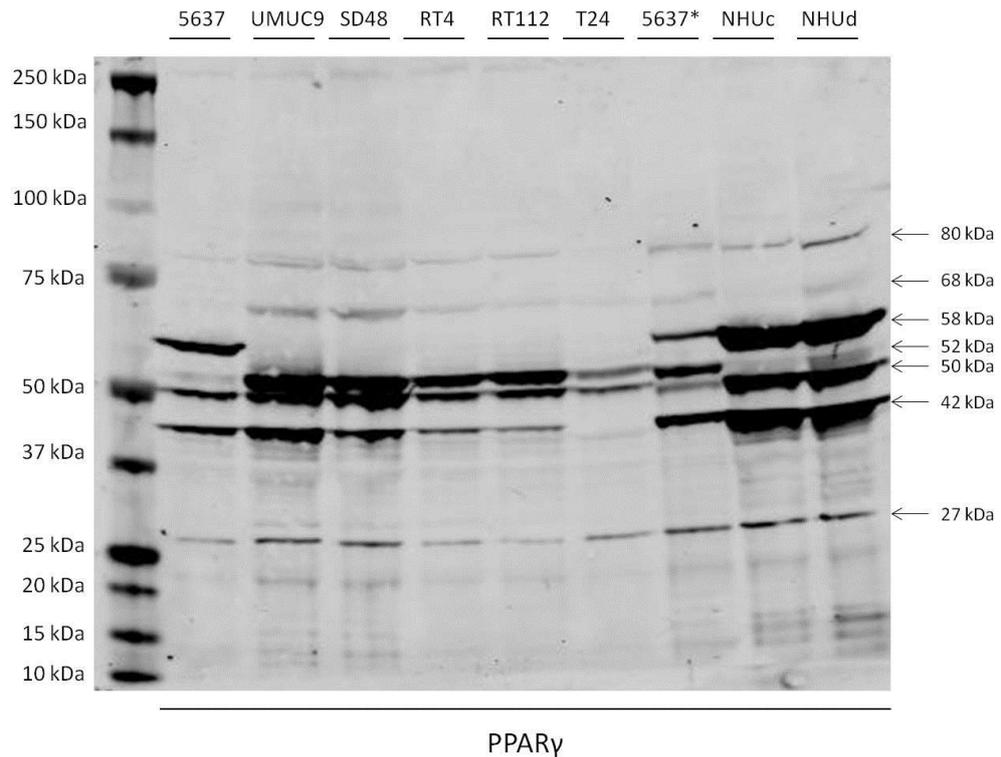
Figure 7.1.5.1. PPARγ in whole cell lysates (20 μg) of bladder-derived tumour cell lines, and proliferative (NHUc) and differentiated (NHUd) NHU cells. 5637* indicates cell line has been adapted to culture in medium containing adult bovine serum (ABS), which instigated an upregulation of 52 kDa PPARγ. 58 kDa is absent in all tumour cell lines with the exception of 5637. A 68 kDa band similar to that seen in immunoprecipitations of NHU cells is observed at 68 kDa, with another faint band at 80 kDa visible. T24 cells are missing both the 58 kDa and the majority of the 42 kDa isoform. Image courtesy Shu Guo, Jack Birch Unit.

7.1.6 Densitometry Calculations for siRNA western blots
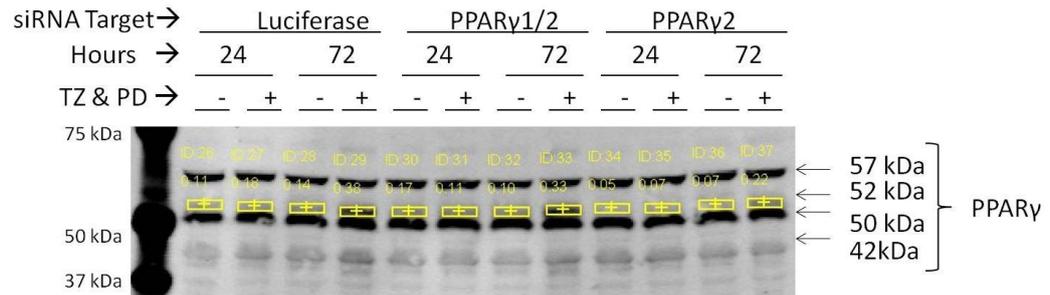
Figure 7.1.6.1. Densitometry measurements for 52 kDa PPARγ from siRNA experiments. Calculations for normalisation in supplementary materials
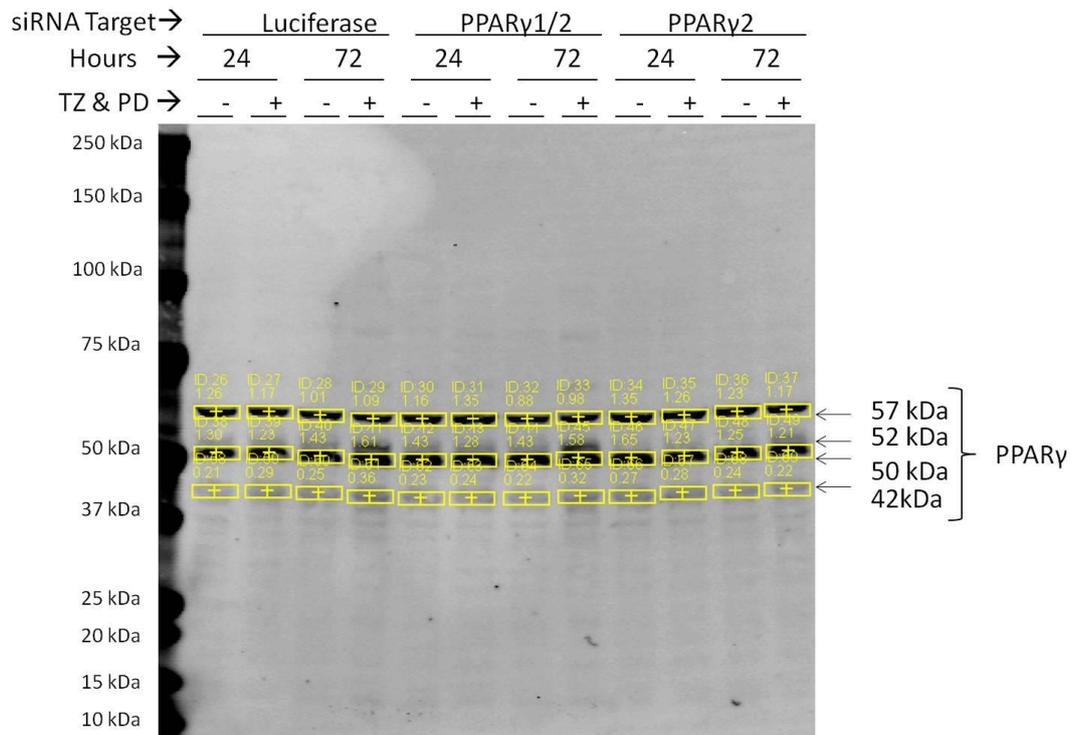
Figure 7.1.6.2 Densitometry measurements for 42, 50 and 58 kDa PPARγ from siRNA experiments. Calculations for normalisation in supplementary materials
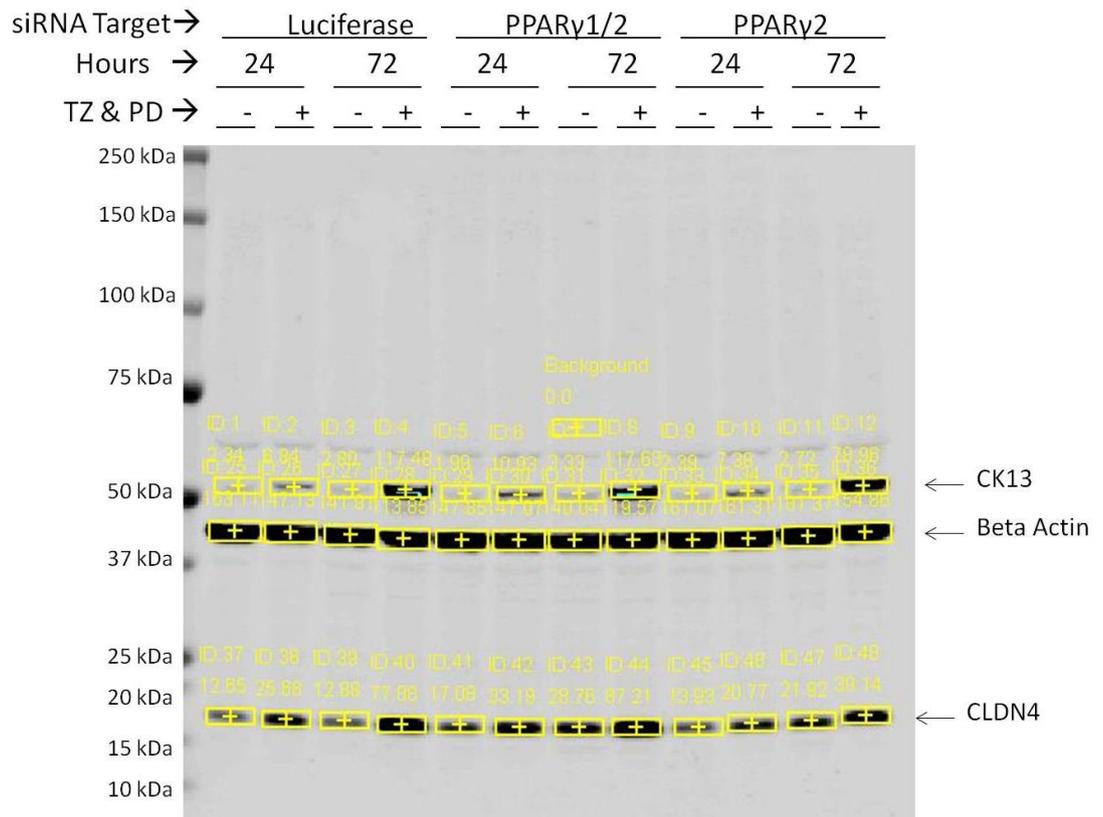
Figure 7.1.6.3. Densitometry measurements for CK13, Beta actin and CLDN4 from siRNA experiments. PPARγ blot from Appendix figures 7.1.6.1./7.1.6.2 was stripped of antibodies to allow re-labelling with above antibodies and labelling of beta actin on the same membrane. Calculations for normalisation in supplementary materials.

7.2 Next Generation-Sequencing

7.2.1 Sequencing Read Numbers

Numbers of sequenced reads obtained and mapped to the genome from RNA-seq, ChIP-seq and FAIRE libraries. HiC sequence numbers had not been communicated by TGAC at the time of writing.

Library name	patient / treatment	reads/ paired reads	library	read length	insert	%mapped UCSC hg19	%mapped GRCh37 ensembl
LIB2636	Y967_24D	23,841,618	single-end	101	391	94.67%	94.78%
LIB2637	Y967_24T	19,383,257	single-end	101	390	94.69%	94.81%
LIB2638_LDI2294	Y967_144D	41,270,129	single-end	51	391	97.96%	98.02%
LIB2639_LDI2295	Y967_144T	34,000,483	single-end	51	393	97.87%	97.94%
LIB2640	Y1192_24D	24,614,609	single-end	101	390	94.83%	94.93%
LIB2641	Y1192_24T	28,390,829	single-end	101	391	80.92%	80.97%
LIB2655_LDI2298	Y1192_144D	28,039,739	single-end	51	392	97.07%	97.13%
LIB2656_LDI2299	Y1192_144T	26,050,398	single-end	51	387	97.16%	97.23%
LIB2657	Y1214_24D	22,602,983	single-end	101	397	93.70%	93.77%
LIB2658	Y1214_24T	17,623,866	single-end	101	385	93.67%	93.74%
LIB2659_LDI2302	Y1214_144D	29,675,689	single-end	51	390	97.46%	97.51%
LIB2660_LDI2303	Y1214_144T	27,423,455	single-end	51	393	97.16%	97.21%

Table 7.2.1.1. RNA-seq read counts. For patient / treatment 24D = 24 h DMSO, 24T = 24 h TZ&PD, 24D = 144 h DMSO, 24T = 144 h TZ&PD.

FAIRE-seq						%mapped UCSC hg19	
Library Name	patient / treatment	reads/ paired reads	library	read length	insert	properly paired	after dedup
LIB2804	Y1192_24D	40,398,642	paired-end	101	248	79.11	61.78
LIB2805	Y1192_24T	46,801,388	paired-end	101	249	93.23	80.09
LIB2806	Y1192_144D	45,618,946	paired-end	101	248	85.73	68.91
LIB2807	Y1192_144T	36,687,886	paired-end	101	248	81.05	60.94
LIB4142	Y967_24D	54,072,365	paired-end	101	280	96.03	80.60
LIB4143	Y967_24T	61,126,547	paired-end	101	284	95.82	86.96
LIB4144	Y967_144D	58,269,690	paired-end	101	289	95.90	77.94
LIB4145	Y967_144T	58,021,821	paired-end	101	268	95.93	82.59
LIB4146	Y1214_24D	62,602,781	paired-end	101	287	97.01	82.00
LIB4147	Y1214_24T	57,897,996	paired-end	101	288	96.07	71.21
LIB4148	Y1214_144D	61,587,468	paired-end	101	292	96.78	85.06
LIB4149	Y1214_144T	55,065,049	paired-end	101	273	96.24	85.51

Table 7.2.1.2. FAIRE-seq read counts. For patient / treatment 24D = 24 h DMSO, 24T = 24 h TZ&PD, 24D = 144 h DMSO, 24T = 144 h TZ&PD.

ChIP-seq					%mapped UCSC hg19	
	patient / treatment	reads/ paired reads	library	read length	properly paired	after dedup
LIB4107	Y1192_24D_ChIP_K4	14,563,316	single-end	51	72.96%	8.14
LIB4108	Y1192_24T_ChIP_K4	15,790,225	single-end	51	83.31%	11.93
LIB4109	Y1192_144D_ChIP_K4	15,458,572	single-end	51	71.26%	10.46
LIB4110	Y1192_144T_ChIP_K4	14,476,060	single-end	51	78.95%	10.76
LIB4111	Y1192_24D_ChIP_K27	13,323,034	single-end	51	90.68%	26.70
LIB4112	Y1192_24T_ChIP_K27	15,000,451	single-end	51	88.80%	30.04
LIB4113	Y1192_144D_ChIP_K27	18,184,951	single-end	51	90.74%	29.86
LIB4114	Y1192_144T_ChIP_K27	15,750,323	single-end	51	90.72%	36.44

Table 7.2.1.3. ChIP-seq read counts. For patient / treatment 24D = 24 h DMSO, 24T = 24 h TZ&PD, 24D = 144 h DMSO, 24T = 144 h TZ&PD. K4 = H3K4me3 and K27 = H3K27 me3 libraries, respectively.

7.2.2 Expression levels at Genes and Promoters Near Peaks from H3K4me3 and H3K27me3 ChIP-seq

24 h Control
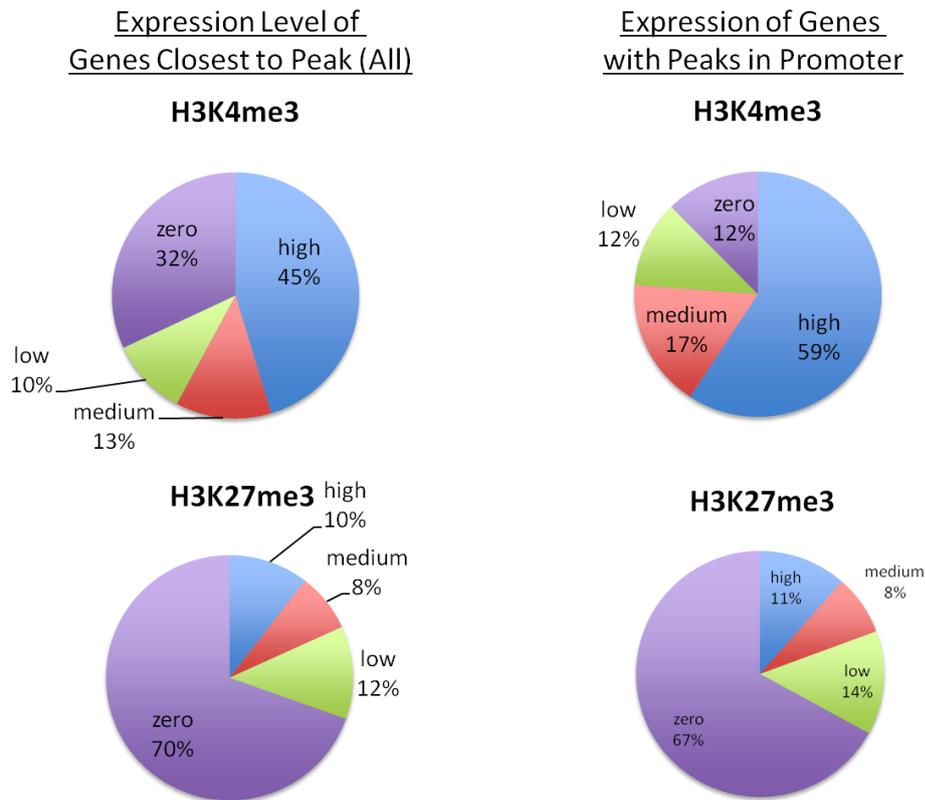
Figure 7.2.2.1. Expression of genes with H3K3me3 and H3K27me3 ChIP-seq peaks. Expression levels of genes nearest to peaks, or expression of genes with peaks within -/+ 500 bp of transcription start site were assessed. Genes were split into zero, low, medium and high expressing categories based on RNA-seq results from matched sample. H3K4me3 clusters around highly expressed genes and promoters, whereas H3K27me3 clusters around genes with low expression. ChIP-seq peaks from 24 h control sample used.

7.3 Mass Spectrometry

7.3.1 Peptides Accepted for Quantitation vs. Total CSK-DNase

Peptide Intensity and Cumulative Frequency of Peptides Accepted for Quantitation CSK-DNase Extracts NHUc vs NHUd
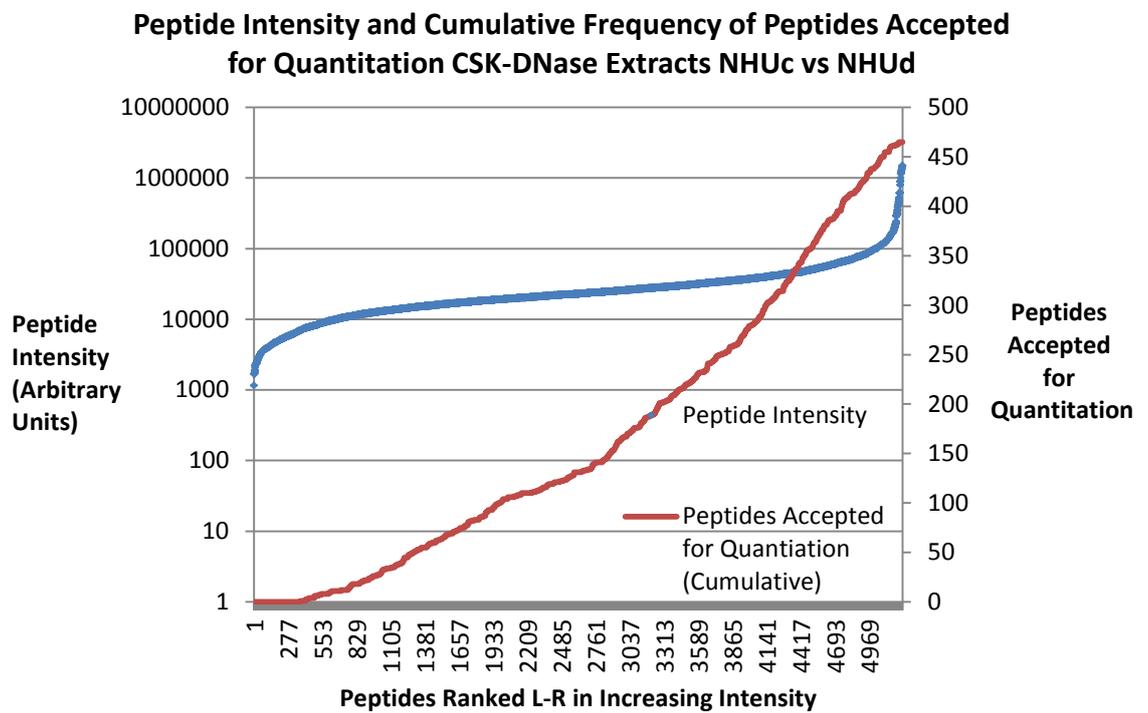
Figure 7.3.1.1. Peptides accepted for quantitation in CSK-DNase extracts. Total peptides in merged DDA of all CSK-DNase extracts ranked L-R (x-axis) in increasing intensity (blue line, left y-axis). Cumulative numbers of peptides accepted for quantitation (red line, right y-axis) were slightly skewed towards high intensity peptides, with over half having intensity of >28,000 in the merged DDA.

7.3.2 Mascot Scores and p-value of LF-MS intensity Differences 0.1-2 M NaCl
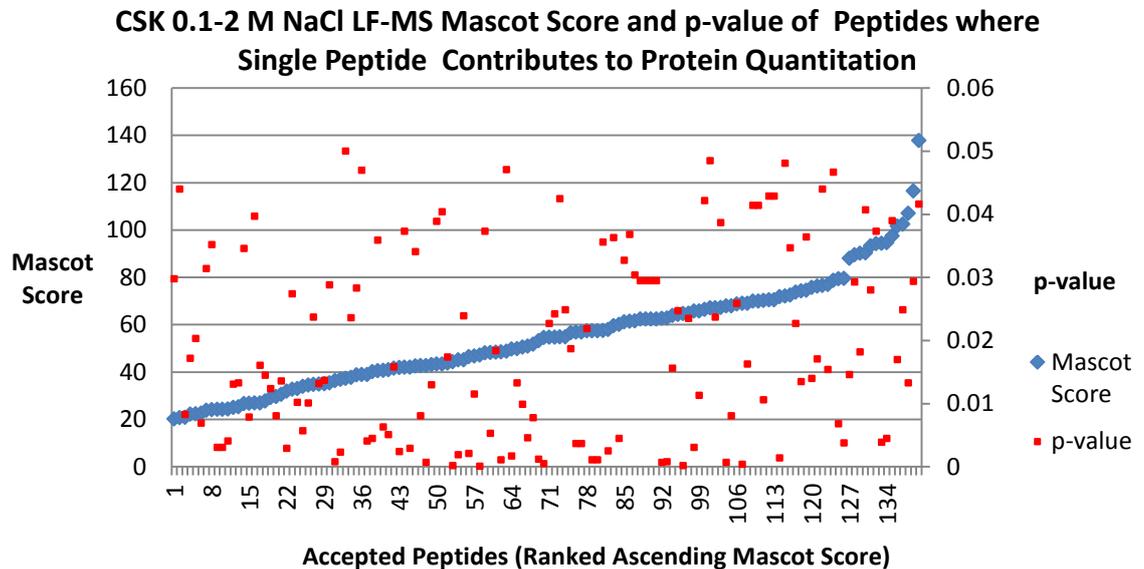
Figure 7.3.2.1. Mascot score of proteins with only one peptide accepted for LF-MS quantitation in CSK 0.1-2 M NaCl. Proteins ranked L-R (x-axis), in order of ascending Mascot score (blue dot, left hand y-axis) accepted for LF-MS quantitation had p-value of ≤0.05 (red dots, right hand axis). Mascot score appeared to bear no relationship with significance of p-value.

7.3.3 Proteins in CSK 0.1-2 M NaCl Extracts Associated with Transcription

Abbreviation, Mascot Score (# Peptides)	Name	Select Description (from UniProt KB)
APEX1 141 (4)	DNA-(apurinic or apyrimidinic site) lyase	Controls DNA binding affinity of FOS/JUN by influencing redox status of the transcription factors (165).
BASP1. 556 (8)	Brain acid soluble protein 1	Transcriptional co-suppressor for Wilms' tumour suppressor protein WT1 (245).
C1QBP, 612 (4)	Complement component 1, q subcomponent binding	Human p32, interacts with B subunit of the CCAAT-binding factor, CBF/NF-Y, and inhibits CBF-mediated transcription activation in vitro (246).
CSDA, 97 (3)	Cold shock domain protein A	Cold shock domain proteins repress transcription from the GM-CSF promoter (247).
CTNNB1, 34 (1)	Catenin (cadherin-associated protein), beta 1, 88kDa	Chromatin-specific regulator of LEF-1 transcription (248).
DDX1, 103 (1)	DEAD (Asp-Glu-Ala-Asp) box helicase 1	Interacts with RelA and enhances nuclear factor kappaB-mediated transcription (249).
DDX5, 671 (11)	DEAD (Asp-Glu-Ala-Asp) box helicase 5	p68 is a androgen receptor coactivator (250). Interacts with Runx2 and regulates osteoblast differentiation (185).
ENO1, 692 (9)	Enolase 1, (alpha)	cMYC binding transcriptional suppressor protein (251).
HMGA1, 136 (1)	High mobility group AT-hook 1	Cooperates with the p16(INK4a) tumour suppressor to promote proliferative arrest and stabilize senescence by contributing to the repression of proliferation-associated genes
JUP, 784 (10)	Junction plakoglobin	Plakoglobin (gamma-catenin) has TCF/LEF family-dependent transcriptional activity (252).
MTDH, 29 (1)	metadherin	Activates NFκB (253).
NACA, 275 (7)	Nascent polypeptide-associated complex alpha subunit	Electronically inferred through sequence similarity to have a role in RNA transcription. Has antiproliferative effects in human CD8+ T-cells (254).

NPM1, 1373 (8)	Nucleophosmin	NFκB coactivator (255).
PBXIP1, 62 (1)	Pre-B-cell leukaemia homeobox interacting protein 1	Inhibits the binding of PBX1-HOX complex to DNA and blocks the transcriptional activity of E2A-PBX1. Tethers estrogen receptor-alpha (ESR1) to microtubules (256, 257).
PDLIM1, 618 (7)	PDZ and LIM domain 1	Protein with both cytoskeletal interacting and transcription factor zinc finger domains, considered to act as scaffold to assemble cellular signal-controlling molecules at certain subcellular sites (258).
RAN, 50 (2)	RAN, member RAS oncogene family	Negative regulator of androgen receptor (259).
RFXAP, 21 (1)	Regulatory factor X-associated protein	Forms cooperative DNA binding complexes with X2BP and CBF/NF-Y. Associates with CIITA to form an active transcriptional complex (260).
SND1, 1309 (21)	Staphylococcal nuclease and tudor domain containing 1	Functions as a bridging factor between STAT6 and the basal transcription factor (261). Plays a role in PIM1 regulation of MYB activity (262). Functions as a transcriptional coactivator for the Epstein-Barr virus nuclear antigen 2 (EBNA2) (263).
SRSF2, 202 (6)	Serine/arginine-rich splicing factor 2	Corepressor of CBF1, binds to PAP-1 and effects alternative splicing.
TRIM28, 673 (8)	Tripartite motif containing 28	Represses transcription factors with KRAB domains (264). Recruits histone deacetylases and methylases.
YWHAB, 282 (8)	Tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein	Adapter protein which modulates signalling activity of tyrosine kinases. Stimulates osteogenesis (265).

Table 7.3.3.1 Proteins in CSK extracts with known transcription factor or co-factor activity. Proteins in CSK 0.1 – 2 M NaCl fraction with GO annotations of transcription factor or co-factor activity are listed. Statements about protein function concerning transcriptional activity and associated publications were taken from the UniProt.org database.

7.3.4 Transcription Related Proteins in CSK-DNase Extracts

Description of transcription-related proteins detected in CSK-DNase extracts. Proteins from all control and DNase extractions are included.

Abbreviation, Mascot Score (# Peptides)	Name	Select Description (from UniProt KB)
CTNNB1, 126 (5)	Catenin, beta 1, 88kDa	Chromatin-specific regulator of LEF-1 transcription (248). Involved in wnt signalling.
CTNND1, 55 (3)	Catenin delta-1	p120 catenin. Binds to and inhibits the transcriptional repressor ZBTB33, which may lead to activation of target genes of the Wnt signaling pathway. Implicated both in cell transformation by SRC and in ligand-induced receptor signaling through the EGF, PDGF, CSF-1 and ERBB2 receptors.
DDX5, 671 (11)	DEAD (Asp-Glu-Ala-Asp) box helicase 5	p68 is a androgen receptor coactivator (250). Interacts with Runx2 and regulates osteoblast differentiation (185).
EEF1A1, 442 (7)	Elongation factor 1-alpha 1	With PARP1 and TXK, forms a complex that acts as a T helper 1 (Th1) cell-specific transcription factor and binds the promoter of IFN-gamma to directly regulate its transcription.
EIF4A3, 93.5 (6)	Eukaryotic initiation factor 4A-III	Core component of the exon junction complex which mediates RNA splicing.
FLNA, 1598.5 (30)	Filamin-A	Actin binding protein which sequesters transcription factors such as FOXC1 (266).
HNRNPAB, 51.3 (3)	Heterogeneous nuclear ribonucleoprotein A/B	Part of mRNP granule complex (267).

Abbreviation, Mascot Score (# Peptides)	Name	Select Description (from UniProt KB)
HNRNPD, 75.1 (2)	Heterogeneous nuclear ribonucleoprotein D	Binds to double- and single-stranded DNA sequences in a specific manner and functions a transcription factor (268).
HNRNPK, 898.5 (7)	Heterogeneous nuclear ribonucleoprotein K	When SUMOylated, acts as a transcriptional coactivator of p53/TP53, playing a role in p21/CDKN1A and 14-3-3 sigma/SFN induction. As far as transcription repression is concerned, acts by interacting with long intergenic RNA p21 (lincRNA-p21), a non-coding RNA induced by p53/TP53.
HNRPDL, 34.1 (2)	Heterogeneous nuclear ribonucleoprotein L	Binds to DNA and acts as a repressor of the COX5 gene by binding to the CATR sequence (269).
HNRNPUL1, 32.6 (1)	Heterogeneous nuclear ribonucleoprotein U-like protein 1	Initiates transcription from glucocorticoid promoters in the absence of ligand when complexed with BRD7 (270).
HSPA8, 2738 (16)	HSPA8	Acts as a repressor of transcriptional activation. Inhibits the transcriptional coactivator activity of CITED1 on SMAD mediated transcription (271).
ILF2, 165.9 (3)	Interleukin enhancer-binding factor 2	Complexes with ILF3. Functions as both a positive and negative regulator of gene expression in mammalian cells (272)
ILF3, 110.9 (4)	Interleukin enhancer-binding factor 3	Complexes with ILF2. Functions as both a positive and negative regulator of gene expression in mammalian cells (272)

Abbreviation, Mascot Score (# Peptides)	Name	Select Description (from UniProt KB)
JUP, 985.1 (11)	Junction plakoglobin	Plakoglobin (gamma-catenin) has TCF/LEF family-dependent transcriptional activity (252).
LMNA, 5491.2 (35)	Prelamin-A/C	Plays an important role in nuclear assembly, chromatin organization, nuclear membrane and telomere dynamics (273).
PHB, 79.4 (4)	Prohibitin.	Inhibits DNA synthesis. Has a role in regulating proliferation (274).
PHB2, 236.3 (5)	Prohibitin 2.	Acts as a mediator of transcriptional repression by nuclear hormone receptors via recruitment of histone deacetylases. Interacts with PHB, ESR1, HDAC1 and HDAC5 (275).
PML, 1481 (20)	Protein PML	Regulates transcription activity of ELF4 (276). Regulates PTEN compartmentalization through the inhibition of USP7-mediated deubiquitination.(277).
POLR2A, 110 (5)	DNA-directed RNA polymerase II subunit RPB1	DNA-dependent RNA polymerase subunit.
POLR2B, 75.4 (3)	DNA-directed RNA polymerase II subunit RPB2	DNA-dependent RNA polymerase subunit.
PSMC5, 35.1 (1)	26S protease regulatory subunit 8	Interacts with the thyroid hormone receptor, retinoid X receptor (RXR), NDC80, PAAF1 and TRIM5 (278).
PTRF, 207.4 (3)	Polymerase I and transcript release factor	PTRF is required for dissociation of the ternary transcription complex

Abbreviation, Mascot Score (# Peptides)	Name	Select Description (from UniProt KB)
RAD21, 34.6 (3)	Double-strand-break repair protein rad21 homolog	Chromatin protein involved in chromosome cohesion during cell cycle, in DNA repair, and in apoptosis.
RBM14, 28.4 (1)	RNA-binding protein 14	Isoform 1 and 2 function as a nuclear receptor coactivator and repressor via interactions with coactivators such as NCOA6 and CITED.
RBMX, 638.9 (8)	RNA-binding motif protein, X chromosome	Interacts with SAFB/SAFB1. Forms a complex with ILF2, ILF3, YLPM1, KHDRBS1, NCOA5 and PPP1CA.
RPS27A, 423.9 (2)	Ubiquitin-40S ribosomal protein S27a	Ubiquitin. Ligation of ubiquitin subunits to proteins leads to their degradation by the proteasome. Degradation plays an important part in receptor recycling.
RPS3, 118.4 (4)	40S ribosomal protein S3	Negative regulation of NFκB transcription.
SAFB, 52.5 (4)	Scaffold attachment factor B1	When associated with RBMX, binds to and stimulates transcription from the SREBF1 promoter. Can function as an estrogen receptor corepressor and can also bind to the HSP27 promoter and decrease its transcription.
SHOX, 27.2 (1)	Short stature homeobox protein	Transcription factor highly expressed in osteoblasts (279).
SMARCA4, 40.3 (2)	Transcription activator BRG1	Transcriptional coactivator cooperating with nuclear hormone receptors to potentiate transcriptional activation.
SMARCC2, 54.9 (2)	SWI/SNF complex subunit SMARCC2	Component of 6 multiprotein chromatin-remodeling complexes: Swi/Snf-A (BAF), Swi/Snf-B (PBAF), Brm, Brg1(I), WINAC and Brg1(II)

Abbreviation, Mascot Score (# Peptides)	Name	Select Description (from UniProt KB)
SNRPB, 30.5 (1)	Small nuclear ribonucleoprotein-associated proteins B and B'	May have a functional role in the pre-mRNA splicing or in snRNP structure. Binds to the downstream cleavage product (DCP) of histone pre-mRNA in a U7 snRNP dependent manner
SP100, 200.1 (5)	Nuclear autoantigen Sp-100	PML body member and co-repressor of ETS family transcription factors.
SRSF6, 35.6 (1)	Serine/arginine-rich splicing factor 6	Plays a role in constitutive splicing and can modulate the selection of alternative splice sites. Represses the splicing of MAPT/Tau exon 10.(280)
SRSF7, 32.7 (1)	Serine/arginine-rich splicing factor 7	Plays a role in constitutive splicing and can modulate the selection of alternative splice sites. Represses the splicing of MAPT/Tau exon 10.(280)
TARDBP, 337 (4)	TAR DNA-binding protein 43	DNA and RNA-binding protein which regulates transcription and splicing. Defects in TARDBP are the cause of amyotrophic lateral sclerosis type 10 (ALS10). Expression is high in pancreas, placenta, lung, genital tract and spleen
TMPO, 27.6 (1)	Lamina-associated polypeptide 2, isoforms beta/gamma	May be involved in the control of initiation of DNA replication through its interaction with NAKAP95. Possible receptor for attachment of lamin filaments to the inner nuclear membrane.
TPR, 119.5 (6)	Nucleoprotein TPR	Component of the cytoplasmic fibrils of the nuclear pore complex implicated in nuclear protein import. Its N-terminus is involved in activation of oncogenic kinases. Plays a role in the mitotic spindle checkpoint.

Abbreviation, Mascot Score (# Peptides)	Name	Select Description (from UniProt KB)
TRIM28, 673 (8)	Tripartite motif containing 28	Represses transcription factors with KRAB domains (264). Recruits histone deacetylases and methylases.
XRCC5, 34.3 (2)	X-ray repair cross-complementing protein 5	Plays a key role in non-homologous end joining. The XRCC5/6 dimer together with APEX1 acts as a negative regulator of transcription in response to extracellular calcium (281).
YBX1, 117.2 (2)	Nuclease-sensitive element-binding protein 1	Regulates the transcription of numerous genes. Its transcriptional activity on the multidrug resistance gene MDR1 is enhanced in presence of the APEX1 acetylated form (282).
ZNF326, 91.1 (1)	DBIRD complex subunit ZNF326	Core component of the DBIRD complex, a multiprotein complex that acts at the interface between core mRNP particles and RNA polymerase II (RNAPII) and integrates transcript elongation with the regulation of alternative splicing. May play a role in neuronal differentiation and is able to bind DNA and activate expression in vitro (283).

Table 7.3.4.1 Proteins associated with DNA dependent transcription in CSK-DNase extracts.

7.3.5 Proteins Identified in SPL Generated Using Progenesis LC-MS

Protein	Peptides	Scores
LAMA3 laminin alpha 3 subunit isoform 2	21	691.1
FLNB Isoform 1 of Filamin-B	18	553
FLNA Uncharacterized protein	10	310.9
SPTAN1 Isoform 1 of Spectrin alpha chain, brain	8	259.1
BAT1 Isoform 2 of Spliceosome RNA helicase BAT1	6	293.8
EVPL Envoplakin	4	69.6
RBMXL1 RNA binding motif protein, X-linked-like 1	4	90.5
ABCD3 Uncharacterized protein	2	45.2
ATAD3A ATPase family AAA domain-containing protein 3A isoform 3	2	39.6
HIST2H3D;HIST2H3C;HIST2H3A Histone H3.2	2	66.3
MAVS Isoform 1 of Mitochondrial antiviral-signaling protein	2	45.3
MCM2 DNA replication licensing factor MCM2	2	55.7
SAFB2 Scaffold attachment factor B2	2	54
SMARCA1 SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 1, isoform CRA_d	2	35
SYNCRIP Isoform 1 of Heterogeneous nuclear ribonucleoprotein Q	2	34.6
TOR1AIP1 Isoform 2 of Torsin-1A-interacting protein 1	2	48.7
CHD2 Isoform 2 of Chromodomain-helicase-DNA-binding protein 2	1	38
EEF1A1 EEF1A protein (Fragment)	1	47.3
EFTUD2 116 kDa U5 small nuclear ribonucleoprotein component	1	48.2
FUS Isoform Short of RNA-binding protein FUS	1	47.5
LCP1 Plastin-2	1	43.3
PKM2 Isoform M1 of Pyruvate kinase isozymes M1/M2	1	46.3
POR cDNA FLJ59656, highly similar to NADPH--cytochrome P450 reductase	1	34.6
PSMD3 26S proteasome non-ATPase regulatory subunit 3	1	36.1
RUVBL1 Isoform 1 of RuvB-like 1	1	73.5
SLC9A1 Isoform 1 of Sodium/hydrogen exchanger 1	1	53.1
SVIL Isoform 2 of Supervillin	1	41.7
TEX10 Testis-expressed sequence 10 protein	1	35.9

H2AFY H2A histone family, member Y isoform 2	9	441.2
H2AFY2 Core histone macro-H2A.2	5	155.2
HNRNPM Isoform 1 of Heterogeneous nuclear ribonucleoprotein M	5	219.5
PML Isoform PML-1 of Probable transcription factor PML	5	95.6
CTNNA1 Isoform 1 of Catenin alpha-1	4	90.8
HIST2H4B;HIST1H4K;HIST1H4H;HIST1H4A;HIST1H4L;HIST1H4B;HIST1H4I;HIST1H4D;HIST1H4E;HIST4H4;HIST1H4F;HIST1H4C;HIST1H4J;HIST2H4A Histone H4	4	154.1
HNRNPUL2 Heterogeneous nuclear ribonucleoprotein U-like protein 2	4	109.8
ILF3 Isoform 5 of Interleukin enhancer-binding factor 3	4	135
DDX21 Isoform 1 of Nucleolar RNA helicase 2	3	34.6
HNRNPF Heterogeneous nuclear ribonucleoprotein F	3	126.5
HNRNPH1 Heterogeneous nuclear ribonucleoprotein H	3	132.4
NUP205 Nuclear pore complex protein Nup205	3	57.9
POLR2B DNA-directed RNA polymerase II subunit RPB2	3	59.7
TRIM28 Isoform 1 of Transcription intermediary factor 1-beta	3	45.8
CTNNB1 Isoform 1 of Catenin beta-1	2	95.8
CTNND1 Isoform 1AB of Catenin delta-1	2	41.8
HIST1H2AL;HIST1H2AM;HIST1H2AJ;HIST1H2AI;HIST1H2AG;HIST1H2AK;HIST1H2AE;HIST1H2AB;HIST1H2AD Histone H2A type 1-B/E	2	216.1
HIST2H2BE Histone H2B type 2-E	2	37.4
HNRNPAB Isoform 2 of Heterogeneous nuclear ribonucleoprotein A/B	2	64.8
HNRNPL Heterogeneous nuclear ribonucleoprotein L	2	42.5
HNRNPU Isoform Short of Heterogeneous nuclear ribonucleoprotein U	2	117.5
ILF2 Interleukin enhancer-binding factor 2	2	56
NOP56 Nucleolar protein 56	2	92.2
PHB2 Prohibitin-2	2	40.9
PTBP1 Isoform 1 of Polypyrimidine tract-binding protein 1	2	37.4
SAFB Scaffold attachment factor B1	2	52.1
TPR Nucleoprotein TPR	2	41.3
UBC;UBB;RPS27A ubiquitin and ribosomal protein S27a precursor	2	128.6
AIFM1 Isoform 1 of Apoptosis-inducing factor 1, mitochondrial	1	47.9
EIF4A3 Eukaryotic initiation factor 4A-III	1	41
FBL rRNA 2'-O-methyltransferase fibrillarin	1	40.4

HNRNPA2B1 Isoform B1 of Heterogeneous nuclear ribonucleoproteins A2/B1	1	43.6
HNRNPC Isoform C1 of Heterogeneous nuclear ribonucleoproteins C1/C2	1	39.7
HNRNPK Isoform 1 of Heterogeneous nuclear ribonucleoprotein K	1	54.2
POLR2A DNA-directed RNA polymerase II subunit RPB1	1	36
PTRF Isoform 1 of Polymerase I and transcript release factor	1	47.7
RAD21 Double-strand-break repair protein rad21 homolog	1	35.4
SF3B1 Splicing factor 3B subunit 1	1	37.5
SF3B3 Isoform 1 of Splicing factor 3B subunit 3	1	35.9
SP100 Isoform Sp100-HMG of Nuclear autoantigen Sp-100	1	51.3
TARDBP TDP43	1	34.9
PLEC1 Isoform 3 of Plectin-1	56	1425.1
DSP Isoform DPI of Desmoplakin	21	247.9
LAMC2 Isoform Long of Laminin subunit gamma-2	21	541.1
LAMB3 Laminin subunit beta-3	17	719
FN1 Isoform 1 of Fibronectin	16	324
ACTB Actin, cytoplasmic 1	14	828.4
SPTBN1 Isoform Long of Spectrin beta chain, brain 1	14	282.9
ATP5A1 ATP synthase subunit alpha, mitochondrial	13	630.4
ACTN1 Alpha-actinin-1	12	315.9
ATP5B ATP synthase subunit beta, mitochondrial	11	623.2
ACTN4 Alpha-actinin-4	10	572.6
LMNA Isoform A of Lamin-A/C	9	287.6
ACTA1 Actin, alpha skeletal muscle	8	289.7
ATP1A1 Isoform Long of Sodium/potassium-transporting ATPase subunit alpha-1	8	318.6
ATP2A2 Isoform SERCA2A of Sarcoplasmic/endoplasmic reticulum calcium ATPase 2	7	441.6
CKAP4 Isoform 1 of Cytoskeleton-associated protein 4	7	252
HSPA8 Isoform 1 of Heat shock cognate 71 kDa protein	7	181.8
KTN1 Isoform 1 of Kinectin	6	120.6
VCP Transitional endoplasmic reticulum ATPase	6	164.8
HSPA5 78 kDa glucose-regulated protein	5	91.8
JUP Junction plakoglobin	4	143.4
NDUFS1 NADH-ubiquinone oxidoreductase 75 kDa subunit	4	184.6
SLC25A5 ADP/ATP translocase 2	4	102.2
ABCD3 Isoform 1 of ATP-binding cassette sub-family D member 3	3	40.8

HSPA1A;HSPA1B Heat shock 70 kDa protein 1	3	66.6
HSPD1 60 kDa heat shock protein, mitochondrial	3	63.9
ITGB4 Isoform Beta-4C of Integrin beta-4	3	43.2
JUP Uncharacterized protein	3	67.3
KRT10 Keratin, type I cytoskeletal 10	3	53.9
KRT19 Keratin, type I cytoskeletal 19	3	41.1
KRT7 Keratin, type II cytoskeletal 7	3	153.6
LIMA1 Isoform Beta of LIM domain and actin-binding protein 1	3	108.4
MATR3 Matrin-3	3	57.2
RPN1 Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 1 precursor	3	158.9
RPN2 Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 2	3	81.5
THBS1 Thrombospondin-1	3	74.8
ATP1B1 Isoform 1 of Sodium/potassium-transporting ATPase subunit beta-1	2	113.5
CLTC Isoform 1 of Clathrin heavy chain 1	2	56.2
FLOT1 Flotillin-1	2	51.8
HADHA Trifunctional enzyme subunit alpha, mitochondrial	2	45.3
HSP90AA1 Isoform 2 of Heat shock protein HSP 90-alpha	2	55.9
IMMT Isoform 1 of Mitochondrial inner membrane protein	2	39.1
KRT18 Keratin, type I cytoskeletal 18	2	38.5
KRT2 Keratin, type II cytoskeletal 2 epidermal	2	87.3
KRT77 keratin 77	2	67.4
LEMD2 LEM domain-containing protein 2	2	33.4
NDUFV1 Isoform 1 of NADH dehydrogenase [ubiquinone] flavoprotein 1, mitochondrial	2	69.1
NNT NAD(P) transhydrogenase, mitochondrial	2	78.6
SLC1A5 Neutral amino acid transporter B(0)	2	90.5
SLC25A24 Isoform 1 of Calcium-binding mitochondrial carrier protein SCaMC-1	2	81.8
SNRNP200 Isoform 1 of U5 small nuclear ribonucleoprotein 200 kDa helicase	2	83.6
STT3A Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit STT3A	2	50
UQCRC2 Cytochrome b-c1 complex subunit 2, mitochondrial	2	34
VIM Vimentin	2	42.4
CD44 Isoform 12 of CD44 antigen	1	73.4

DDOST Dolichyl-diphosphooligosaccharide--protein glycosyltransferase 48 kDa subunit	1	32.1
DKC1 H/ACA ribonucleoprotein complex subunit 4	1	55.4
DLD Dihydrolipoyl dehydrogenase, mitochondrial	1	67.1
EEF1G;TUT1 cDNA FLJ56389, highly similar to Elongation factor 1-gamma	1	38.2
HSP90B1 Endoplasmin	1	69.4
KRT9 Keratin, type I cytoskeletal 9	1	90.9
LAMP2 Isoform LAMP-2A of Lysosome-associated membrane glycoprotein 2	1	35.4
LGALS3BP Galectin-3-binding protein	1	64.5
NUP155 Isoform 1 of Nuclear pore complex protein Nup155	1	53.3
PKP3 Plakophilin 3b (Fragment)	1	44.8
UQCRC1 Cytochrome b-c1 complex subunit 1, mitochondrial	1	51.1

Table 7.3.5 Proteins identified in SPL lists generated by Progenesis LC-MS. Proteins which were in the SPL data that had not been observed in DDA data are highlighted in bold. All potential nuclear proteins are in italics.

7.3.6 Proteins Identified by SPL from Bruker Software

NHUc DNase SPL			NHUd DNase SPL		
Protein	Peptides	Scores	Protein	Peptides	Scores
ACTB	6	437.2	ACTB	1	76.6
ATP5A1	4	724.2	ACTN1	1	42.9
ATP5B	6	714.1	ATP2A2	1	33.2
BAT1	**5**	**1530.3**	ATP5A1	4	1130.4
DLST	**1**	**33.3**	ATP5B	4	402.3
FN1	**1**	**28**	DDX5	1	30.1
H2AFY	1	56.4	FLNB	2	64.8
HNRNPA2B1	1	188.5	H2AFY	2	666.7
HNRNPCL1	**1**	**33.4**	H2AFY2	2	274
HSPA8	1	29.8	HNRNPF	2	77
LAMA3	6	223.9	HNRNPH1	2	62.8
LAMB2	4	330.9	HNRNPL	1	28.7
RBMXL1	**2**	**339.8**	HNRNPM	2	164.6
SDHA	**1**	**30.4**	JUP	1	51.3
SLC25A24	3	120.2	NDUFS1	1	27.6
SNRNP70	1	25.6	RBMX	1	31.9
STOML2	**1**	**38.7**			
THBS1	1	290.2			

Table 7.3.6.1. Proteins identified in SPLs generated from Bruker software. Proteins not previously identified in DDAs are displayed in bold text.

7.4 Buffer Recipes

2% SDS Western Blot Buffer

2% SDS western blot solution: 20% (v/v) glycerol, 2% (w/v) SDS, 125 mM Tris-HCl (pH 6.8), 200 mM NaF, 0.1 mM Na_3PO_4, 33 mM Na_3PO_4, with freshly added 13 mM DTT and 1:100 dilution of protease inhibitor cocktail (Sigma Aldrich, Cat # P8340).

Tris Buffered Saline

10 mM Tris (adjusted to pH 7.4 with HCl) and 140 mM NaCl in H_2O

CSK

Stock cytoskeletal (CSK) buffer was made with 10 mM PIPES-KOH (pH 6.8), 300 mM Sucrose, 1 mM EGTA and 1 mM $MgCl_2$. Aliquots were made with 0.1 M NaCl added, or 0.1% Triton-X100 and 0.1, 0.2 0.4, 0.5, 1 or 2 M NaCl.

Transfer Buffer

20 % (v/v) methanol and 80 % (v/v) H_2O with final concentrations of 12 mM Tris and 96 mM glycine.

Pierce Direct IP Wash Buffer

25 mM Tris (pH 7.4), 150 mM NaCl, 1 mM EDTA, 1% (v/v) NP-40 substitute (Sigma Aldrich, Cat # 74385) and 5% (v/v) glycerol

TBS for IF

50mM Tris-HCl (pH 7.6), 150mM NaCl, 0.1% (w/v) NaN_3 and 0.1% (w/v) BSA.

Antifade for IF

5% N-propyl gallate (w/v) in 95% glycerol and 5% PBS)

HiC Lysis Buffer

10 mM Tris-HCl (pH 8.0), 10 mM NaCl, 0.2% NP-40 substitute (Sigma Aldrich, Cat # 74385)

NEBuffer 2

10 mM Tris-HCl (pH 7.9), 50 mM NaCl, 10 mM MgCl2, 1 mM DTT

Ligation Buffer

50 mM Tris-HCl (pH 7.5), 10 mM $MgCl_2$, 10 mM DTT

TE Buffer

10 mM Tris (pH 8.0), 1 mM EDTA

Tween Wash Buffer

5 mM Tris-HCl (pH8.0), 0.5 mM EDTA, 1 M NaCl and 0.05% Tween-20

ChIP Swelling Buffer

5mM PIPES (pH8) with 85 mM KCl in H_2O

ChIP Modified TE Buffer

2 mM EDTA, 20 mM Tris-HCl pH 8, 150 mM NaCl, 1% w/v SDS and 1% v/v Triton-X100

Radioimmunoprecipitation assay (RIPA) Buffer

150 mM NaCl, 1.0 % v/v NP-40 substitute, 0.5 % w/v sodium deoxycholate, 0.1% w/v SDS, and 50 mM Tris, pH 8.0

8 List of Abbreviations

3'/5' RACE	-	3 prime / 5 prime rapid amplification of cDNA ends
3C	-	Chromosome conformation capture
BHPrE	-	Benign human prostate epithelial cells
cDNA	-	Complementary DNA
ChIP-seq	-	Chromatin immunoprecipitation sequencing
CIS	-	Carcinoma *in situ*
Co-IP	-	Co-Immunoprecipitation
CSK	-	Cytoskeletal
DDA	-	Data directed analysis
DEPC	-	Diethylpyrocarbonate
DNA	-	Deoxyribonucleic acid
DNase	-	Deoxyribonuclease
DNTP	-	Deoxyribonucleotide
EGF	-	Epidermal growth factor
EGFR	-	Epidermal growth factor receptor
emPAI	-	Exponentially modified protein abundance index
ENCODE	-	Encyclopedia of DNA elements
FAIRE	-	Formaldehyde assisted isolation of regulatory elements
FASP	-	Filter aided sample preparation
FMF	-	Find molecular features
GO	-	Gene ontology
H3K27me3	-	Histone H3 trimethylated at lysine 27
H3K4me3	-	Histone H3 trimethylated at lysine 4
HiC	-	High throughput chromosome conformation capture
IgG	-	Immunoglobulin G
KSFM	-	Keratinocyte serum free medium
KSM	-	Keratinising squamous metaplasia
LC-MS	-	Liquid chromatography-mass spectrometry
LDS	-	Lithium dodecyl sulfate
LF-MS	-	Label free mass spectrometry
MNase	-	Micrococcal nuclease
MRES	-	Multiple regions of epigenetic silencing
mRNA	-	Messenger ribonucleic acid
MS	-	Mass spectrometry

MS/MS	-	Tandem mass spectrometry
NCBI	-	The National Center for Biotechnology Information
NHU	-	Normal Human Urothelial
NHUc	-	Normal human urothelial - control
NHUd	-	Normal human urothelial - differentiation
NM	-	Nuclear Matrix
NR	-	Nuclear receptor
PBS	-	Phosphate buffered saline
PCR	-	Polymerase chain reaction
PD	-	PD153035
RIPA	-	Radioimmunoprecipitation
RNA	-	Ribonucleic acid
RNase	-	Ribonuclease
RT-PCR	-	Reverse transcriptase polymerase chain reaction
SDS	-	Sodium dodecyl sulfate
SDS-PAGE	-	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
shRNA	-	Short hairpin RNA
siRNA	-	Small interfering RNA
SPL	-	Scheduled precursor list
TBS	-	Tris buffered saline
TLE	-	Tris low EDTA buffer
TWB	-	Tween wash buffer
TZ	-	Troglitazone

9 **References**

1.   Hicks R. The mammalian urinary bladderan accommodating organ. Biological Reviews. 1975;50:215-46.

2.   Varley CL, Garthwaite MAE, Cross W, Hinley J, Trejdosiewicz LK, Southgate J. PPAR -regulated tight junction development during human urothelial cytodifferentiation. Journal of cellular physiology. 2006;208(2):407.

3.   Sun TT, Liang FX, Wu XR. Uroplakins as markers of urothelial differentiation. Advances in experimental medicine and biology. 1999;462:7.

4.   Olsburgh J, Harnden P, Weeks R, Smith B, Joyce A, Hall G, et al. Uroplakin gene expression in normal human tissues and locally advanced bladder cancer. The Journal of pathology. 2003;199(1):41-9.

5.   Atiyeh BS, Costagliola M. Cultured epithelial autograft (CEA) in burn treatment: three decades later. Burns. 2007;33(4):405-13.

6.   Rama P, Matuska S, Paganoni G, Spinelli A, De Luca M, Pellegrini G. Limbal stem-cell therapy and long-term corneal regeneration. New England Journal of Medicine. 2010;363(2):147-55.

7.   Macchiarini P, Jungebluth P, Go T, Asnaghi M, Rees LE, Cogan TA, et al. Clinical transplantation of a tissue-engineered airway. The Lancet. 2008;372(9655):2023-30.

8.   Atala A, Bauer SB, Soker S, Yoo JJ, Retik AB. Tissue-engineered autologous bladders for patients needing cystoplasty. The Lancet. 2006;367(9518):1241-6.

9.   Shin K, Lee J, Guo N, Kim J, Lim A, Qu L, et al. Hedgehog/Wnt feedback supports regenerative proliferation of epithelial stem cells in bladder. Nature. 2011;472(7341):110-4.

10.   Kurzrock EA, Lieu DK, Chan CW, Isseroff RR. Label-retaining cells of the bladder: candidate urothelial stem cells. American Journal of Physiology-Renal Physiology. 2008;294(6):F1415-F21.

11.   Cairns DA, Barrett JH, Billingham LJ, Stanley AJ, Xinarianos G, Field JK, et al. Sample size determination in clinical proteomic profiling experiments using mass spectrometry for class comparison. Proteomics. 2009;9(1):74-86.

12.   Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, et al. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell. 2007;131(5):861-72.

13.   Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell. 2006;126(4):663-76.

14.   Oottamasathien S, Wang Y, Williams K, Franco OE, Wills ML, Thomas JC, et al. Directed differentiation of embryonic stem cells into bladder tissue. Developmental biology. 2007;304(2):556-66.

15.   Gandhi KS, McKay FC, Diefenbach E, Crossett B, Schibeci SD, Heard RN, et al. Novel approaches to detect serum biomarkers for clinical response to interferon-β treatment in multiple sclerosis. PLoS ONE. 2010;5(5):e10484.

16.   Anumanthan G, Makari JH, Honea L, Thomas JC, Wills ML, Bhowmick NA, et al. Directed differentiation of bone marrow derived mesenchymal stem cells into bladder urothelium. The Journal of urology. 2008;180(4):1778-83.

17.   Mauney JR, Ramachandran A, Yu RN, Daley GQ, Adam RM, Estrada CR. All-Trans Retinoic Acid Directs Urothelial Specification of Murine Embryonic Stem Cells via GATA4/6 Signaling Mechanisms. PLoS ONE. 2010;5(7):e11513.

18.   Burridge PW, Keller G, Gold JD, Wu JC. Production of de novo cardiomyocytes: human pluripotent stem cell differentiation and direct reprogramming. Cell stem cell. 2012;10(1):16-28.

19.   Murry CE, Keller G. Differentiation of embryonic stem cells to clinically relevant populations: lessons from embryonic development. Cell. 2008;132(4):661-80.

20.   Wichterle H, Lieberam I, Porter JA, Jessell TM. Directed differentiation of embryonic stem cells into motor neurons. Cell. 2002;110(3):385-97.

21.   Popov B, Zaĭchik A, Bud'ko M, Zlobina O, Tolkunova E, Zhidkova O, et al. Epithelial intestine cells transdifferentiate into bladder urothelium in experiments in vivo]. Tsitologiia. 2011;53(4):332.

22.   Mattout A, Biran A, Meshorer E. Global epigenetic changes during somatic cell reprogramming to iPS cells. Journal of molecular cell biology. 2011;3(6):341-50.

23.   Kim K, Doi A, Wen B, Ng K, Zhao R, Cahan P, et al. Epigenetic memory in induced pluripotent stem cells. Nature. 2010;467(7313):285-90.

24.   Li H, Collado M, Villasante A, Strati K, Ortega S, Cañamero M, et al. The Ink4/Arf locus is a barrier for iPS cell reprogramming. Nature. 2009;460(7259):1136-9.

25.   Southgate J, Hutton K, Thomas D, Trejdosiewicz L. Normal human urothelial cells in vitro: proliferation and induction of stratification. Laboratory investigation; a journal of technical methods and pathology. 1994;71(4):583.

26.   Varley C, Stahlschmidt J, Lee W, Holder J, Diggle C, Selby P, et al. Role of PPARgamma and EGFR signalling in the urothelial terminal differentiation programme. Journal of cell science. 2004;117(Pt 10):2029.

27.   Cross W, Eardley I, Leese H, Southgate J. A biomimetic tissue from cultured normal human urothelial cells: analysis of physiological function. American Journal of Physiology- Renal Physiology. 2005;289(2):F459.

28.   Reznikoff C, Johnson M, Norback D, Bryan G. Growth and characterization of normal human urothelium in vitro. In Vitro Cellular & Developmental Biology - Plant. 1983;19(4):326-43.

29.   Sen A, Lea-Currie YR, Sujkowska D, Franklin DM, Wilkison WO, Halvorsen YDC, et al. Adipogenic potential of human adipose derived stromal cells from multiple donors is heterogeneous. Journal of cellular biochemistry. 2001;81(2):312-9.

30.   Chapman EJ, Kelly G, Knowles MA. Genes involved in differentiation, stem cell renewal, and tumorigenesis are modulated in telomerase-immortalized human urothelial cells. Molecular Cancer Research. 2008;6(7):1154.

31.   Georgopoulos NT, Kirkwood LA, Varley CL, MacLaine NJ, Aziz N, Southgate J. Immortalisation of normal human urothelial cells compromises differentiation capacity. European urology. 2011;60(1):141-9.

32.   Ferrai C, Xie SQ, Luraghi P, Munari D, Ramirez F, Branco MR, et al. Poised transcription factories prime silent uPA gene prior to activation. PLoS biology. 2010;8(1):e1000270.

33.   Schoenfelder S, Sexton T, Chakalova L, Cope N, Horton A, Andrews S, et al. Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. Nature genetics. 2009.

34.   Varley C, Bacon E, Holder J, Southgate J. FOXA1 and IRF-1 intermediary transcriptional regulators of PPAR -induced urothelial cytodifferentiation. Cell Death & Differentiation. 2008;16(1):103-14.

35.   Yu Z, Mannik J, Soto A, Lin KK, Andersen B. The epidermal differentiation-associated Grainyhead gene Get1/Grhl3 also regulates urothelial differentiation. The EMBO Journal. 2009;28(13):1890-903.

36.   Bell SM, Zhang L, Mendell A, Xu Y, Haitchi HM, Lessard JL, et al. Kruppel-like factor 5 is required for formation and differentiation of the bladder urothelium. Developmental biology. 2011;358(1):79-90.

37.   Jain S, Pulikuri S, Zhu Y, Qi C, Kanwar YS, Yeldandi AV, et al. Differential expression of the peroxisome proliferator-activated receptor γ (PPARγ) and its coactivators steroid receptor coactivator-1 and PPAR-binding protein PBP in the brown fat, urinary bladder, colon, and breast of the mouse. The American journal of pathology. 1998;153(2):349-54.

38.   Gerstein MB, Kundaje A, Hariharan M, Landt SG, Yan K-K, Cheng C, et al. Architecture of the human regulatory network derived from ENCODE data. Nature. 2012;489(7414):91-100.

39. Hamza M, Pott S, Vega V, Thomsen J, Kandhadayar G, Ng P, et al. De-novo identification of PPAR /RXR binding sites and direct targets during adipogenesis. PLOS one. 2009;4(3).

40. Song L, Zhang Z, Grasfeder LL, Boyle AP, Giresi PG, Lee B-K, et al. Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity. Genome research. 2011;21(10):1757-67.

41. Messier B, Leblond C. Cell proliferation and migration as revealed by radioautography after injection of thymidine-H3 into male rats and mice. American Journal of Anatomy. 1960;106(3):247-85.

42. Locher G, Cooper E. Repair of rat urinary bladder epithelium following injury by cyclophosphamide. Investigative urology. 1970;8(1):116-23.

43. Knowles MA, Finesilver A, Harvey AE, Berry RJ, Hicks RM. Long-term organ culture of normal human bladder. Cancer Research. 1983;43(1):374-85.

44. Reedy E, Heatfield B. Histomorphometry and cell kinetics of normal human bladder mucosa in vitro. Urological research. 1987;15(6):321-7.

45. Walker BE. Radioautographic observations on regeneration of transitional epithelium. Tex Rep Biol Med. 1959;17(375):1C.

46. Strand DW, DeGraff DJ, Jiang M, Sameni M, Franco OE, Love HD, et al. Deficiency in Metabolic Regulators PPARγ and PTEN Cooperates to Drive Keratinizing Squamous Metaplasia in Novel Models of Human Tissue Regeneration. The American Journal of Pathology. 2012.

47. Liebert M, Wedemeyer G, Chang J, Stein J, McKeever P, Carey T, et al. Comparison of antigen expression on normal urothelial cells in tissue section and tissue culture. The Journal of urology. 1990;144(5):1288-92.

48. Varley C, Hill G, Pellegrin S, Shaw N, Selby P, Trejdosiewicz L, et al. Autocrine regulation of human urothelial cell proliferation and migration during regenerative responses in vitro. Experimental cell research. 2005;306(1):216-29.

49. Southgate J, Harnden P, Trejdosiewicz L. Cytokeratin expression patterns in normal and malignant urothelium: a review of the biological and diagnostic implications. Histology and histopathology. 1999;14(2):657.

50. Harnden P, Southgate J. Cytokeratin 14 as a marker of squamous differentiation in transitional cell carcinomas. Journal of clinical pathology. 1997;50(12):1032.

51. Cattan V, Bernard G, Rousseau A, Bouhout S, Chabaud S, Auger FA, et al. Mechanical stimuli-induced urothelial differentiation in a human tissue-engineered tubular genitourinary graft. European urology. 2011;60(6):1291-8.

52. Liang F-X, Bosland MC, Huang H, Baptiste S, Deng F-M, Wu X-R, et al. Cellular basis of urothelial squamous metaplasia roles of lineage heterogeneity and cell replacement. The Journal of cell biology. 2005;171(5):835-44.

53. Sporn MB, Clamon GH, Smith JM, Dunlop NM, Newton DL, Saffiotti U. The reversal of keratinized squamous metaplastic lesions of vitamin A deficiency in tracheobronchial epithelium by vitamin A and vitamin A analogs in organ culture: a model system for anti-carcinogenesis studies. Experimental Lung Cancer: Springer; 1974. p. 575-82.

54. Heyman RA, Mangelsdorf DJ, Dyck JA, Stein RB, Eichele G, Evans RM, et al. 9-cis retinoic acid is a high affinity ligand for the retinoid X receptor. Cell. 1992;68(2):397-406.

55. Petkovich M, Brand NJ, Krust A, Chambon P. A human retinoic acid receptor which belongs to the family of nuclear receptors. Nature. 1987;330(6147):444-50.

56. Varley CL, Stahlschmidt J, Smith B, Stower M, Southgate J. Activation of peroxisome proliferator-activated receptor-γ reverses squamous metaplasia and induces transitional differentiation in normal human urothelial cells. The American Journal of Pathology. 2004;164(5):1789-98.

57. Tontonoz P, Graves RA, Budavari AI, Erdjument-Bromage H, Lui M, Hu E, et al. Adipocyte-specific transcription factor ARF6 is a heterodimeric complex of two nuclear hormone receptors, PPAR7 and RXRa. Nucleic Acids Research. 1994;22(25):5628-34.

58.   Camp HS, Tafuri SR. Regulation of peroxisome proliferator-activated receptor γ activity by mitogen-activated protein kinase. Journal of Biological Chemistry. 1997;272(16):10811-6.

59.   Adams M, Reginato MJ, Shao D, Lazar MA, Chatterjee VK. Transcriptional activation by peroxisome proliferator-activated receptor γ is inhibited by phosphorylation at a consensus mitogen-activated protein kinase site. Journal of Biological Chemistry. 1997;272(8):5128-32.

60.   Hu E, Kim JB, Sarraf P, Spiegelman BM. Inhibition of adipogenesis through MAP kinase-mediated phosphorylation of PPARgamma. Science (New York, NY). 1996;274(5295):2100.

61.   Deng FM, Liang FX, Tu L, Resing KA, Hu P, Supino M, et al. Uroplakin IIIb, a urothelial differentiation marker, dimerizes with uroplakin Ib as an early step of urothelial plaque assembly. The Journal of cell biology. 2002;159(4):685.

62.   Stenoien D, Sharp ZD, Smith CL, Mancini MA. Functional subnuclear partitioning of transcription factors. Journal of cellular biochemistry. 1998;70(2):213-21.

63.   Munkley J, Copeland NA, Moignard V, Knight JRP, Greaves E, Ramsbottom SA, et al. Cyclin E is recruited to the nuclear matrix during differentiation, but is not recruited in cancer cells. Nucleic Acids Research. 2011;39(7):2671.

64.   Ainscough JF-X, Rahman FA, Sercombe H, Sedo A, Gerlach B, Coverley D. C-terminal domains deliver the DNA replication factor Ciz1 to the nuclear matrix. Journal of cell science. 2007;120(1):115-24.

65.   Elbashir SM, Harborth J, Lendeckel W, Yalcin A, Weber K, Tuschl T. Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. Nature. 2001;411(6836):494-8.

66.   Persengiev SP, Zhu X, Green MR. Nonspecific, concentration-dependent stimulation and repression of mammalian gene expression by small interfering RNAs (siRNAs). Rna. 2004;10(1):12-8.

67.   Sakurai H, Suzuki S, Kawasaki N, Nakano H, Okazaki T, Chino A, et al. Tumor necrosis factor-α-induced IKK phosphorylation of NF-κB p65 on serine 536 is mediated through the TRAF2, TRAF5, and TAK1 signaling pathway. Journal of Biological Chemistry. 2003;278(38):36916-23.

68.   Schroeder A, Mueller O, Stocker S, Salowsky R, Leiber M, Gassmann M, et al. The RIN: an RNA integrity number for assigning integrity values to RNA measurements. BMC molecular biology. 2006;7(1):3.

69.   Dekker J, Rippe K, Dekker M, Kleckner N. Capturing chromosome conformation. Science. 2002;295(5558):1306.

70.   Belton J-M, McCord RP, Gibcus JH, Naumova N, Zhan Y, Dekker J. Hi–C: A comprehensive technique to capture the conformation of genomes. Methods. 2012;58(3):268-76.

71.   Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science. 2009;326(5950):289.

72.   Barber RD, Harmer DW, Coleman RA, Clark BJ. GAPDH as a housekeeping gene: analysis of GAPDH mRNA expression in a panel of 72 human tissues. Physiological genomics. 2005;21(3):389-95.

73.   Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome biology. 2013;14(4):R36.

74.   Rashid NU, Giresi PG, Ibrahim JG, Sun W, Lieb JD. ZINBA integrates local covariates with DNA-seq data to identify broad and narrow regions of enrichment, even within amplified genomic regions. Genome Biol. 2011;12(7):R67.

75.   Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nature methods. 2012;9(4):357-9.

76.   Feng J, Liu T, Qin B, Zhang Y, Liu XS. Identifying ChIP-seq enrichment using MACS. nature protocols. 2012;7(9):1728-40.

77.   Feng J, Liu T, Zhang Y. Using MACS to Identify Peaks from ChIP-Seq Data. Current Protocols in Bioinformatics. 2011:2.14. 1-2.. .

78.   Wisniewski JR, Zougman A, Nagaraj N, Mann M. Universal sample preparation method for proteome analysis. Nat Meth. 2009;6(5):359-62.

79.   Podwojski K, Fritsch A, Chamrad DC, Paul W, Sitek B, Stühler K, et al. Retention time alignment algorithms for LC/MS data must consider non-linear shifts. Bioinformatics. 2009;25(6):758-64.

80.   Fajas L, Auboeuf D, Raspé E, Schoonjans K, Lefebvre A-M, Saladin R, et al. The organization, promoter analysis, and expression of the human PPARγ gene. Journal of Biological Chemistry. 1997;272(30):18779-89.

81.   Fajas L, Fruchart J-C, Auwerx J. PPARgamma3 mRNA: a distinct PPARgamma mRNA subtype transcribed from an independent promoter. FEBS letters. 1998;438(1-2):55.

82.   Knouff C, Auwerx J. Peroxisome proliferator-activated receptor-γ calls for activation in moderation: lessons from genetics and pharmacology. Endocrine reviews. 2004;25(6):899-918.

83.   Chen Y, Jimenez AR, Medh JD. Identification and regulation of novel PPAR-γ splice variants in human THP-1 macrophages. Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression. 2006;1759(1):32-43.

84.   Park SJ, Jung S-H, Jogeswar G, Ryoo H-M, Yook JI, Choi HS, et al. The transcription factor snail regulates osteogenic differentiation by repressing Runx2 expression. Bone. 2010;46(6):1498-507.

85.   Kallenberger BC, Love JD, Chatterjee VKK, Schwabe JW. A dynamic mechanism of nuclear receptor activation and its perturbation in a human disease. Nature Structural & Molecular Biology. 2003;10(2):136-40.

86.   Bugge A, Grontved L, Aagaard M, Borup R, Mandrup S. The PPAR {gamma} 2 A/B-domain plays a gene-specific role in transactivation and cofactor recruitment. Molecular Endocrinology. 2009;23(6):794.

87.   Chandra V, Huang P, Hamuro Y, Raghuram S, Wang Y, Burris TP, et al. Structure of the intact PPAR-&ggr;–RXR-&agr; nuclear receptor complex on DNA. Nature. 2008;456(7220):350-6.

88.   Tontonoz P, Hu E, Graves RA, Budavari AI, Spiegelman BM. mPPAR gamma 2: tissue-specific regulator of an adipocyte enhancer. Genes & development. 1994;8(10):1224-34.

89.   Harris P, Kletzien RF. Localization of a pioglitazone response element in the adipocyte fatty acid-binding protein gene. Molecular pharmacology. 1994;45(3):439-45.

90.   Lehmann JM, Moore LB, Smith-Oliver TA, Wilkison WO, Willson TM, Kliewer SA. An antidiabetic thiazolidinedione is a high affinity ligand for peroxisome proliferator-activated receptor γ (PPARγ). Journal of Biological Chemistry. 1995;270(22):12953-6.

91.   Takada I, Kouzmenko AP, Kato S. Wnt and PPARγ signaling in osteoblastogenesis and adipogenesis. Nature Reviews Rheumatology. 2009;5(8):442-7.

92.   Forman B, Tontonoz P, Chen J, Brun R, Spiegelman B, Evans R. 15-Deoxy-[Delta] 12, 14-Prostaglandin J2 is a ligand for the adipocyte determination factor PPAR [gamma]. Cell. 1995;83(5):803-12.

93.   Waki H, Nakamura M, Yamauchi T, Wakabayashi K, Yu J, Hirose-Yotsuya L, et al. Global mapping of cell type–specific open chromatin by FAIRE-seq reveals the regulatory role of the NFI family in adipocyte differentiation. PLoS genetics. 2011;7(10):e1002311.

94.   Siersbæk R, Nielsen R, Mandrup S. PPAR [gamma] in Adipocyte Differentiation and Metabolism-Novel Insights from Genome-wide Studies. FEBS letters. 2010.

95.   Varley C, Southgate J. Effects of PPAR agonists on proliferation and differentiation in human urothelium. Experimental and Toxicologic Pathology. 2008;60(6):435-41.

96.   Sugii S, Evans RM. Epigenetic codes of PPARγ in metabolic disease. FEBS letters. 2011;585(13):2121-8.

97. Yu S, Reddy JK. Transcription coactivators for peroxisome proliferator-activated receptors. Biochimica et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids. 2007;1771(8):936-51.

98. Viswakarma N, Jia Y, Bai L, Vluggens A, Borensztajn J, Xu J, et al. Coactivators in PPAR-regulated gene expression. PPAR research. 2010;2010.

99. Berrabah W, Aumercier P, Lefebvre P, Staels B. Control of nuclear receptor activities in metabolism by post-translational modifications. FEBS letters. 2011;585(11):1640-50.

100. Burgermeister E, Seger R. MAPK-Kinases as Nucleo-Cytoplasmic Shuttles for PPARγ. Cell Cycle. 2007;6(13):1539-48.

101. Wadosky KM, Willis MS. The story so far: post-translational regulation of peroxisome proliferator-activated receptors by ubiquitination and SUMOylation. American Journal of Physiology-Heart and Circulatory Physiology. 2012;302(3):H515-H26.

102. Brun R, Spiegelman B. PPAR gamma and the molecular control of adipogenesis. Journal of endocrinology. 1997;155(2):217-8.

103. IJpenberg A, Jeannin E, Wahli W, Desvergne B. Polarity and specific sequence requirements of peroxisome proliferator-activated receptor (PPAR)/retinoid X receptor heterodimer binding to DNA A functional analysis of the malic enzyme gene PPAR response element. Journal of Biological Chemistry. 1997;272(32):20108-17.

104. Siersbæk R, Nielsen R, John S, Sung M-H, Baek S, Loft A, et al. Extensive chromatin remodelling and establishment of transcription factor 'hotspots' during early adipogenesis. The EMBO Journal. 2011;30(8):1459-72.

105. Rosenfeld MG, Lunyak VV, Glass CK. Sensors and signals: a coactivator/corepressor/epigenetic code for integrating signal-dependent programs of transcriptional response. Genes & development. 2006;20(11):1405-28.

106. Perissi V, Scafoglio C, Zhang J, Ohgi KA, Rose DW, Glass CK, et al. TBL1 and TBLR1 phosphorylation on regulated gene promoters overcomes dual CtBP and NCoR/SMRT transcriptional repression checkpoints. Molecular cell. 2008;29(6):755-66.

107. Bugge A, Mandrup S. Molecular mechanisms and genome-wide aspects of PPAR subtype specific transactivation. PPAR research. 2010;2010.

108. Yuan C-X, Ito M, Fondell JD, Fu Z-Y, Roeder RG. The TRAP220 component of a thyroid hormone receptor-associated protein (TRAP) coactivator complex interacts directly with nuclear receptors in a ligand-dependent fashion. Proceedings of the National Academy of Sciences. 1998;95(14):7939-44.

109. Zhu Y, Qi C, Calandra C, Rao M, Reddy J. Cloning and identification of mouse steroid receptor coactivator-1 (mSRC-1), as a coactivator of peroxisome proliferator-activated receptor gamma. Gene expression. 1996;6(3):185.

110. Puigserver P, Wu Z, Park CW, Graves R, Wright M, Spiegelman BM. A cold-inducible coactivator of nuclear receptors linked to adaptive thermogenesis. Cell. 1998;92(6):829.

111. Castillo G, Brun RP, Rosenfield JK, Hauser S, Park CW, Troy AE, et al. An adipogenic cofactor bound by the differentiation domain of PPARγ. The EMBO Journal. 1999;18(13):3676-87.

112. Rosen ED, Hsu C-H, Wang X, Sakai S, Freeman MW, Gonzalez FJ, et al. C/EBPα induces adipogenesis through PPARγ: a unified pathway. Genes & development. 2002;16(1):22-6.

113. Cornelius P, MacDougald OA, Lane MD. Regulation of adipocyte development. Annual review of nutrition. 1994;14(1):99-129.

114. Tontonoz P, Hu E, Spiegelman BM. Stimulation of adipogenesis in fibroblasts by PPARγ2, a lipid-activated transcription factor. Cell. 1994;79(7):1147-56.

115. Berger J, Patel HV, Woods J, Hayes NS, Parent SA, Clemas J, et al. A PPARγ mutant serves as a dominant negative inhibitor of PPAR signaling and is localized in the nucleus. Molecular and cellular endocrinology. 2000;162(1):57-67.

116. Sabatino L, Casamassimi A, Peluso G, Barone MV, Capaccio D, Migliore C, et al. A novel peroxisome proliferator-activated receptor γ isoform with dominant negative

activity generated by alternative splicing. Journal of Biological Chemistry. 2005;280(28):26517-25.

117.   Agostini M, Schoenmakers E, Mitchell C, Szatmari I, Savage D, Smith A, et al. Non-DNA binding, dominant-negative, human PPARγ mutations cause lipodystrophic insulin resistance. Cell metabolism. 2006;4(4):303-11.

118.   Hummasti S, Tontonoz P. The peroxisome proliferator-activated receptor N-terminal domain controls isotype-selective gene expression and adipogenesis. Molecular endocrinology. 2006;20(6):1261-75.

119.   Pascual G, Fong AL, Ogawa S, Gamliel A, Li AC, Perissi V, et al. A SUMOylation-dependent pathway mediates transrepression of inflammatory response genes by PPAR-γ. Nature. 2005;437(7059):759-63.

120.   Cunha GR. Epithelio-mesenchymal interactions in primordial gland structures which become responsive to androgenic stimulation. The Anatomical Record. 1972;172(2):179-95.

121.   Southgate J, Varley CL, Garthwaite MAE, Hinley J, Marsh F, Stahlschmidt J, et al. Differentiation potential of urothelium from patients with benign bladder dysfunction. BJU international. 2007;99(6):1506.

122.   Fleming JM. Regulation of Growth and Differentiation in Human Urothelium: University of York; 2008.

123.   Fleming JM, Shabir S, Varley CL, Kirkwood LA, White A, Holder J, et al. Differentiation-Associated Reprogramming of the Transforming Growth Factor β Receptor Pathway Establishes the Circuitry for Epithelial Autocrine/Paracrine Repair. PLOS one. 2012;7(12):e51404.

124.   Wang Y, Hayward SW, Cao M, Thayer KA, Cunha GR. Cell differentiation lineage in the prostate. Differentiation. 2001;68(4-5):270-9.

125.   Floyd ZE, Stephens JM. Interferon-γ-mediated activation and ubiquitin-proteasome-dependent degradation of PPARγ in adipocytes. Journal of Biological Chemistry. 2002;277(6):4062-8.

126.   Lefterova MI, Steger DJ, Zhuo D, Qatanani M, Mullican SE, Tuteja G, et al. Cell-specific determinants of peroxisome proliferator-activated receptor γ function in adipocytes and macrophages. Molecular and cellular biology. 2010;30(9):2078-89.

127.   Furuyashiki T, Nagayasu H, Aoki Y, Bessho H, Hashimoto T, Kanazawa K, et al. Tea Catechin Suppresses Adipocyte Differentiation Accompanied by Down-regulation of PPARγ2 and C/EBPα in 3T3-L1 Cells. Bioscience, Biotechnology, and Biochemistry. 2004;68(11):2353-9.

128.   Gibbings S, Elkins ND, Fitzgerald H, Tiao J, Weyman ME, Shibao G, et al. Xanthine oxidoreductase promotes the inflammatory state of mononuclear phagocytes through effects on chemokine expression, peroxisome proliferator-activated receptor-γ sumoylation, and HIF-1α. Journal of Biological Chemistry. 2011;286(2):961-75.

129.   Ohshima T, Koga H, Shimotohno K. Transcriptional activity of peroxisome proliferator-activated receptor γ is modulated by SUMO-1 modification. Journal of Biological Chemistry. 2004;279(28):29551-7.

130.   Yamashita D, Yamaguchi T, Shimizu M, Nakata N, Hirose F, Osumi T. The transactivating function of peroxisome proliferator-activated receptor γ is negatively regulated by SUMO conjugation in the amino-terminal domain. Genes to Cells. 2004;9(11):1017-29.

131.   Chung S, Ahn B, Kim M, Kho J, Jung H, Park K. SUMO modification selectively regulates transcriptional activity of peroxisome-proliferator-activated receptor gamma in C2C12 myotubes. Biochem J. 2011;433:155-61.

132.   Floyd ZE, Stephens JM. Control of Peroxisome Proliferator-Activated Receptor γ2 Stability and Activity by SUMOylation. Obesity research. 2012;12(6):921-8.

133.   Chopra B, Hinley J, Oleksiewicz M, Southgate J. Trans-species comparison of PPAR and RXR expression by rat and human urothelial tissues. Toxicologic pathology. 2008;36(3):485.

134.   Baumann CT, Lim CS, Hager GL. Intracellular localization and trafficking of steroid receptors. Cell biochemistry and biophysics. 1999;31(2):119-27.

135.   Baumann CT, Maruvada P, Hager GL, Yen PM. Nuclear cytoplasmic shuttling by thyroid hormone receptors. Journal of Biological Chemistry. 2001;276(14):11237-45.

136.   Guan H-P, Ishizuka T, Chui PC, Lehrke M, Lazar MA. Corepressors selectively control the transcriptional activity of PPARγ in adipocytes. Genes & development. 2005;19(4):453-61.

137.   Feige JN, Gelman L, Tudor C, Engelborghs Y, Wahli W, Desvergne B. Fluorescence imaging reveals the nuclear behavior of peroxisome proliferator-activated receptor/retinoid X receptor heterodimers in the absence and presence of ligand. Journal of Biological Chemistry. 2005;280(18):17880-90.

138.   Akiyama TE, Baumann CT, Sakai S, Hager GL, Gonzalez FJ. Selective intranuclear redistribution of PPAR isoforms by RXRα. Molecular Endocrinology. 2002;16(4):707-21.

139.   Barrack ER, Hawkins EF, Allen SL, Hicks L, Coffey DS. Concepts related to salt resistant estradiol receptors in rat uterine nuclei: nuclear matrix. Biochemical and biophysical research communications. 1977;79(3):829.

140.   Stenoien DL, Mancini MG, Patel K, Allegretto EA, Smith CL, Mancini MA. Subnuclear trafficking of estrogen receptor-α and steroid receptor coactivator-1. Molecular Endocrinology. 2000;14(4):518-34.

141.   Debril M, Dubuquoy L, Feige J, Wahli W, Desvergne B, Auwerx J, et al. Scaffold attachment factor B1 directly interacts with nuclear receptors in living cells and represses transcriptional activity. Journal of molecular endocrinology. 2005;35(3):503-17.

142.   Su J-L, Winegar DA, Wisely GB, Sigel CS, Hull-Ryde EA. Use of a PPAR gamma-specific monoclonal antibody to demonstrate thiazolidinediones induce PPAR gamma receptor expression in vitro. Hybridoma. 1999;18(3):273-80.

143.   Fukuda I, Ito A, Hirai G, Nishimura S, Kawasaki H, Saitoh H, et al. Ginkgolic acid inhibits protein SUMOylation by blocking formation of the E1-SUMO intermediate. Chemistry & biology. 2009;16(2):133-40.

144.   Rubenwolf PC, Georgopoulos NT, Clements LA, Feather S, Holland P, Thomas DF, et al. Expression and localisation of aquaporin water channels in human urothelium in situ and in vitro. European urology. 2009;56(6):1013-24.

145.   Balasubramanyam K, Swaminathan V, Ranganathan A, Kundu TK. Small molecule modulators of histone acetyltransferase p300. Journal of Biological Chemistry. 2003;278(21):19134-40.

146.   van Steensel B, Jenster G, Damm K, Brinkmann AO, van Driel R. Domains of the human androgen receptor and glucocorticoid receptor involved in binding to the nuclear matrix. Journal of cellular biochemistry. 1995;57(3):465-78.

147.   Henikoff S, Henikoff JG, Sakai A, Loeb GB, Ahmad K. Genome-wide profiling of salt fractions maps physical properties of chromatin. Genome research. 2009;19(3):460-9.

148.   Guilherme A, Tesz GJ, Guntur KV, Czech MP. Tumor necrosis factor-α induces caspase-mediated cleavage of peroxisome proliferator-activated receptor γ in adipocytes. Journal of Biological Chemistry. 2009;284(25):17082-91.

149.   He F, Doucet JA, Stephens JM. Caspase-mediated degradation of PPARγ proteins in adipocytes. Obesity. 2008;16(8):1735-41.

150.   Kim HJ, Woo IS, Kang ES, Eun SY, Kim HJ, Lee JH, et al. Identification of a truncated alternative splicing variant of human PPARγ1 that exhibits dominant negative activity. Biochemical and biophysical research communications. 2006;347(3):698-706.

151.   Choi JH, Banks AS, Estall JL, Kajimura S, Boström P, Laznik D, et al. Anti-diabetic drugs inhibit obesity-linked phosphorylation of PPAR [ggr] by Cdk5. Nature. 2010;466(7305):451-6.

152.   Rappsilber J, Ryder U, Lamond AI, Mann M. Large-scale proteomic analysis of the human spliceosome. Genome research. 2002;12(8):1231-45.

153.   Ishihama Y, Oda Y, Tabata T, Sato T, Nagasu T, Rappsilber J, et al. Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in

proteomics by the number of sequenced peptides per protein. Molecular & Cellular Proteomics. 2005;4(9):1265.

154.   Beck M, Schmidt A, Malmstroem J, Claassen M, Ori A, Szymborska A, et al. The quantitative proteome of a human cell line. Molecular systems biology. 2011;7(1).

155.   Addona TA, Abbatiello SE, Schilling B, Skates SJ, Mani D, Bunk DM, et al. Multi-site assessment of the precision and reproducibility of multiple reaction monitoring–based measurements of proteins in plasma. Nature biotechnology. 2009;27(7):633-41.

156.   Lu P, Vogel C, Wang R, Yao X, Marcotte EM. Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. Nature biotechnology. 2006;25(1):117-24.

157.   Bantscheff M, Schirle M, Sweetman G, Rick J, Kuster B. Quantitative mass spectrometry in proteomics: a critical review. Analytical and Bioanalytical Chemistry. 2007;389(4):1017-31.

158.   Neilson KA, Ali NA, Muralidharan S, Mirzaei M, Mariani M, Assadourian G, et al. Less label, more free: Approaches in label-free quantitative mass spectrometry. Proteomics. 2011;11(4):535-53.

159.   Takata H, Nishijima H, Ogura Si, Sakaguchi T, Bubulya PA, Mochizuki T, et al. Proteome analysis of human nuclear insoluble fractions. Genes to Cells. 2009;14(8):975-90.

160.   Dutta B, Adav SS, Koh CG, Lim SK, Meshorer E, Sze SK. Elucidating the Temporal Dynamics of Chromatin-Associated Protein Release Upon DNA Digestion by Quantitative Proteomics Approach. Journal of Proteomics. 2012.

161.   Mosley AL, Florens L, Wen Z, Washburn MP. A label free quantitative proteomic analysis of the Saccharomyces cerevisiae nucleus. Journal of Proteomics. 2009;72(1):110-20.

162.   Zhu Z, Boobis AR, Edwards RJ. Identification of estrogen-responsive proteins in MCF-7 human breast cancer cells using label-free quantitative proteomics. PROTEOMICS. 2008;8(10):1987-2005.

163.   Vissers J, Langridge J, Aerts J. Analysis and quantification of diagnostic serum markers and protein signatures for Gaucher disease. Molecular & Cellular Proteomics. 2007;6(5):755.

164.   Huang J, McKenna T, Hughes C, Leweke F, Schwarz E, Bahn S. CSF biomarker discovery using label free nano LC MS based proteomic profiling: Technical aspects. Journal of separation science. 2007;30(2):214-25.

165.   Xanthoudakis S, Miao G, Wang F, Pan Y, Curran T. Redox activation of Fos-Jun DNA binding activity is mediated by a DNA repair enzyme. The EMBO Journal. 1992;11(9):3323.

166.   Huarte M, Guttman M, Feldser D, Garber M, Koziol MJ, Kenzelmann-Broz D, et al. A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. Cell. 2010;142(3):409-19.

167.   Yang Z, Bowles NE, Scherer SE, Taylor MD, Kearney DL, Ge S, et al. Desmosomal dysfunction due to mutations in desmoplakin causes arrhythmogenic right ventricular dysplasia/cardiomyopathy. Circulation research. 2006;99(6):646-55.

168.   Takimoto M, Tomonaga T, Matunis M, Avigan M, Krutzsch H, Dreyfuss G, et al. Specific binding of heterogeneous ribonucleoprotein particle protein K to the human c-myc promoter, in vitro. Journal of Biological Chemistry. 1993;268(24):18249-58.

169.   Angelov D, Molla A, Perche P-Y, Hans F, Côté J, Khochbin S, et al. The histone variant macroH2A interferes with transcription factor binding and SWI/SNF nucleosome remodeling. Molecular cell. 2003;11(4):1033-41.

170.   Cottrell J, London U. Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis. 1999;20(18):3551-67.

171.   Savitski MM, Lemeer S, Boesche M, Lang M, Mathieson T, Bantscheff M, et al. Confident phosphorylation site localization using the Mascot Delta Score. Molecular & Cellular Proteomics. 2011;10(2).

172.   Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized ppb-range mass accuracies and proteome-wide protein quantification. Nature biotechnology. 2008;26(12):1367-72.

173.   MacCoss MJ, Wu CC, Yates JR. Probability-based validation of protein identifications using a modified SEQUEST algorithm. Analytical chemistry. 2002;74(21):5593-9.

174.   Steen H, Mann M. The ABC's (and XYZ's) of peptide sequencing. Nature Reviews Molecular Cell Biology. 2004;5(9):699-711.

175.   The_Global_Proteome_Machine. http://gpmdb.thegpm.org/index.html.

176.   Soleimani VD, Yin H, Jahani-Asl A, Ming H, Kockx CE, van Ijcken WF, et al. Snail regulates MyoD binding-site occupancy to direct enhancer switching and differentiation-specific transcription in myogenesis. Molecular cell. 2012.

177.   Hindermann W, Berndt A, Haas KM, Wunderlich H, Katenkamp D, Kosmehl H. Immunohistochemical demonstration of the γ2 chain of laminin-5 in urinary bladder urothelial carcinoma: Impact for diagnosis and prognosis. Cancer detection and prevention. 2003;27(2):109-15.

178.   Harburger DS, Calderwood DA. Integrin signalling at a glance. Journal of cell science. 2009;122(2):159-63.

179.   Hozák P, Sasseville A, Raymond Y, Cook PR. Lamin proteins form an internal nucleoskeleton as well as a peripheral lamina in human cells. Journal of cell science. 1995;108(2):635-44.

180.   Dechat T, Pfleghaar K, Sengupta K, Shimi T, Shumaker DK, Solimando L, et al. Nuclear lamins: major factors in the structural organization and function of the nucleus and chromatin. Genes & development. 2008;22(7):832-53.

181.   Ockleford C, Bright N, Hubbard A, D'Lacey C, Smith J, Gardiner L, et al. Micro-Trabeculae, Macro-Plaques or Mini-Basement Membranes in Human Term Fetal Membranes? Philosophical Transactions of the Royal Society of London Series B: Biological Sciences. 1993;342(1300):121-36.

182.   Clarke C, Henry M, Doolan P, Kelly S, Aherne S, Sanchez N, et al. Integrated miRNA, mRNA and protein expression analysis reveals the role of post-transcriptional regulation in controlling CHO cell growth rate. BMC genomics. 2012;13(1):656.

183.   Landthaler M, Gaidatzis D, Rothballer A, Chen PY, Soll SJ, Dinic L, et al. Molecular characterization of human Argonaute-containing ribonucleoprotein complexes and their bound target mRNAs. Rna. 2008;14(12):2580-96.

184.   Warner DR, Bhattacherjee V, Yin X, Singh S, Mukhopadhyay P, Pisano MM, et al. Functional interaction between Smad, CREB binding protein, and p68 RNA helicase. Biochemical and Biophysical Research Communications. 2004;324(1):70-6.

185.   Jensen ED, Niu L, Caretti G, Nicol SM, Teplyuk N, Stein GS, et al. p68 (Ddx5) interacts with Runx2 and regulates osteoblast differentiation. Journal of cellular biochemistry. 2008;103(5):1438-51.

186.   Wang Q, Li W, Liu X, Carroll J, Jänne O, Keeton E, et al. A hierarchical network of transcription factors governs androgen receptor-dependent prostate cancer growth. Molecular cell. 2007;27(3):380-92.

187.   Ross-Innes C, Stark R, Holmes K, Schmidt D, Spyrou C, Russell R, et al. Cooperative interaction between retinoic acid receptor- and estrogen receptor in breast cancer. Genes & development. 2010;24(2):171.

188.   Lupien M, Eeckhoute J, Meyer C, Wang Q, Zhang Y, Li W, et al. FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. Cell. 2008;132(6):958-70.

189.   Carroll J, Meyer C, Song J, Li W, Geistlinger T, Eeckhoute J, et al. Genome-wide analysis of estrogen receptor binding sites. Nature genetics. 2006;38(11):1289-97.

190.   Arvey A, Agius P, Noble WS, Leslie C. Sequence and chromatin determinants of cell-type–specific transcription factor binding. Genome research. 2012;22(9):1723-34.

191.   Charos AE, Reed BD, Raha D, Szekely AM, Weissman SM, Snyder M. A highly integrated and complex PPARGC1A transcription factor binding network in HepG2 cells. Genome research. 2012;22(9):1668-79.

192.   Wang J, Zhuang J, Iyer S, Lin X, Whitfield TW, Greven MC, et al. Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. Genome research. 2012;22(9):1798-812.

193.   Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, et al. The accessible chromatin landscape of the human genome. Nature. 2012;489(7414):75-82.

194.   Wu C. The 5'ends of Drosophila heat shock genes in chromatin are hypersensitive to DNase I. Nature. 1980;286(5776):854.

195.   Giresi PG, Kim J, McDaniell RM, Iyer VR, Lieb JD. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. Genome research. 2007;17(6):877-85.

196.   Giresi PG, Lieb JD. Isolation of active regulatory elements from eukaryotic chromatin using FAIRE (Formaldehyde Assisted Isolation of Regulatory Elements). Methods. 2009;48(3):233-9.

197.   Boyle AP, Davis S, Shulha HP, Meltzer P, Margulies EH, Weng Z, et al. High-resolution mapping and characterization of open chromatin across the genome. Cell. 2008;132(2):311-22.

198.   Crawford GE, Davis S, Scacheri PC, Renaud G, Halawi MJ, Erdos MR, et al. DNase-chip: a high-resolution method to identify DNase I hypersensitive sites using tiled microarrays. Nature Methods. 2006;3(7):503-9.

199.   Giannopoulou EG, Elemento O. Inferring chromatin-bound protein complexes from genome-wide binding assays. Genome research. 2013.

200.   Fraser P, Bickmore W. Nuclear organization of the genome and the potential for gene regulation. Nature. 2007;447(7143):413-7.

201.   Lanctôt C, Cheutin T, Cremer M, Cavalli G, Cremer T. Dynamic genome architecture in the nuclear space: regulation of gene expression in three dimensions. Nature Reviews Genetics. 2007;8(2):104-15.

202.   Misteli T. Beyond the sequence: cellular organization of genome function. Cell. 2007;128(4):787.

203.   Green EM, Jiang Y, Joyner R, Weis K. A negative feedback loop at the nuclear periphery regulates GAL gene expression. Molecular biology of the cell. 2012;23(7):1367-75.

204.   Li G, Fullwood M, Xu H, Mulawadi F, Velkov S, Vega V, et al. ChIA-PET tool for comprehensive chromatin interaction analysis with paired-end tag sequencing. Genome Biology. 2010;11(2):R22.

205.   Fullwood M, Liu M, Pan Y, Liu J, Xu H, Mohamed Y, et al. An oestrogen-receptor-&agr;-bound human chromatin interactome. Nature. 2009;462(7269):58-64.

206.   Eskiw CH, Fraser P. Ultrastructural study of transcription factories in mouse erythroblasts. Journal of cell science. 2011;124(21):3676-83.

207.   Rickman DS, Soong TD, Moss B, Mosquera JM, Dlabal J, Terry S, et al. Oncogene-mediated alterations in chromatin conformation. Proceedings of the National Academy of Sciences. 2012;109(23):9083-8.

208.   Yu J, Yu J, Mani R-S, Cao Q, Brenner CJ, Cao X, et al. An integrated network of androgen receptor, polycomb, and TMPRSS2-ERG gene fusions in prostate cancer progression. Cancer cell. 2010;17(5):443-54.

209.   Lin YC, Benner C, Mansson R, Heinz S, Miyazaki K, Miyazaki M, et al. Global changes in the nuclear positioning of genes and intra-and interdomain genomic interactions that orchestrate B cell fate. Nature immunology. 2012.

210.   Nishiyama N, Arai E, Chihara Y, Fujimoto H, Hosoda F, Shibata T, et al. Genome-wide DNA methylation profiles in urothelial carcinomas and urothelia at the precancerous stage. 2009.

211.   Vallot C, Stransky N, Bernard-Pierrot I, Hérault A, Zucman-Rossi J, Chapeaublanc E, et al. A Novel Epigenetic Phenotype Associated With the Most Aggressive Pathway of Bladder Tumor Progression. Journal of the National Cancer Institute. 2011;103(1):47.

212.   Frigola J, Song J, Stirzaker C, Hinshelwood RA, Peinado MA, Clark SJ. Epigenetic remodeling in colorectal cancer results in coordinate gene suppression across an entire chromosome band. Nature genetics. 2006;38(5):540-9.

213.   Stransky N, Vallot C, Reyal F, Bernard-Pierrot I, de Medina SGD, Segraves R, et al. Regional copy number–independent deregulation of transcription in cancer. Nature.200:6.

214.   Dyrskjot L, Kruhoffer M, Thykjaer T, Marcussen N, Jensen JL, Moller K, et al. Gene expression in the urinary bladder: a common carcinoma in situ gene expression signature exists disregarding histopathological classification. Cancer research. 2004;64(11):4040.

215.   Maruyama R, Toyooka S, Toyooka KO, Harada K, Virmani AK, Zöchbauer-Müller S, et al. Aberrant promoter methylation profile of bladder cancer and its relationship to clinicopathological features. Cancer research. 2001;61(24):8659-63.

216.   Catto JW, Azzouzi A-R, Rehman I, Feeley KM, Cross SS, Amira N, et al. Promoter hypermethylation is associated with tumor location, stage, and subsequent progression in transitional cell carcinoma. Journal of clinical oncology. 2005;23(13):2903-10.

217.   Nakagawa T, Kanai Y, Ushijima S, Kitamura T, Kakizoe T, Hirohashi S. DNA hypermethylation on multiple CpG islands associated with increased DNA methyltransferase DNMT1 protein expression during multistage urothelial carcinogenesis. The Journal of urology. 2005;173(5):1767.

218.   Dhawan D, Hamdy F, Rehman I, Patterson J, Cross S, Feeley K, et al. Evidence for the early onset of aberrant promoter methylation in urothelial carcinoma. The Journal of Pathology. 2006;209(3):336-43.

219.   Dudziec E, Gogol-Döring A, Cookson V, Chen W, Catto J. Integrated epigenome profiling of repressive histone modifications, DNA methylation and gene expression in normal and malignant urothelial cells. PLOS one. 2012;7(3):e32750.

220.   Liu CL, Schreiber SL, Bernstein BE. Development and validation of a T7 based linear amplification for genomic DNA. BMC genomics. 2003;4(1):19.

221.   Adli M, Zhu J, Bernstein BE. Genome-wide chromatin maps derived from limited numbers of hematopoietic progenitors. Nature Methods. 2010;7(8):615-8.

222.   Shankaranarayanan P, Mendoza-Parra M-A, Walia M, Wang L, Li N, Trindade LM, et al. Single-tube linear DNA amplification (LinDA) for robust ChIP-seq. Nature Methods. 2011;8(7):565-7.

223.   Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol. 2010;11(10):R106.

224.   Zhou VW, Goren A, Bernstein BE. Charting histone modifications and the functional organization of mammalian genomes. Nature Reviews Genetics. 2010;12(1):7-18.

225.   Kidder BL, Hu G, Zhao K. ChIP-Seq: technical considerations for obtaining high-quality data. Nature immunology. 2011;12(10):918-22.

226.   Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 2008;9(9):R137.

227.   Giannopoulou E, Elemento O. An integrated ChIP-seq analysis platform with customizable workflows. BMC bioinformatics. 2011;12(1):277.

228.   Simon JM, Giresi PG, Davis IJ, Lieb JD. Using formaldehyde-assisted isolation of regulatory elements (FAIRE) to isolate active regulatory DNA. Nature Protocols. 2012;7(2):256-67.

229.   van Berkum NL, Lieberman-Aiden E, Williams L, Imakaev M, Gnirke A, Mirny LA, et al. Hi-C: A Method to Study the Three-dimensional Architecture of Genomes. Journal of Visualized Experiments. 2010(39).

230.   Southgate J, Masters JR, Trejdosiewicz LK. Culture of human urothelium. Culture of epithelial cells. 2002:381-400.

231.   Blouin K, Veilleux A, Tchernof A. Androgen metabolism in adipose tissue: Recent advances. Molecular and cellular endocrinology. 2009;301(1):97-103.

232.   Wrann CD, Eguchi J, Bozec A, Xu Z, Mikkelsen T, Gimble J, et al. FOSL2 promotes leptin gene expression in human and mouse adipocytes. The Journal of clinical investigation. 2012;122(3):1010.

233.   Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B (Methodological). 1995:289-300.

234.   Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nature biotechnology. 2010;28(5):511-5.

235.   Boada LD, Zumbado M, del Río I, Blanco A, Torres S, Monterde JG, et al. Steroid hormone progesterone induces cell proliferation and abnormal mitotic processes in rat liver. Archives of toxicology. 2002;75(11-12):707-16.

236.   Sirigu P, Perra M, Turno F, Usai E. Immunohistochemical demonstration of secretory IgA in human urothelium. 1995.

237.   d'Addessi A, Racioppi M, Fanasca A, Menchinelli P, Alcini A, Delicato G, et al. Long-term behavior of secretory immunity in ileocecal and ileal orthotopic neobladders. Urologia internationalis. 2001;67(1):41-5.

238.   Asano M, Komiyama K. Polymeric immunoglobulin receptor. Journal of oral science. 2011;53(2):147-56.

239.   Rossel M, Billerey C, Bittard H, Ksiazek P, Alber D, Revillard J-P, et al. Alterations in polymeric immunoglobulin receptor expression and secretory component levels in bladder carcinoma. Urological research. 1991;19(6):361-6.

240.   Vadigepalli R, Chakravarthula P, Zak DE, Schwaber JS, Gonye GE. PAINT: a promoter analysis and interaction network generation tool for gene regulatory network identification. Omics A Journal of Integrative Biology. 2003;7(3):235-52.

241.   Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature. 2012;485(7398):376-80.

242.   Bernstein B, Mikkelsen T, Xie X, Kamal M, Huebert D, Cuff J, et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell. 2006;125(2):315-26.

243.   Nora EP, Dekker J, Heard E. Segmental folding of chromosomes: A basis for structural and regulatory chromosomal neighborhoods? BioEssays. 2013.

244.   Harney AS, Meade TJ, LaBonne C. Targeted Inactivation of Snail Family EMT Regulatory Factors by a Co(III)-Ebox Conjugate. PLOS one. 2012;7(2):e32318.

245.   Carpenter B, Hill KJ, Charalambous M, Wagner KJ, Lahiri D, James DI, et al. BASP1 is a transcriptional cosuppressor for the Wilms' tumor suppressor protein WT1. Molecular and cellular biology. 2004;24(2):537-49.

246.   Chattopadhyay C, Hawke D, Kobayashi R, Maity SN. Human p32, interacts with B subunit of the CCAAT-binding factor, CBF/NF-Y, and inhibits CBF-mediated transcription activation in vitro. Nucleic Acids Research. 2004;32(12):3632-41.

247.   Coles LS, Diamond P, Occhiodoro F, Vadas MA, Shannon MF. Cold shock domain proteins repress transcription from the GM-CSF promoter. Nucleic Acids Research. 1996;24(12):2311-7.

248.   Tutter AV, Fryer CJ, Jones KA. Chromatin-specific regulation of LEF-1–β-catenin transcription activation and inhibition in vitro. Genes & development. 2001;15(24):3342-54.

249.   Ishaq M, Ma L, Wu X, Mu Y, Pan Ja, Hu J, et al. The DEAD-box RNA helicase DDX1 interacts with RelA and enhances nuclear factor kappaB-mediated transcription. Journal of cellular biochemistry. 2008;106(2):296-305.

250.   Clark EL, Coulson A, Dalgliesh C, Rajan P, Nicol SM, Fleming S, et al. The RNA helicase p68 is a novel androgen receptor coactivator involved in splicing and is overexpressed in prostate cancer. Cancer research. 2008;68(19):7938-46.

251.   Ray R, Miller DM. Cloning and characterization of a human c-myc promoter-binding protein. Molecular and cellular biology. 1991;11(4):2154-61.

252.   Maeda O, Usami N, Kondo M, Takahashi M, Goto H, Shimokata K, et al. Plakoglobin (γ-catenin) has TCF/LEF family-dependent transcriptional activity in β-catenin-deficient cell line. Oncogene. 2003;23(4):964-72.

253.   Emdad L, Sarkar D, Su Z-z, Randolph A, Boukerche H, Valerie K, et al. Activation of the nuclear factor κB pathway by astrocyte elevated gene-1: implications for tumor progression and metastasis. Cancer research. 2006;66(3):1509-16.

254.   Al-Shanti N, Steward C, Garland R, Rowbottom A. Investigation of alpha nascent polypeptide-associated complex functions in a human CD8+ T cell ex vivo expansion model using antisense oligonucleotides. Immunology. 2004;112(3):397-403.

255.   Dhar SK, Lynn BC, Daosukho C, Clair DKS. Identification of nucleophosmin as an NF-κB co-activator for the induction of the human SOD2 gene. Journal of Biological Chemistry. 2004;279(27):28209-19.

256.   Abramovich C, Shen W-F, Pineault N, Imren S, Montpetit B, Largman C, et al. Functional cloning and characterization of a novel nonhomeodomain protein that inhibits the binding of PBX1-HOX complexes to DNA. Journal of Biological Chemistry. 2000;275(34):26172-7.

257.   Manavathi B, Acconcia F, Rayala SK, Kumar R. An inherent role of microtubule network in the action of nuclear receptor. Proceedings of the National Academy of Sciences. 2006;103(43):15981-6.

258.   Ito J, Hashimoto T, Nakamura S, Aita Y, Yamazaki T, Schlegel W, et al. Splicing transitions of the anchoring protein ENH during striated muscle development. Biochemical and biophysical research communications. 2012.

259.   Lin D-Y, Fang H-I, Ma A-H, Huang Y-S, Pu Y-S, Jenster G, et al. Negative modulation of androgen receptor transcriptional activity by Daxx. Molecular and cellular biology. 2004;24(24):10529-41.

260.   Laird KM, Briggs LL, Boss JM, Summers MF, Garvie CW. Solution structure of the heterotrimeric complex between the interaction domains of RFX5 and RFXAP from the RFX gene regulatory complex. Journal of molecular biology. 2010;403(1):40-51.

261.   Yang J, Aittomäki S, Pesu M, Carter K, Saarinen J, Kalkkinen N, et al. Identification of p100 as a coactivator for STAT6 that bridges STAT6 with RNA polymerase II. The EMBO Journal. 2002;21(18):4950-8.

262.   Leverson JD, Koskinen PJ, Orrico FC, Rainio E-M, Jalkanen KJ, Dash AB, et al. Pim-1 kinase and p100 cooperate to enhance c-Myb activity. Molecular cell. 1998;2(4):417-25.

263.   Tong X, Drapkin R, Yalamanchili R, Mosialos G, Kieff E. The Epstein-Barr virus nuclear protein 2 acidic domain forms a complex with a novel cellular coactivator that can interact with TFIIE. Molecular and cellular biology. 1995;15(9):4735-44.

264.   Moosmann P, Georgiev O, Le Douarin B, Bourquin J-P, Schaffner W. Transcriptional repression by RING finger protein TIF1β that interacts with the KRAB repressor domain of KOX1. Nucleic Acids Research. 1996;24(24):4859-67.

265.   Liu Y, Ross JF, Bodine PV, Billiard J. Homodimerization of Ror2 tyrosine kinase receptor induces 14-3-3β phosphorylation and promotes osteoblast differentiation and bone formation. Molecular Endocrinology. 2007;21(12):3050-61.

266.   Berry FB, O'Neill MA, Coca-Prados M, Walter MA. FOXC1 transcriptional regulatory activity is impaired by PBX1 in a filamin A-mediated manner. Molecular and cellular biology. 2005;25(4):1415-24.

267.   Jønson L, Vikesaa J, Krogh A, Nielsen LK, Borup R, Johnsen AH, et al. Molecular composition of IMP1 ribonucleoprotein granules. Molecular & Cellular Proteomics. 2007;6(5):798-811.

268.   Tay N, Chan S, Ren E. Identification and cloning of a novel heterogeneous nuclear ribonucleoprotein C-like protein that functions as a transcriptional activator of the hepatitis B virus enhancer II. Journal of virology. 1992;66(12):6841-8.

269. Boopathi E, Lenka N, Prabu SK, Fang J-K, Wilkinson F, Atchison M, et al. Regulation of Murine Cytochrome c Oxidase Vb Gene Expression during Myogenesis. Journal of Biological Chemistry. 2004;279(34):35242-54.

270. Kzhyshkowska J, Rusch A, Wolf H, Dobner T. Regulation of transcription by the heterogeneous nuclear ribonucleoprotein E1B-AP5 is mediated by complex formation with the novel bromodomain-containing protein BRD7. Biochemical Journal. 2003;371(Pt 2):385.

271. Yahata T, de Caestecker MP, Lechleider RJ, Andriole S, Roberts AB, Isselbacher KJ, et al. The MSG1 non-DNA-binding transactivator binds to the p300/CBP coactivators, enhancing their functional link to the Smad transcription factors. Journal of Biological Chemistry. 2000;275(12):8825-34.

272. Reichman TW, Muñiz LC, Mathews MB. The RNA binding protein nuclear factor 90 functions as both a positive and negative regulator of gene expression in mammalian cells. Molecular and cellular biology. 2002;22(1):343-56.

273. De Vos WH, Houben F, Hoebe RA, Hennekam R, Van Engelen B, Manders EM, et al. Increased plasticity of the nuclear envelope and hypermobility of telomeres due to the loss of A–type lamins. Biochimica et Biophysica Acta (BBA)-General Subjects. 2010;1800(4):448-58.

274. Coates P, Nenutil R, McGregor A, Picksley S, Crouch D, Hall P, et al. Mammalian prohibitin proteins respond to mitochondrial stress and decrease during cellular senescence. Experimental cell research. 2001;265(2):262-73.

275. Montano MM, Ekena K, Delage-Mourroux R, Chang W, Martini P, Katzenellenbogen BS. An estrogen receptor-selective coregulator that potentiates the effectiveness of antiestrogens and represses the activity of estrogens. Proceedings of the National Academy of Sciences. 1999;96(12):6947-52.

276. Suico MA, Yoshida H, Seki Y, Uchikawa T, Lu Z, Shuto T, et al. Myeloid Elf-1-like factor, an ETS transcription factor, up-regulates lysozyme transcription in epithelial cells through interaction with promyelocytic leukemia protein. Journal of Biological Chemistry. 2004;279(18):19091-8.

277. Song MS, Salmena L, Carracedo A, Egia A, Lo-Coco F, Teruya-Feldstein J, et al. The deubiquitinylation and localization of PTEN are regulated by a HAUSP–PML network. Nature. 2008;455(7214):813-7.

278. Lee JW, Ryan F, Swaffield JC, Johnston SA, Moore DD. Interaction of thyroid-hormone receptor with a conserved transcriptional mediator. 1995.

279. Ellison JW, Wardak Z, Young MF, Robey PG, Laig-Webster M, Chiong W. PHOG, a candidate gene for involvement in the short stature of Turner syndrome. Human molecular genetics. 1997;6(8):1341-7.

280. Wang J, Gao QS, Wang Y, Lafyatis R, Stamm S, Andreadis A. Tau exon 10, whose missplicing causes frontotemporal dementia, is regulated by an intricate interplay of cis elements and trans factors. Journal of neurochemistry. 2004;88(5):1078-90.

281. Chung U-i, Igarashi T, Nishishita T, Iwanari H, Iwamatsu A, Suwa A, et al. The interaction between Ku antigen and REF1 protein mediates negative gene regulation by extracellular calcium. Journal of Biological Chemistry. 1996;271(15):8593-8.

282. Chattopadhyay R, Das S, Maiti AK, Boldogh I, Xie J, Hazra TK, et al. Regulatory role of human AP-endonuclease (APE1/Ref-1) in YB-1-mediated activation of the multidrug resistance gene MDR1. Molecular and cellular biology. 2008;28(23):7066-80.

283. Close P, East P, Dirac-Svejstrup AB, Hartmann H, Heron M, Maslen S, et al. DBIRD complex integrates alternative mRNA splicing with RNA polymerase II transcript elongation. Nature. 2012;484(7394):386-9.