THE UNIVERSITY
of EDINBURGH

Transcriptomic changes during differentiation of the leukaemia cell line THP-1 and the role of chromatin modifying enzymes

Iveta Gažová

Thesis submitted for the degree of Doctor of Philosophy,
College of Medicine and Veterinary Medicine
University of Edinburgh

April 2018

Declaration

I declare that the work presented in this thesis is my own, except when clearly stated otherwise by reference and/or acknowledgement. Any work performed in collaboration or conducted by others is explicitly acknowledged. This work has not been submitted for any other academic degree or professional qualification.

Iveta Gažová

… April 2018

Abstract

During normal cell development, many genes are activated and repressed, usually through epigenetic mechanisms. These are modifications of the DNA and protein within the nucleus that result in changes in gene expression without alteration in DNA sequence. Key proteins for epigenetic modifications are the histone proteins bound to DNA in the nucleus. The best-characterised epigenetic complexes that modify histone proteins are the polycomb group proteins (PcG), comprising polycomb repressive complexes 1 (PRC1) and 2 (PRC2). The repressive modifications generated by these complexes can be removed, and the blocked genes reactivated, by enzymes that are the subject of this project. PRC1 repressive marks are removed by deubiquitinases USP12, USP16 and BAP1, whereas PRC2 marks are removed by demethylases KDM6A, KDM6B and potentially UTY. During the development of cancer, the regulation of many genes becomes abnormal, allowing the cells to escape normal growth restrictions. In this thesis, the expression of this set of chromatin-modifying enzymes in a leukaemia cell line was investigated. The FANTOM consortium has been helping to understand patterns of gene expression for over 10 years. The FANTOM4 dataset described changes in gene expression and promoter usage during differentiation of the THP-1 acute monocytic leukaemia cell line, using CAGE (Cap Analysis of Gene Expression) technology. This human monocyte-like cancer cell line can be stimulated with phorbol esters to halt proliferation and differentiate into macrophages. However, the FANTOM4 time course did not capture detailed mechanisms of regulatory factors in macrophage differentiation due to sparse time points and low read coverage.

The main aim of this project was therefore to repeat the time course with tighter time points and deeper sequencing of the transcriptome to develop a very precise picture of sequential activation of gene expression, transcription start site (TSS) usage and the activity of enhancers during the transition from proliferating monocytes to differentiated macrophage phenotype of the THP-1 leukaemia cell line, using CAGE. The focus of this research was on the chromatin-modifying enzymes, but other key cell cycle and macrophage genes have also been examined. The differentiation time course was repeated in triplicate. RNA was extracted and CAGE libraries generated

for 18 time points, including the 6 originally studied in FANTOM4. Sequencing results were analysed and normalised using bioinformatics tools. It was shown that analysing 8 samples on one Illumina HiSeq 2500 lane yielded enough read coverage to detect activity from even low expression TSSs, such as those associated with enhancer activity. Clusters of genes which were up- and downregulated at different time points during the differentiation process were identified and characterised. CAGE results for key genes encoding chromatin modifying enzymes and macrophage markers were validated by qRT-PCR. There was a rapid increase of histone demethylase *KDM6B* mRNA once differentiation was initiated. Histone deubiquitinase *USP12* mRNA was also upregulated early in the process. Histone deubiquitinase *BAP1* mRNA shows an interesting cyclic regulation pattern which was not seen in the more limited samples of FANTOM4.

These interesting chromatin-modifying enzymes and their close paralogues (deubiquitinases USP12, USP16 and BAP1, together with demethylases KDM6A, KDM6B and UTY) were investigated by bioinformatics and genetic tools. *USP16* knockout THP-1 cell line was successfully created using CRISPR-Cas9 and its ability to differentiate into macrophages was examined using cell cycle analysis and CAGE sequencing. The *USP16* knockout cell line, along with siRNA knock downs of *USP12*, *USP16* and *BAP1*, was also compared to wildtype THP-1 differentiation using CAGE. Unfortunately, creating other mutant THP-1 cell lines was unsuccessful due to low THP-1 viability after single cell sorting. Investigating KDM6A, KDM6B and UTY using bioinformatics showed that *UTY* and *KDM6A* gene expression is positively correlated and this is disrupted in cancer samples. Gene expression and sequence comparison suggested that *KDM6A* and *UTY* are coregulated and may act in a similar way in histone demethylation.

In summary, the results in this thesis show the transcriptomic changes as the leukaemia cell line ceases proliferation and commences differentiation. Detailed examination suggests that histone modifications are important in the transition between proliferation and differentiation and provide better understanding of regulatory factors in macrophage differentiation and leukaemia.

Lay summary

All the cells which make up our body contain the same genetic material (DNA), but perform different functions. For example, muscle cells are specialised to contract and expand, whereas white blood cells need to fight against pathogens. They are able to do different things from the same DNA by expressing different genes to make different proteins for their specialised functions.

These changes in gene expression are controlled by various mechanisms, one of which is by altering the histone proteins that bind to DNA, forming chromatin. The presence of chemical modifications of the histones tells the cell to either start or stop expressing the gene. In this thesis, the enzymes responsible for removing some of these modifications are investigated.

The removal of these chromatin marks is especially important in the process of cell differentiation, in which cells are transformed from general precursors into specialised cells. Monocytes are precursor cells found in the blood which can stop multiplying and differentiate into macrophages. These specialised macrophages are able to recognise and dispose of any dangerous pathogens or cancer cells by a process called phagocytosis. Studying macrophages is important for understanding how the body can fight against bacteria, viruses and abnormal cancer cells.

For this study, a human leukaemia cell line was used as it is very similar to monocytes and can be induced to differentiate from monocyte to macrophage. Changes in gene expression across the entire DNA were looked at hour by hour in this experiment. By examining the activity of the DNA (gene expression) at close time points, it was possible to understand what was happening in the cell in great detail as differentiation progressed. In particular, the study looked at how chromatin modifying enzymes affected differentiation.

This study also looked at what happened to cancer cells when the chromatin-modifying enzymes were inactivated and how this affected differentiation into macrophages. This research helps in the understanding how cancer progresses and how we can prevent it.

Acknowledgments

I would like to dedicate this thesis to my wonderful supervisor Professor Kim Summers. It is safe to say that without her this thesis would never exist. She has been the best supervisor I could have ever wished for, with her never-ending optimism and energy, her doors always open for when I needed it. Kim, thank you so much for everything you've ever done for me, coping with my ups and downs and always supporting and caring. I seriously don't know how you can do it all, but I think I will always strive to be more like you. I will miss our weekly chats, and I will always think you're very cool!

I would also like to thank my second supervisor Dr Andreas Lengeling and my unofficial supervisor Prof David Hume for proofreading my thesis chapters and all the support throughout my PhD. Thank you to my sponsors, Lady Tata Memorial Trust for allowing me to study this PhD and Roslin SMG for allowing me to finish this PhD.

There's so many people to thank, but I would like to mention everyone in Maclab and for making my last three years so enjoyable (cake is the most important feature of any lab), especially Ailsa Carlisle (the core of Summers lab), Dr Clare Pridans (very important person if you need to know anything about macrophages), Dr Sara Clohisey (for being my yoga and desk buddy – until she moved desks. Never forget.), Dr Tim Regan (for making me laugh during hard times and showing me cute videos of animals), Dr Anna Raper (for being the nicest person in the world) and Dr Steve Bush (the pun master and food spotter). Also to (now) Dr Rocio Rojo thanks for all those tea breaks, where we needed to just vent about our PhDs.

I also have to mention Lucas Lefevre, Lel Eory and again Steve Bush for helping me with CAGE, Professor Bernadette Dutia for giving me the much needed hot chocolate that day, Liz Archibald for being so caring and Dr Doug Vernimmen PhD for helping me realise I can survive anything, even this PhD!

Thank you Meadowlark Yoga studio (especially the best yoga teacher Frankie), your wonderful hot classes were essential for helping me cope with the stress of writing (together with green tea).

I could not have done this without my friends who kept me sane throughout the years, Veverice squad (Ola, Michelle, Tereza, Helga and Zuz), CathNik and Hein. Thank you so much, Dr Mari Pattison, for the much needed tea breaks and wine nights. Thank you for everything from Barcelona to Gibraltar, Ania Dowbaj, it's you next year! And thank you, my long distance besties, Peťa Kňazeová and Ivana Nemčoková, for being the best friends anyone could ever ask for!

We. Did. It. Writing two PhDs in one year is not the easiest task, but we managed! Thank you so much Dr Alex Brown for the never-ending support, all the wonderful food you cook and always believing in me. Your support and love means everything to me, and I would not be able to do this without you. I can't wait what the future without PhD stress holds for us!

Chcela by som sa tiež poďakovať mojej Mamine a Ocinovi, vďaka ktorým som mohla ísť študovať tak "ďaleko", haha. Bez vašej lásky a podpory by som to určite nezvládla. Maťulo, teraz si ty na rade, pamätáš ako si vždy chcel mať o jeden titul viac ako ja? :D Ďakujem aj Babine a Dedovi, ktorí ma vždy sledujú ako prilietam. Už sa na vás teším! Ľúbim vás všetkých.

Manuscripts in preparation

Gazova I, Lengeling A and Summers K. “KDM6A and UTY: the X and Y of histone methylation”
Based on Chapter 5 data.

Gazova I, Lengeling A, Hume D and Summers K. “Detailed transcriptomic analysis of THP-1 differentiation”
Based on Chapter 3 data.

Gazova I, Lengeling A, Hume D and Summers K “The role on USP16 in macrophage differentiation”
Based on Chapter 6 data.

Contents

List of Figures

List of Tables

List of abbreviations

A	Adenine
ac	Acetylation
BLAST	Basic local alignment search tool
bp	Base-pairs
BRCA1	Breast cancer 1
C	Cytosine
CAGE	Cap analysis of gene expression
CD	Cluster of differentiation
cDNA	Complementary DNA
ChIPseq	Chromatin immunoprecipitation sequencing
CRISPR	Clustered regularly interspaced short palindromic repeats
CTSS	CAGE-detected transcription start site
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
ESCs	Embryonic stem cells
FAB	The French-American-British classification system of haematologic disorders
FACS	Fluorescence-activated cell sorting
FANTOM	Functional annotation of mammalian genome
FBS	Foetal bovine serum
FDR	False discovery rate
G	Guanine
GFP	Green fluorescent protein
GO	Gene Ontology
H2Aub	Ubiquitination on histone 2A
H3K27me3	Tri-methylation of lysine 27 on histone 3
KO	Knockout
LPS	Lipopolysaccharide
m	Mili
mRNA	Messenger RNA

MTT	3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide
n	Nano
nt	Nucleotide
OMIM	Online Mendelian inheritance in man
p	Pico
PBS	Phosphate-buffered saline
PCR	Polymerase Chain Reaction
PMA	Phorbol 12-myristate 13-acetate
qRT-PCR	Quantitative Reverse Transcriptase PCR
RACE	Rapid amplification of cDNA ends
RNA	Ribonucleic acid
RNAi	RNA interference
RNAseq	RNA sequencing
RPMI	Roswell Park Memorial Institute medium
rRNA	Ribosomal RNA
SD	Standard deviation
SE	Standard error
sgRNA	Single guide RNA
siRNA	Small interfering RNA
T	Thymine
TF	Transcription factor
TPR	Tetratricopeptide
TSS	Transcription start site
U	Uracil
USP	Ubiquitin specific peptidase
µ	Micro

Chapter 1: Introduction

All organisms store information about how to build their cells in deoxyribonucleic acid (DNA). This molecule contains data encrypted in a four-letter code using adenine (A), thymine (T), guanine (G) and cytosine (C). Although elements of this code have been known for decades, all the ways in which the genetic material is implemented are still not fully understood. Almost every cell in the body of a multicellular organism contains the same genetic (DNA) material, but each cell only transcribes a part of this DNA into ribonucleic acid (RNA) that has specific functions, including translation into proteins which are essential for most tasks carried out by a cell. The overall aim of this thesis is to expand knowledge on the differential expression of RNA in different types of cells and cellular states.

1.1 Transcriptional control of gene expression

The term gene has been defined as a genomic sequence encoding a functional product, be it protein or non-translated RNA (Gerstein *et al*., 2007). Gene expression (the conversion of this information into a form that can be interpreted by the cell) is an extremely complex and dynamic process. It is highly specific to the cell and situation, and thus is influenced by many different factors.

1.1.1 Gene promoters

Protein-coding genes and some types of non-coding RNA genes are transcribed into mRNA by RNA polymerase II, an enzyme which initiates transcription at the beginning of the gene, called the transcription start site (TSS) (as reviewed by Jonkers and Lis (2015)). The region around the TSS is defined as the promoter, which is the site at which the transcriptional machinery is assembled, bringing together the RNA polymerase with the associated regulatory and processing molecules (Jonkers and Lis, 2015). Promoter structures can be identified as narrow (sharp) or broad, each associated with different characteristics, such as promoter architecture (Carninci *et al*., 2006). Textbooks usually describe an AT rich sequence (the TATA box) as a key component of eukaryotic promoters with transcript initiation beginning a precise

distance upstream, typically 25 nucleotides (Alberts *et al*., 2013). This type of TATA box promoter is associated with a sharp promoter structure and tends to show tissue specific gene expression (Carninci *et al*., 2006). In contrast, recent studies have shown that the majority of promoters in vertebrates lack the TATA box and are found in regions of the genome called CpG islands (Carninci *et al*., 2006, Forrest *et al*., 2014). These regions contain many doublets of cytosine followed by guanine (CpG) and higher average GC content compared to frequencies in the genome overall (Saxonov *et al*., 2006). Around 70% of human promoters overlap with CpG islands, and these allow transcription to initiate from many different nucleotides within a broad region (Carninci *et al*., 2006, Forrest *et al*., 2014). These broad promoters tend to be associated with genes that are widely expressed in multiple tissues, such as housekeeping genes (Saxonov *et al*., 2006, Deaton and Bird, 2011, Carninci *et al*., 2006, Forrest *et al*., 2014).

In addition to transcript initiation occurring at diverse points within a promoter region, many genes can have alternative promoters which express different transcripts from various TSS, depending on the context (Ayoubi and Van De Ven, 1996, Davuluri *et al*., 2008). This can result in transcripts with alternative 5' regions which may be subject to different post-transcription regulation, and sometimes omit protein-coding exons thereby potentially changing the function of the resulting protein (Forrest *et al*., 2014). Moreover, 10% of human promoters are bidirectional (Trinklein *et al*., 2004), so they are transcribed from both DNA strands - this is discussed further in section 1.1.2.

A cell must respond immediately to environmental cues, so the activation and inactivation of genes and the choice of promoter must be rapid and flexible. For example, when treated with lipopolysaccharide (LPS), macrophages produce an early response of rapid changes in gene expression (Baillie *et al*., 2017). This is achieved by multiple regulatory elements, which either enhance or repress transcription from specific genes. The main regulatory elements are transcription factors (TFs), which recognise and bind to 6-12 bp DNA sequences within the promoter region (called motifs), These TFs work with other elements to recruit the RNA polymerase and allow

transcription to start (Spitz and Furlong, 2012). The processes that facilitate the binding of transcription factors are discussed in more detail below.

1.1.2 Enhancers

As well as the proximal regulatory elements, such as transcription factor binding motifs within promoter sequences, gene expression is influenced by more distal DNA sequences – enhancers (Fukaya *et al*., 2016). Enhancers play a crucial role in driving specific gene expression by delivering the transcriptional machinery to the promoters (Calo and Wysocka, 2013). Studies using different methodologies and criteria have identified between 40,000 and 400,000 enhancers in the human genome, as many as 20 enhancers per gene (ENCODE, 2012, Andersson *et al*., 2014).

Not all enhancers are considered equal in strength and importance. "Super-enhancers" are large clusters of enhancers within up to 50 kb of DNA that drive expression of those genes that define the identity of the cell (Hnisz *et al*., 2013). Enhancers act through binding both master transcription factors and mediator coactivators. For example, super-enhancers in embryonic stem cells (ESCs) are occupied by well-known stem cell transcription factors such as OCT4 (also known as POU5F1), SOX2 or NANOG (Whyte *et al*., 2013). Differences between normal enhancers and super-enhancers include size, transcription factor density, activation of transcription and sensitivity to perturbation (Whyte *et al*., 2013).

Another category of enhancers is the latent enhancers, which are found in differentiated cells. Latent enhancers are defined as regions of the genome which do not have any transcription factors bound to them, nor do they normally have enhancer activity but they can acquire it in response to stimuli (Ostuni *et al*., 2013). Once activated, these enhancers do not go back to their inactive state even after the removal of the stimulus. This could allow this type of enhancer to provide a memory of the environment. These latent enhancers illustrate the dynamism and versatility of enhancer elements in regulating gene expression.

Discovering new enhancers is difficult as no consensus sequence exists and their distance from their target gene is highly variable (Pennacchio *et al*., 2013). They can

be found upstream or downstream of the gene, in close proximity or many kilobases away. One way to characterise new enhancers is to look for specific histone modifications known to be associated with enhancers (Pennacchio *et al*., 2013). This was the approach used by ENCODE to identify putative enhancers (ENCODE, 2012). In addition, a number of reports have recently shown that enhancers are characterised by bidirectional transcription which can be used to identify candidate enhancers (Andersson *et al*., 2014, Kaikkonen *et al*., 2013, Li *et al*., 2013b, Kim *et al*., 2010, Hah *et al*., 2015). This method, notably used by the FANTOM5 consortium (discussed in section 1.4.1), detected far fewer enhancers than the ENCODE project.

1.1.3 Epigenetic control of gene expression

In a broad sense, epigenetic modifications are stable heritable alterations of the DNA and proteins within the nucleus that result in changes in gene expression without alteration in DNA sequence (Berger *et al*., 2009). Genetic information is efficiently stored by packing the DNA around proteins into a structure called chromatin. The basic unit of chromatin is the nucleosome, which is comprised of 2 copies of each of the core histones, H2A, H2B, H3 and H4. Wrapped around this histone octamer is a 147 bp long stretch of DNA (Richmond and Davey, 2003, Struhl and Segal, 2013). Gene expression requires enzymes such as RNA polymerase II to access the DNA, so packaging into nucleosomes could potentially obstruct cellular processes. Therefore, epigenetic modifications are used to disrupt nucleosomes and open the chromatin in the region of a gene that needs to be expressed. These modifications mark the functional elements of DNA to guide enzymes and other proteins that regulate gene expression, allowing activation or repression of gene expression depending on the requirements, state of differentiation and environmental stimuli to the cell.

The importance of these chromatin modifications is highlighted by studies which indicate that the epigenetic state of enhancers dictates the regulation of gene expression (as reviewed in Ong and Corces (2011)). This epigenetic patterning occurs before cell differentiation, suggesting that the information needed to determine the fate of the cell is located on the enhancers themselves (Ong and Corces, 2011). This information is dynamic and may differ from one enhancer to another. Further research

could help us to determine how various developmental pathways are regulated and several key epigenetic marks are discussed in the next section.

1.2 Epigenetic marks

Epigenetic marks can be found on both the DNA and proteins of chromatin. The DNA can be modified, usually by methylation of cytosine residues, which usually results in repression of activity when found in a promoter region (Long *et al*., 2016, Deaton and Bird, 2011). Methylation of cytosine is essential for proper development, as it is involved in many cellular processes including genomic imprinting, which leads to differently expressed alleles from mother and father, and X-inactivation, which ensures that only one copy of X chromosome is expressed in mammalian females (Sharp *et al*., 2011, Elhamamsy, 2017). Although the human genome is widely methylated, one functional feature usually contains long stretches of unmethylated cytosines – the CpG islands (Deaton and Bird, 2011) characteristic of broad promoters (section 1.1.1). It is likely that the low level of methylation in these elements is necessary to ensure that the promoter is accessible to RNA polymerase, representing one form of epigenetic regulation of gene expression.

A major epigenetic target, and the focus of the present study, are histone proteins. Unlike epigenetic modifications of DNA, there are hundreds of potential modifications for histones including methylation, ubiquitination, acetylation and others (Gates *et al*., 2017). Ubiquitination involves attaching the highly-conserved 76 amino acid protein, ubiquitin, which is an important regulator in various cell processes, such as the protein degradation and the cell cycle (Hard *et al*., 2010). Epigenetic marks are primarily found on the N-terminal ends of the histone molecules and each has different meaning for the cell, for example, histone acetylation almost always means gene activation (Bannister and Kouzarides, 2011). The focus of this thesis is on chromatin modifying enzymes which remove repressive histone marks to pave the way to gene activation.

1.2.1 Polycomb group proteins

The best-characterised epigenetic complexes that modify histone proteins are the polycomb group proteins, comprising polycomb repressive complexes 1 (PRC1) and

2 (PRC2) (Lanzuolo and Orlando, 2012). The repressive modifications generated by these complexes can be removed by chromatin modifying enzymes, deubiquitinases and demethylases, allowing the blocked genes to be reactivated (Figure 1-1).

PRC1 is composed of many proteins in mammals, most importantly RING1A and RING1B ubiquitin ligases (Lanzuolo and Orlando, 2012). PRC1 ubiquitinates the lysine at position 119 of histone 2A (H2AK119) most usually via RING1B, creating a repressed state of chromatin (Lanzuolo and Orlando, 2012) (Figure 1-1). H2A ubiquitination (H2Aub) blocks transcription through preventing RNA polymerase II from progressing along the DNA, for example at the promoters of developmental genes needed for differentiation in ESCs (Embryonic stem cells) (Stock *et al*., 2007). Recent studies showed that H2Aub also marks bivalent genes (genes containing both repressive and active histone marks) and regulates their expression in ESCs (Yang *et al*., 2014, Stock *et al*., 2007, Ku *et al*., 2008). Depletion of RING1A and RING1B causes the loss of ESC identity (Stock *et al*., 2007). PRC1 may also contribute to gene repression by directly compacting chromatin, blocking remodelling and inhibiting transcription initiation (Yang *et al*., 2014).

The best-known repressive epigenetic mark is trimethylation of histone 3 at lysine 27 (H3K27me3), which is generated and maintained by PRC2 (Lanzuolo and Orlando, 2012) (Figure 1-1). Mammalian PRC2 is composed of EED, EZH1/2, SUZ12 and RbAp46/48 proteins (Lanzuolo and Orlando, 2012) and H3K27 trimethylation is produced by the EZH1 or EZH2 methyltransferases (Yoshimi and Kurokawa, 2011). H3K27me3 is responsible for gene silencing and is implicated in bivalent gene domains in ESCs (Ku *et al*., 2008). The activating epigenetic mark, H3K4me3, is generated by MLL1 (Schuettengruber *et al*., 2007).

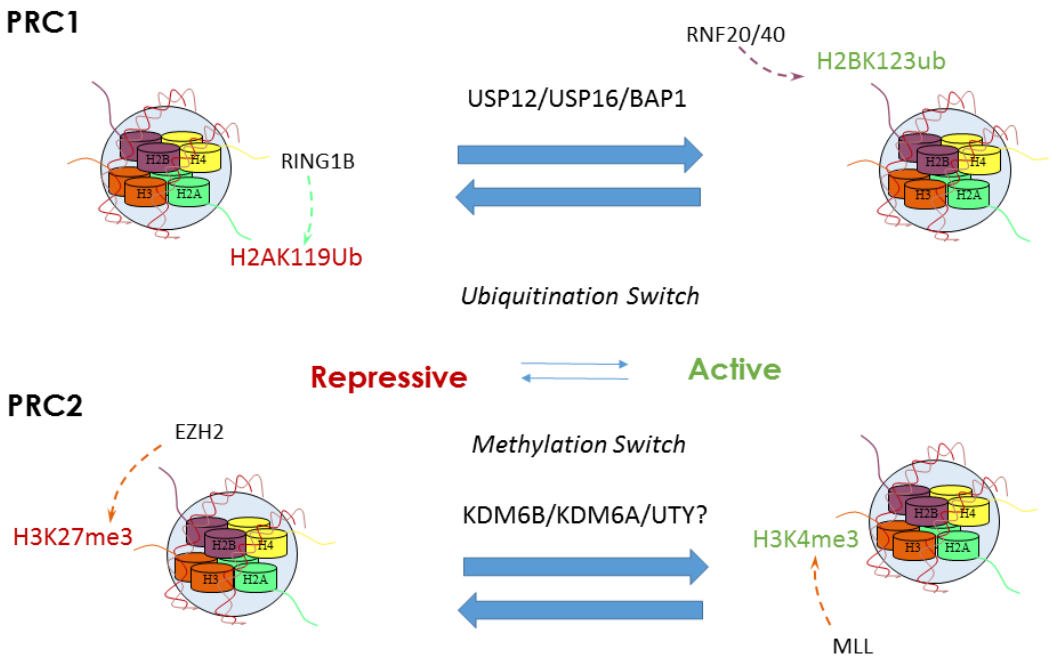

Figure 1-1 PRC1 and PRC2 create a repressive state through methylation and ubiquitination of histones. USP12/USP16/BAP1 erasers remove repressive mark H2AK119Ub and RNF20/40 writers create an active H2BK123ub. KDM6A and KDM6B (possibly also UTY) erasers remove repressive mark H3K27me3 and MLL1 writer creates an active H3K4me3 mark.

1.2.2 Removal of PRC1 associated marks

As mentioned above, during transcriptional activation, H2A deubiquitinases remove ubiquitin from H2AK119, and H2B histone is ubiquitinated on K123 (Abdel-Wahab and Dey, 2013). This ubiquitination switch is very important for the balance between repression and activation of gene expression. A small number of H2AK119 deubiquitinases have been identified, among as many as 100 deubiquitinases encoded in the human genome (Belle and Nijnik, 2014, Komander *et al*., 2009). These are the six ubiquitin specific proteases (USPs; USP3, USP12, USP16, USP21, USP22, USP44), one ubiquitin C-terminal hydrolase (BAP1) and one Zn^{2+} metalloprotease (MYSM1).

Three of these enzymes were selected for detailed examination in this study – BAP1, USP12 and USP16, which were found to control *Hox* gene expression during embryonic development in the mouse. *USP12* and *USP16* are less well-known in the scientific literature than *BAP1*, but all three have been implicated as possible tumour

suppressor genes. Another deubiquitinase, MYSM1, was found to be a regulator of B cell development and defects in *Mysm1* expression were detrimental to proper differentiation of other hematopoietic cells in the mouse (Jiang *et al*., 2011, Forster *et al*., 2015). USP21 has a possible role in T cell development and the cell cycle (Li *et al*., 2016, Urbe *et al*., 2012). The roles of USP3, USP22 and USP44 were mostly in DNA damage repair, with USP22 also implicated in carcinogenesis (Sharma *et al*., 2014, Ramachandran *et al*., 2016, Schrecengost *et al*., 2014, Mosbech *et al*., 2013).

1.2.2.1 BAP1 (BRCA1-associated protein 1)

BAP1 was first identified as a ubiquitin carboxy-terminal hydrolase, which binds to Breast/Ovarian Cancer Susceptibility Gene product (BRCA1) (Jensen *et al*., 1998). BAP1 deubiquitinase activity (against H2A though not H2B) was discovered much later, when *BAP1* was identified as a human homologue of *Drosophila calypso* (Scheuermann *et al*., 2010). BAP1 may also take part in many cellular pathways and processes, such as cell cycle progression, promoting cell proliferation and repair of DNA double strand breaks (Yu *et al*., 2014).

BAP1 is a component of Polycomb Repressive Deubiquitinase (PR-DUB) complexes, which have an important role in changes of gene expression. FOXK2, a forkhead transcription factor, was shown to recruit PR-DUB complexes (and thus BAP1) to chromatin, where BAP1 can promote local histone deubiquitination and causes changes in gene expression (Ji *et al*., 2014). This shows how BAP1 could be recruited to the regulatory regions of the genome.

In humans, heterozygous mutation of the *BAP1* gene is associated with tumour predisposition syndrome (OMIM 614327) (Abdel-Rahman *et al*., 2011, Rai *et al*., 2016). The *BAP1* gene is commonly deleted or mutated in multiple cancers, such as renal cell carcinoma, hepatic cancer and others (Battaglia, 2014, Popova *et al*., 2013, Testa *et al*., 2011, Wiesner *et al*., 2011). The loss of BAP1 function is usually considered a mark of poor prognosis in cancer patients (Joseph *et al*., 2014). Downregulation of *BAP1* was also found in chronic myeloid leukaemia patients with Philadelphia chromosome, which contains oncogenic *BCR-ABL1* fusion (Dkhissi *et*

al., 2015). This downregulation of *BAP1*, a direct effect of BCR-ABL, further upregulates ubiquitination of BRCA1, which is then targeted for degradation. Biallelic *BAP1* expression loss has also been associated with metastasis formation in uveal melanoma (Gezgin *et al*., 2017), where lack of BAP1 drives T cell influx and thus adds to the development of an inflammatory environment. *Bap1*$^{-/-}$ (knockout; KO) mice are embryonically lethal, where the embryos are reabsorbed by E9.5 (Dey *et al*., 2012). Conditional knockout in haematopoietic lineages in mice resulted in human myelodysplastic syndrome-like phenotype, emphasising the role of BAP1 as a tumour suppressor (Dey *et al*., 2012). These mice developed various haematological disorders, such as splenomegaly, monocytosis and neutrophilia.

Myelodysplastic syndrome and acute myeloid leukaemia patients have increased incidence of mutations of ASXL1 (Abdel-Wahab *et al*., 2012), which directly associates with BAP1 (Scheuermann *et al*., 2010). BAP1 (Calypso)-ASXL1 complex was found on the *HOX* genes, where it deubiquitinated H2A (Scheuermann *et al*., 2010). Interestingly, Abdel-Wahab *et al*. (2012) showed that ASXL1 promotes myeloid transformation through H3K27me3 loss, independently of BAP1. However, *Bap1* loss drives an increase in H3K27me3 levels due to elevated *Ezh2* in mesothelioma in mice (LaFave *et al*., 2015). Inhibitor against EZH2 was successful in treating mesothelioma cells, but caution must be taken to extend these results to other cancers such as uveal melanoma, where *BAP1* loss did not increase *EZH2* (Schoumacher *et al*., 2016).

In summary, there is substantial evidence that BAP1 acts as a tumour suppressor, both in experimental models (*Bap1* knockout mice) and cancer patients. However, the mechanism by which BAP1 represses cancer is still unknown, as is the question of whether it is linked to H2A deubiquitinase activity. Conflicting reports about the effect of BAP1 in various cancers show that the role of BAP1 is likely to be highly cell-specific.

1.2.2.2 USP12 (Ubiquitin-specific peptidase 12)

USP12 H2A deubiquitinase activity was first identified by Joo *et al*. (2011), together with that of USP46, which shares over 88% amino acid homology. H2A ubiquitin is removed by a complex between USP12 and the nucleosomal substrate, WDR48 (Joo *et al*., 2011). USP12 was found to deubiquitinate H2A and H2B with the same efficiency *in vitro* and *in vivo* (Joo *et al*., 2011). The expression of *USP12* and *USP46* was found to be coordinated, so that when *USP12* expression was knocked down, *USP46* gene expression increased, suggesting that they may have similar functions. Long term knock down of *usp12* (over 4 days) severely affected the development of *Xenopus laevis*, which suggests that USP12 is required for proper development (Joo *et al*., 2011).

USP12 has not been well-studied, but it has been found to serve multiple functions in the cell aside from the H2A deubiquitinase activity. For example, USP12 is a negative regulator of NOTCH signalling (Moretti *et al*., 2012) and together with WDR48 it negatively regulates AKT activation and promotes cellular apoptosis (Gangula and Maddika, 2013). Knock down of *USP12* was shown to increase proliferation of tumour cells and thus *USP12* was a potential tumour suppressor gene (Gangula and Maddika, 2013). USP12, together with WDR20 and UAF1, was found to be a co-activator of the androgen receptor, which is a transcription factor commonly altered in prostate cancer (Burska *et al*., 2013). These proteins and their mRNAs were found to be upregulated in prostate cancer (McClurg *et al*., 2015), with higher levels of USP12 when the other two were also present, suggesting a positive feedback loop. In contrast with findings of Gangula and Maddika (2013), knock down of *USP12* in prostate cancer cells reduced cancer proliferation (McClurg *et al*., 2015). USP12 also regulated cross talk between androgen receptor and AKT pathways in prostate cancer (McClurg *et al*., 2014). A super-enhancer (section 1.1.2) close to *USP12*, was found in colorectal carcinoma, which increased the expression of *USP12* (Zhang *et al*., 2016). USP12 is also involved in regulating cell cycle progression by involving cMyc, cyclin D2 and BMI1 (Tang *et al*., 2016), and its knock down induced a cell cycle arrest in HeLa cells. USP12 also acts as a crucial component of the T-cell receptor, and the Jurkat cell line

without *USP12* expression is defective in important cell signalling such as through NFκB, NFAT and MAPK (Jahan *et al*., 2016).

1.2.2.3 USP16 (Ubiquitin-specific peptidase 16)

USP16 was first mentioned by Cai *et al*. (1999) as UBP-M, and this 823 amino acid protein was localised to cytoplasm. Further research revealed that USP16 forms a tetramer, which is actively exported from the nucleus during interphase (Xu *et al*., 2013). However, at the onset of mitosis USP16 is phosphorylated at Serine 552, by CDK1 (Xu *et al*., 2013) and localises to nucleus, where it is required for G2/M progression. In mitosis, USP16 deubiquitinates PLK1, which is needed for proper chromosome alignment and without which cell cycle progression is stopped (Joo *et al*., 2007).

USP16 can also act as a negative regulator of the process of DNA damage repair. A DNA double strand break induces H2A ubiquitination at the site, and once the break is repaired the ubiquitin is removed by USP16 and HERC2 (Zhang *et al*., 2014). Levels of *USP16* mRNA are upregulated directly after DNA damage, and both downregulation and overexpression of *USP16* leads to reduced response to DNA damage (Zhang *et al*., 2014). Disruption of *USP16* expression may contribute to the inability of cancer cells to resolve the DNA repair process, and has been found downregulated or mutated in several human cancers, such as leukaemia, lung adenocarcinoma and hepatocellular carcinoma (Fernandez *et al*., 2004, Qian *et al*., 2016, Gelsi-Boyer *et al*., 2008).

The function of USP16 which is most important for this project is gene activation. USP16 binds to promoter regions in various genes on all chromosomes of ESCs, and this was correlated with low H2A ubiquitination levels but high gene expression (Yang *et al*., 2014). USP16 was also shown to bind to active genes and promoters, taking on a role in shifting mouse B-cells from a quiescent to an active state (Frangini *et al*., 2013). In addition, USP16 was shown to regulate developmental *hox* genes in *Xenopus laevis* (Joo *et al*., 2007). All three of these functions may contribute to the embryonic lethality of *Usp16* $^{-/-}$ mice, where defects in the development were found as early as

E7.5 stage embryos after implantation (Yang *et al*., 2014). USP16 catalytic function was shown to be required for ESC differentiation, but not ESC viability (Yang *et al*., 2014). In addition, knock down of *USP16* in HeLa cells caused growth rates to slow, as there was a sharp decrease in G2/M cells and *HOX* gene expression. (Joo *et al*., 2007).

USP16 is located on human chromosome 21. Trisomy of this chromosome results in Down syndrome (Adorno *et al*., 2013). A study by Adorno *et al*. (2013) in mice showed that USP16 is the main candidate in causing accelerated senescence and aging-related phenomena in Down syndrome patients. Consequences of overexpression of *Usp16* in mice include reduction of hematopoietic stem cells and their self-renewal ability, cellular defects owing to increased removal of H2A ubiquitins and decreased proliferation (Adorno *et al*., 2013). Down syndrome patients also have increased incidence of leukaemia and decreased rate of solid tumours (Mateos *et al*., 2015). *USP16* overexpression was also shown to cause reduced DNA damage response in human Down syndrome cell lines (Zhang *et al*., 2014).

Taken together, these reports suggest that cells are sensitive to any changes in USP16 dosage, as either increase (Down syndrome) or decrease (mutations in cancers) causes decreased DNA damage response and higher incidence of some cancers.

1.2.3 Removal of PRC2 associated marks

Trimethylation of lysine 27 of histone 3 (H3K27me3) is catalysed by PRC2 and is associated with gene repression. Histone demethylases, which remove di- and trimethyl groups on H3K27, are responsible for radical alteration of epigenetic states to create new cell-specific gene expression profiles (Miller *et al*., 2010). Two demethylases that contain a Jumonji C (JmjC) catalytic domain are responsible for removing the di-and trimethyl groups on H3K27, as described below (Hong *et al*., 2007).

1.2.3.1 KDM6A (UTX) and UTY

Like the deubiquitinating chromatin modifying enzymes discussed in the previous section, the demethylase KDM6A has also been identified as a tumour suppressor protein, which was deleted in several human tumours, including multiple myeloma, oesophageal, pancreatic and renal cancer (Van der Meulen *et al*., 2014, van Haaften *et al*., 2009, Waddell *et al*., 2015, Ler *et al*., 2017). Using an inhibitor against EZH2 seemed to antagonise the carcinogenesis caused by loss of KDM6A in bladder cancer, which shows a potential new therapeutic target for treating this type of cancer (Ler *et al*., 2017).

The *KDM6A* gene is located on the X chromosome in mammals, and it has a homologue on the Y chromosome called *UTY,* which shows similar patterns of expression to *KDM6A* (Xu *et al*., 2008). *UTY* shares 84% sequence homology to *KDM6A*, but the protein was thought to be enzymatically inactive. However, a recent report suggests that UTY might also have a demethylase activity (Walport *et al*., 2014). All *Kdm6a*$^{-/-}$ females but only some *Kdm6a*$^{-}$ males (which carry *Uty* on the Y chromosome) die in utero. *Uty* KO mice are viable, but *Kdm6a/Uty* double KO mice are embryonically lethal (Shpargel *et al*., 2012). Most likely UTY can compensate to some extent for loss of KDM6A in adults and partially during development. UTY has not been widely recognised as sharing full redundancy with KDM6A; these observations suggested that further analysis of the similarities between KDM6A and UTY was necessary, which is presented in Chapter 5 of this thesis.

Although KDM6A seems to have no role in ESC renewal, it is needed for bivalency resolution and activation of retinoic acid inducible genes during differentiation of mouse ESCs (Dhar *et al*., 2016). Furthermore, Morales Torres *et al*. (2013) discovered that KDM6A is required for proper differentiation of ectoderm and mesoderm in mouse ESC, but not their proliferation. Remarkably, ectoderm and mesoderm are established in the absence of the demethylase activity of KDM6A (Morales Torres *et al*., 2013). UTY and KDM6B have been shown to partially compensate for KDM6A loss in male ESCs. During their knock down using RNAi, there was a significant but small downregulation of genes needed for ectoderm and mesoderm differentiation (Morales Torres *et al*., 2013).

In humans, expression levels of the *KDM6A* gene are the highest in haematopoietic stem and progenitor cells than decline during haematopoietic differentiation (Van der Meulen *et al*., 2014). KDM6A has been identified as a novel regulator for haematopoietic cell migration using conditional *Kdm6a* KO mice (Thieme *et al*., 2013). KDM6A was also found to epigenetically activate expression of cytokines IL-6 and IFN-β in primary macrophages by removing H3K27me3 from their promoters (Li *et al*., 2017).

1.2.3.2 KDM6B (JMJD3)

The related histone demethylase KDM6B lacks the tetratricopeptide repeats (TPRs) that are present at N-terminal regions of the KDM6A and UTY proteins (see section 5.1). These TPRs are important for protein-protein interactions (van den Meulen *et al*., 2014), which might suggest a lack of redundant functions between KDM6A and KDM6B.

No correlation between KDM6B recruitment and changes in H3K27me3 levels has been observed (De Santa *et al*., 2009, De Santa *et al*., 2007). KDM6B interacts with the p53 tumour suppressor and is transcriptionally upregulated following DNA damage (Williams *et al*., 2014). The authors speculated that KDM6B could act as a fail-safe mechanism to maintain low levels of H3K27me3, so the chromatin stays permissive for activation by H3K27 acetylation at p53 response elements, instead of directly activating target genes. Recently, KDM6B was found to downregulate OCT4 expression at both mRNA and protein level without its demethylase activity (Xun *et al*., 2017). Therefore, this data further supports the theory that both KDM6A and KDM6B and possibly UTY have a secondary function independent of their demethylase activity.

Both KDM6A and KDM6B were found to catalyse the removal of the repressive histone mark at genes important for T cell differentiation (Manna *et al*., 2015). Interestingly, a recent study identified *KDM6B* as an oncogene and *KDM6A* as a tumour suppressor gene in T-cell acute lymphoblastic leukaemia, but did not completely address the role of *UTY* (Ntziachristos *et al*., 2014). KDM6B is also

upregulated in macrophages, where it is important for the activation of inflammatory cytokine genes such as *IL23A* (also known as IL-23p19), *CSF3* (*GCSF*) and *TREM1* (Yan *et al*., 2014).

Deletion of *Kdm6b* results in perinatal lethality at E18.5 of mouse development, where KDM6B has been shown to have a role in the maintenance of the embryonic respiratory neuronal network, breathing and survival (Burgold *et al*., 2012).

1.2.3.3 Link between PRC1 and PRC2

Bivalent domains in mouse ESCs containing both PRC1 and PRC2 modifications appear to retain H3K27me3 more efficiently upon differentiation, show high evolutionary conservation and associate with a large number of gene promoters controlling development (Ku *et al*., 2008). According to the study by Ku *et al*. (2008), both sets of PRCs are required for proper gene silencing and for differentiation. It is not yet known how ubiquitously expressed chromatin-modifying enzymes are capable of establishing and maintaining specific patterns of gene expression for differentiating cells. Most probably, cell-type specific gene expression patterns require distinct combinations of chromatin regulators at different times, with PRC1 and PRC2 working together (DeVilbiss *et al*., 2013).

There are several instances of this happening. Blackledge *et al*. (2014) discovered that H2Aub deposited by PRC1 leads to recruitment of PRC2 and trimethylation of H3K27 to establish a repressive state of chromatin. BAP1 interacts with ASLX1, as discussed above (section 1.2.2.1), and expression of ASXL1 was associated with abundance of the H3K27me3 repressive PRC2 mark, suggesting that ASXL1 could be responsible for recruitment and stabilisation of PRC2 complex (Abdel-Wahab *et al*., 2012). Both BAP1 and ASLX1 are probably involved in a large number of interactions, but these observations show that they could form a link between PRC1 and PRC2 gene silencing, even if they do not occur together. These results present a mechanism by which PRC1 and PRC2 might interact to regulate chromatin state.

1.3 Cellular differentiation

The change from totipotent single celled zygote to a multicellular adult organism with hundreds of different cell types involves a series of developmental and differential processes. Studying cell differentiation is useful for determining the cause of developmental anomalies of the organism and for understanding disease processes, particularly those caused by inappropriate expression or repression of genes, such as cancer. Chromatin plasticity is tested to an extreme degree during differentiation, as various genes need to be blocked and others turned on in quick succession. It is to be expected that chromatin modifying enzymes will have an important role in influencing these transitions in gene expression during differentiation.

The aim of this thesis is to explore the transcriptomic changes involving chromatin-modifying enzymes in relation to cell differentiation. All of these genes are implicated as having an effect in cancers, with most of them confirmed or suspected as tumour suppressor genes. The system chosen for transcriptomic analysis was monocyte to macrophage differentiation, as haematopoiesis involves rapidly dividing and proliferating cells which undergo transitions in cell state as the final mature phenotype is developed (Gordon and Taylor, 2005).

1.3.1 Monocyte-macrophage differentiation

Monocytes originate from a common myeloid progenitor in bone marrow, where they are released into peripheral blood (Gordon and Taylor, 2005). They circulate for several days, before differentiating into macrophages, characterised by growing in size, increasing the number of lysosomes and becoming adherent (Figure 1-2). This differentiation is dependent on stimulation by macrophage colony stimulating factor (CSF1), which binds to the extracellular domain of the CSF1 receptor (CSF1R) (Wallner *et al*., 2016). Further changes are orchestrated by a cascade of different transcriptomic modifications enabled by upregulation or downregulation by various transcription factors, such as PU.1 (also known as SPI1) or FOXP1 (Shi *et al*., 2008, Zhang *et al*., 1994, Hume and Freeman, 2014). Monocytes undergo maturation in the circulation to generate subsets of cells that differ in expression of marker genes including CD14 and CD16 (Schmidl *et al*., 2014). Further monocyte differentiation

in response to CSF1 produces major changes in the transcriptome, analysed using CAGE (see section 1.4.1), including down-regulation of pattern recognition receptors and proinflammatory mediators (Baillie *et al*., 2017, Hume *et al*., 2016).

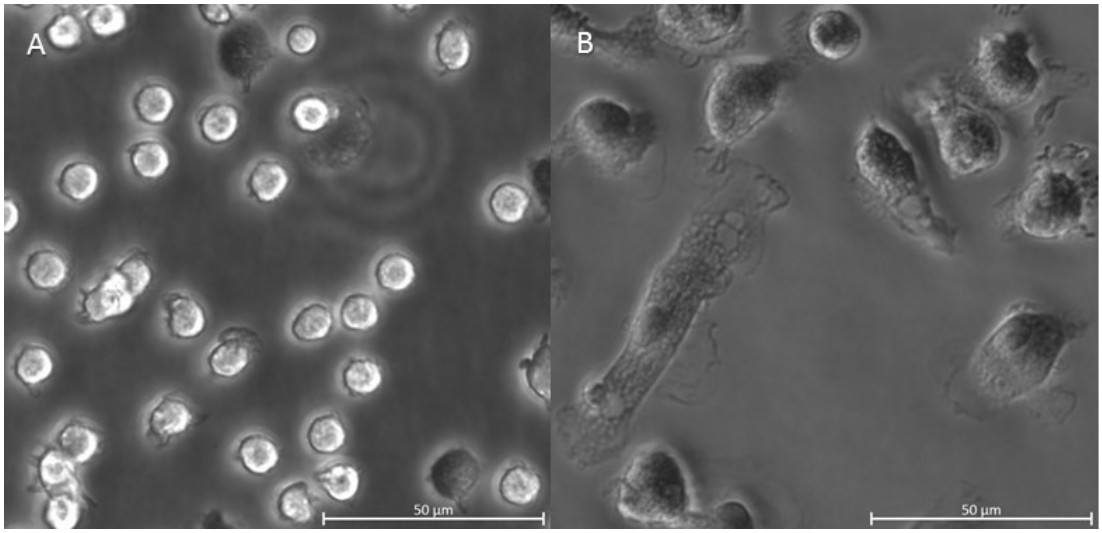

Figure 1-2 A) Human primary peripheral blood monocytes harvested from a female donor. B) The same cells differentiated into macrophages after 5 days in culture with CSF1. Scale bar = 50 µm. Images provided by Dr Tim Regan, The Roslin Institute, UK.

Monocytes and macrophages differ immensely in their epigenetic signatures (Wallner *et al*., 2016, Hoeksema and de Winther, 2016). Even the chromatin landscapes of macrophages in different tissues vary, based on their microenvironment (Lavin *et al*., 2014). This data suggests that the role of chromatin modifying enzymes must be critical in their dynamic transcriptomic change in differentiation of monocytes into macrophages. Thus, monocyte to macrophage differentiation was chosen as a model for studying these changes.

1.3.2 Leukaemia cell lines as an experimental model

Most experiments using genome-wide approaches to study gene expression require a large number of cells to be processed and analysed, which is difficult to achieve with primary cells from a single donor. There are many cell lines which were established from human cancer patients to provide a good source of monocyte or macrophage-like cells for researchers. Some of these are outlined below.

U-937 is an acute myeloid leukaemia cell line, established from a 37-year male patient (Sundstrom and Nilsson, 1976). At establishment, U-937 grew as a single-cell suspension with identical morphology to the tumour cells it was established from. Its karyotype years later was found to be quite variable with ~56-60 chromosomes, two X chromosomes and no Y chromosome (Lee *et al*., 2002). Even though this cell line was established from lymphoma, it expressed myelocytic markers, and can be stimulated to express monocyte-like phenotype in response to stimuli such as interferon γ or retinoic acid (Tanaka *et al*., 1995, Harris *et al*., 1985). U-937 cells can also be differentiated into macrophages by the protein kinase C agonist phorbol 12-myristate 13-acetate (PMA) (Yang *et al*., 2017), Another cell line is HL-60, which was established from an acute promyelocytic leukaemia. The morphology of HL-60 cells is predominantly that of the neutrophilic promyelocyte, but they appear to be able to phagocytose (Gallagher *et al*., 1979). Their karyotype is close to the human diploid complement of 46 chromosomes, with few chromosomal deletions (Gallagher *et al*., 1979). The Mono Mac 6 cell line was also derived from patient with monoblastic leukaemia (Ziegler-Heitbrock *et al*., 1988). This cell line exhibits characteristics of more mature monocytes, and can also be stimulated to acquire macrophage-like function by addition of PMA (Ziegler-Heitbrock *et al*., 1994).

The human cell line most widely used as a model to study macrophage differentiation and function is THP-1, which currently boasts 9083 papers indexed by PubMed (accessed November 2017). The THP-1 cell line was originally established from a 1-year-old male patient with acute monocytic leukaemia (Tsuchiya *et al*., 1980). This cell line differentiates into a macrophage-like phenotype in response to treatment with PMA, which initially inhibits cell growth prior to differentiation by upregulation of the cyclin-dependent kinase inhibitor CDKN1A (also known as $p21^{WAF1/CIP1}$); this, in turn inhibits the activating phosphorylation of CDK2. The increase of CDKN1A was mediated by an increase in binding of the transcription factor SP1 to the promoter and accompanied by an immediate increase in reactive oxygen species (ROS) generation. PKC signalling was essential for this response (Traore *et al*., 2005). Another chemical widely used to induce THP-1 differentiation is vitamin D (1,25 dihydroxyvitamin D3). However, this results in a less differentiated phenotype based on phagocytosis, loss of

proliferation and downregulation of *CD11B* and *CD14*, as vitamin D and PMA stimulate different cellular pathways in THP-1 cells (Schwende *et al*., 1996). Thus, PMA stimulation is usually used to study macrophage function (Qin, 2012). THP-1 cells are considered to have a more mature monocytic phenotype than U937 (Altieri and Edgington, 1988), but less than Mono Mac 6 cell line (Ziegler-Heitbrock *et al*., 1988).

In the parent THP-1 cell line available from the American Type Culture Collection (ATCC® TIB-202™), around 50% of cells become adherent in response to PMA (Traore *et al*., 2005). To enable a detailed study of the transcriptomic response to PMA, the FANTOM consortium isolated a clonal line that was highly responsive to PMA and uniformly differentiated into macrophage-like state (Suzuki *et al*., 2009). The clonal line's genome has not been sequenced, but its karyotype was established by microarray based comparative genomic hybridisation (Adati *et al*., 2009). This method concluded that the karyotype of the THP-1 cell line was 46XY, with minimal chromosome aberrations, which are deletions at 6p, 12p, 17p and a deletion on the X-chromosome, which coincidentally deleted the first few exons of *KDM6A* (Adati *et al*., 2009).

THP-1 cells are also very useful for the study of leukaemia. Their origin from an acute monocytic leukaemia predisposes them to certain genetic changes naturally occurring in cancer, such as mutations in multiple tumour suppressor genes (for example *TP53* and *PTEN*), and *MLL* fusions (Adati *et al*., 2009). Information on gene expression changes during differentiation could be used to discover drug targets and understand more of the biology of this form of cancer.

The THP-1 cell line was selected to use as an experimental system for this project due to its high responsiveness to PMA stimulation, homogeneity in genetic background in comparison with primary cells from donors, documented successful transfection with small interfering RNA (siRNA) (Suzuki *et al*., 2009, Maess *et al*., 2011), published previous transcriptomic data (Suzuki *et al*., 2009), and a wide ranging scientific audience potentially interested in detailed characterisation of this differentiation

process, from those using THP-1 cells as models for macrophages to those interested in the biology and treatment of haematopoietic cell cancers.

1.4 Transcriptomic analysis of monocytes and macrophages

Many genome-wide methods of analysis of gene expression have been implemented in study of monocytes and macrophages. Such studies are usually part of large datasets studying various primary cells, for example HaemAtlas examining microarray expression data in blood cells (Watkins *et al*., 2009). Both the epigenome and the transcriptome of primary monocytes compared with primary macrophages has been analysed extensively (Wallner *et al*., 2016, Saeed *et al*., 2014, Martens and Stunnenberg, 2013, ENCODE, 2012). The THP-1 transcriptome was also previously studied, for example recently after vitamin D3 stimulation by RNA sequencing (RNAseq) (Neme *et al*., 2016).

However, previous studies did not address the dynamic process of change in gene expression during monocyte to macrophage differentiation, as in general only the difference between these two types of cells and no intermediate states were investigated.

1.4.1 FANTOM (Functional annotation of mammalian genome)

One study which looked at the progression of changes in gene expression during THP-1 differentiation time course in response to PMA was that of the FANTOM4 consortium. This uncovered the transcriptional network using genome scale 5' RACE (rapid amplification of cDNA ends) (Suzuki *et al*., 2009). They termed this method cap analysis of gene expression (CAGE). CAGE is a high-throughput method for transcriptome analysis utilising a cap-trapping technique, in which short tags (27 nt long) at the 5' end of RNA transcript are sequenced (Takahashi *et al*., 2012). This technique relies on biotinylation of the methylguanosine cap, which is characteristic for RNA polymerase II transcripts (Carninci *et al*., 1996). By isolating RNA containing the 5' cap, the TSS of individual transcripts can be precisely tracked (see Figure 2-1; further details of the CAGE methodology are given in section 2.2.6). The

analysis of CAGE data provides information not only on the base-pair resolution of TSS across the genome but also on the number of individual transcripts at TSSs, (Forrest *et al*., 2014). Thus, CAGE data can be used to characterise various functional elements, such as promoters (Forrest *et al*., 2014), and enhancers (Andersson *et al*., 2014, Arner *et al*., 2015), as well as quantifying expression from specific promoters to develop a transcriptomic atlas of gene expression across cell types or states.

In an initial analysis, the FANTOM4 dataset of THP-1 differentiation in response to PMA (Suzuki *et al*., 2009) was used to investigate the expression pattern of the chromatin modifying enzyme genes. The datasets are available at http://fantom.gsc.riken.jp/zenbu/. The time course examined samples taken at 0, 1, 4, 12, 24 and 96 hours after stimulation with PMA. These results showed a peak in the mRNA levels for *KDM6B* and *USP12* around four hours after initiation of differentiation, with almost no change in levels of expression for *BAP1, USP16* or *UTY* (Figure 1-3, Figure 1-4). However, the FANTOM4 time course lacked time points at critical stages in the differentiation, between one and four hours (when both *KDM6B* and *USP12* were upregulated), between four and 12 hours (when *KDM6B* was downregulated) and between 12 and 24 hours (when *USP12* was downregulated).

The most recent phase of the FANTOM collaboration released a set of promoter and enhancer expression patterns for single time point static samples for nearly 1,000 human samples as well as time courses of cellular state changes in various systems (Forrest *et al*., 2014, Arner *et al*., 2015). In this data, *USP12* and *KDM6B* mRNA levels also increased drastically in a time course of human primary macrophages (derived from peripheral blood monocytes) stimulated by lipopolysaccharide (LPS) (Baillie *et al*., 2017) (Figure 1-5). These results support the results of an earlier study, where THP-1 derived macrophages stimulated by LPS also showed more than 2-fold increase in *USP12* (Iglesias *et al*., 2012). The FANTOM5 data was also explored in relation to epigenetic regulators in haematopoiesis (Prasad *et al*., 2014).

These results from the two FANTOM studies indicate that at least some of the chromatin modifying enzymes are regulated in monocytes undergoing differentiation

and macrophages responding to LPS. However, the depth of study for the THP-1 time course was hampered by the lack of key time points and the relatively low depth of sequencing available at that time. Therefore, further detailed study of the THP-1 gene expression in differentiation and the effect of chromatin modifying enzymes is likely to reveal additional information about key processes relating to cessation of proliferation and maturation into macrophages.

The main hypothesis of this project is that the activity and certain level of various chromatin modifying enzymes is crucial for monocyte to macrophage differentiation. Due to the fact that the gene expression of these enzymes seems to be tightly regulated in the THP-1 differentiation, inhibition of their activity or regulation of their level of expression should prevent the repressive epigenetic marks to be removed efficiently, and the THP-1 cells could remain in their monocyte-like state (non-adherent).

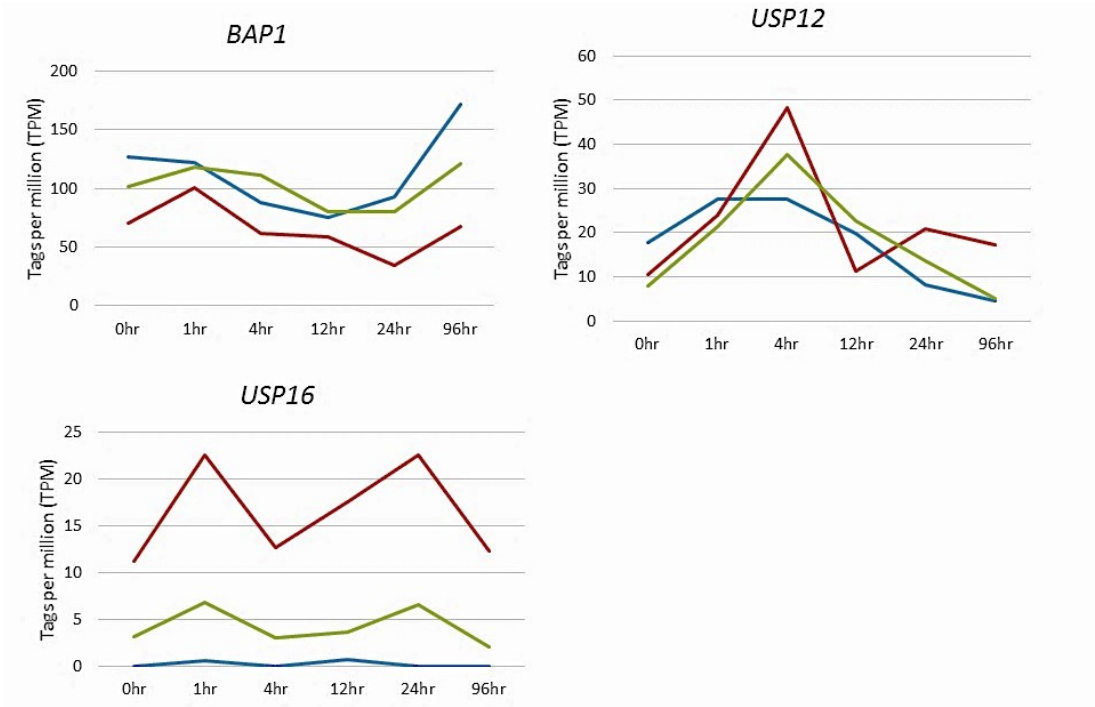

Figure 1-3 FANTOM4 time course of PMA-induced THP-1 differentiation for deubiquitinases analysed in this study. X axis shows the time points. Y axis shows normalised expression levels, in Tags per million. Three replicates are shown. mRNA levels increase for USP12 at around 4 hours after initiation of differentiation, whereas the levels for BAP1 and USP16 do not appear to change significantly.

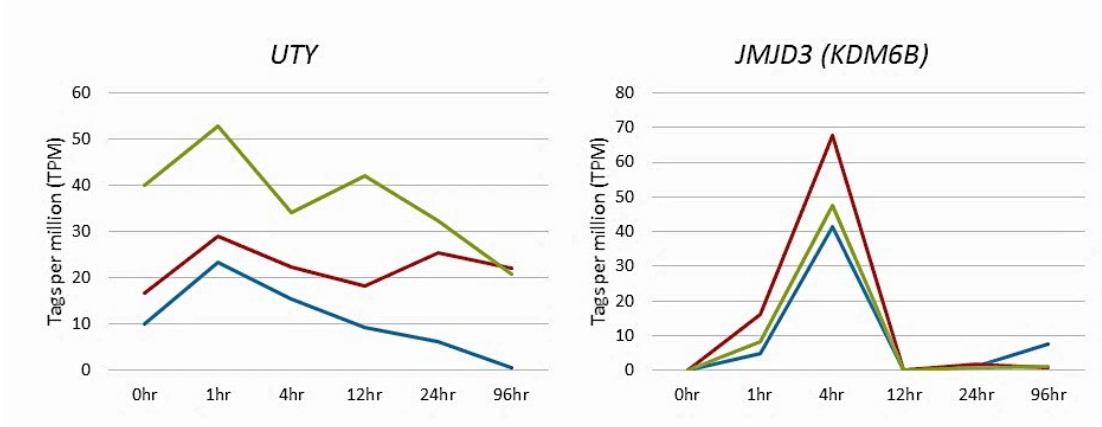

Figure 1-4 FANTOM4 time course of PMA-induced differentiation of THP-1 for demethylases analysed in this study. X axis shows the time points. Y axis shows normalised expression levels, in Tags per million. Three replicates are shown. mRNA levels increase for KDM6B but not for UTY. KDM6A is not expressed in THP-1 cells due to a deletion of first few exons.

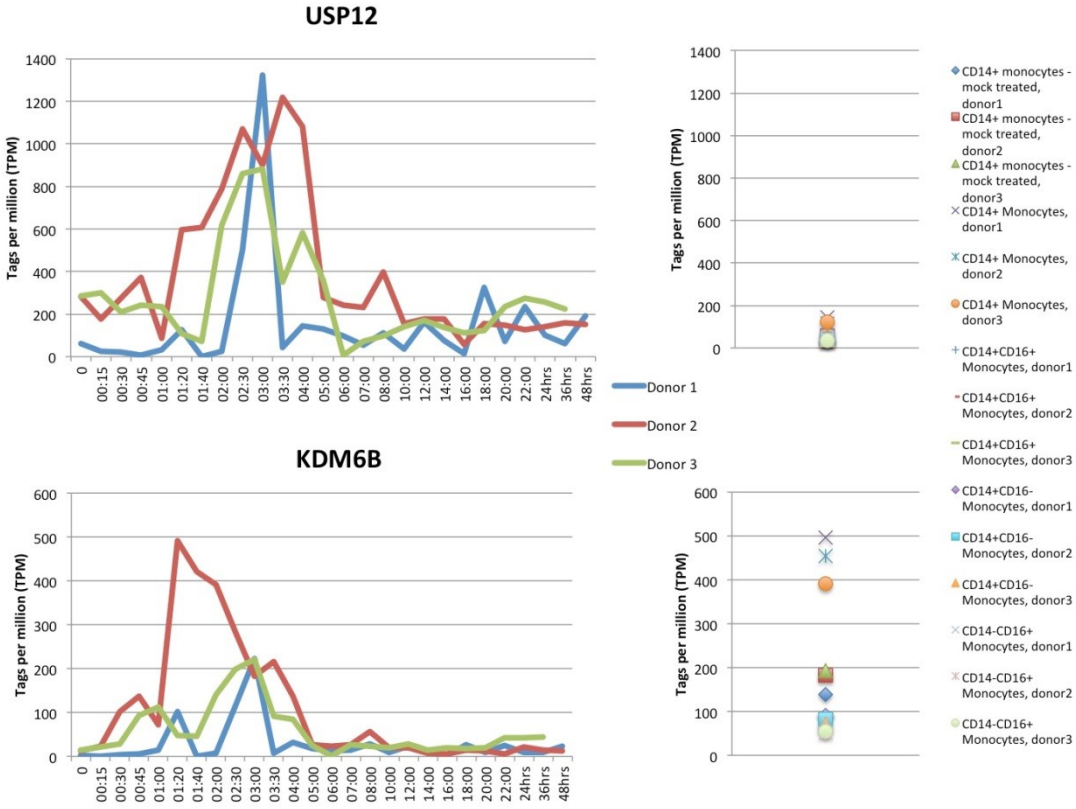

Figure 1-5 LPS treatment of monocyte-derived macrophages. X axis shows the time points. Y axis shows normalised expression levels, in Tags per million. Results for three different human donors are shown. mRNA levels for USP12 peak at around 3 hours after LPS treatment, whereas KDM6B mRNA levels are highest slightly earlier for donor 2 and similar to USP12 for the other donors. On the right, various unstimulated monocyte controls from FANTOM5 are shown at the same scale as the time course, with KDM6B having variable levels for various monocyte samples, while USP12 levels are less than 20% of the peak level in activated monocytes.

1.5 Aims and objectives of this thesis

A primary objective of this thesis was to obtain detailed knowledge of the transcriptomic changes involved differentiation of the cells. This study used THP-1 monocyte to macrophage differentiation as an experimental model, with data primarily generated by the CAGE technique, with the emphasis on the different chromatin modifying enzymes (Chapter 3). Addition of multiple time points not covered in the FANTOM4 study, especially in the 24 hours of the time course, and the greater depth of sequencing possible following technical advances in next generation sequencing, bring more resolution into the whole picture of the interplay between repression and activation of genes in differentiating cells.

The second part of the study (Chapters 4, 5 and 6) details the bioinformatics analyses of publicly available database data for H2Aub deubiquitinases (Chapter 4) and H3K27me3 demethylases (Chapter 5). This knowledge is used to assess the impact on the monocyte to macrophage differentiation when the level of activity of these chromatin modifying enzymes is affected either by RNA interference or enzyme-specific inhibitors (Chapters 4 and 5). In addition, the relationship between UTY and KDM6A is explored in Chapter 5 using various bioinformatic methods. Due to the fact that using RNA interference to knock down gene expression produced variable results (Chapters 4 and 5), the last experimental chapter asks the question whether knocking out these genes using the CRISPR-Cas9 genome editing system affects the THP-1 differentiation and if it is suitable to use in this type of cell line.

The impact of this study is to characterise the monocyte to macrophage transition in this model system in depth and highlight the potential roles of chromatin modifying deubiquitinases and demethylases, which may lead to therapeutic targets as well as greater understanding of the transcriptomic changes in the innate immune system.

Chapter 2: General Materials and Methods

This chapter presents the materials and methods that were used throughout the experiments presented in this thesis. More detailed and experiment-specific materials and methods are provided in the relevant chapters. All water used, if not mentioned otherwise, was Ultra Pure™ DNase/RNase free distilled water by Gibco (now Thermo Fischer Scientific/Invitrogen, Paisley, UK).

2.1 Cell line and THP-1 differentiation assay

THP-1 cells (clone 5, from FANTOM4 consortium, passage number 8, provided by Dr Mark Barnett, The Roslin Institute, UK) were cultured in sterile-filtered (0.2 μm) RPMI 1680 medium with HEPES modification (25 mM HEPES + $NaHCO_3$) (R5886 Sigma-Aldrich, Gillingham, UK), 10% heat inactivated foetal bovine serum (FBS, GE Healthcare, PAA laboratories, Pasching, Austria), 1x Glutamax (35050-038, Gibco), 1 mM sodium pyruvate (11360-039, Gibco), 1x MEM non-essential amino acids solution (11140-035, Gibco), 50 μM 2-mercaptoethanol (31350-038, Gibco) and penicillin-streptomycin (16 U/ml and 16 μg/ml, Gibco) (later referred to as THP-1 medium). Cells were incubated at 37°C, 5% CO_2 and their concentration was maintained between $2x10^5$ – $8x10^5$ cells per ml by splitting in half three times a week.

For THP-1 differentiation assays, THP-1 cells were grown up at a low passage number (typically p10-13). The day before the start of the differentiation, cells were counted by haemocytometer and for each time point between $3x10^6$ and $5x10^6$ cells were pelleted and resuspended in 10 ml fresh medium. THP-1 cells were then differentiated by adding 30 ng/ml (48.6 nM) phorbol 12-myristate 13-acetate (PMA; P1585, Sigma-Aldrich) in DMSO. The cells were plated on a Sterilin plate (without tissue culture treatment of the plastic), and after 4-6 hours of differentiation, the cells were lifted off by flushing with a blunt-end needle syringe.

2.2 Gene expression analysis

2.2.1 RNA extraction and DNase I treatment

RNA was extracted from pelleted cells using RNABee (AMS Biosciences, Frienswood, TX, USA) or TRIzol (ThermoFischer Scientific) according to the manufacturers' instructions. Cells were pelleted and resuspended in 1 ml RNABee or TRIzol, then frozen down to -80°C. The protocol was resumed later at a suitable time. The cells in RNABee or TRIzol were defrosted on ice, 0.2 ml pure chloroform was added and the tube was vortexed for 30 seconds. The samples were stored on ice for 5 minutes and centrifuged at 12,000 x *g* for 15 minutes at 4°C. The top aqueous phase was transferred to a clean tube, 0.5 ml isopropanol was added and the sample was incubated for 10 minutes at room temperature. Samples were again centrifuged at 12,000 x *g* for 15 minutes at 4°C. The pellet was washed twice with 1 ml 75% ethanol and centrifuged again for 5 minutes at 4°C. The pellet was air-dried and resuspended in 50 µl water.

The RNA samples were then treated with DNase I according to manufacturer's instructions (Ambion DNase kit AM1906, Thermo Fischer Scientific) by adding 5 µl 10x DNase I buffer and 1 µl rDNase I to the RNA. Samples were incubated at 37°C for 20 minutes, and then mixed with 5 µl DNase inactivation reagent to stop the reaction. RNA was incubated for 2 minutes at room temperature and centrifuged at 10,000 x *g* for 1.5 minutes at 4°C. The supernatant containing the RNA was moved to a fresh tube.

The RNA was precipitated with 5 µl (0.1 of the volume) 3 M sodium acetate and 100 µl (2 volumes) 100% ethanol and frozen at -20°C (short term, usually overnight) or -80°C (long term, more than one day). Finally, the sample was centrifuged to pellet the RNA which was then washed with 75% ethanol. The pellet was air-dried and again resuspended in 50 µl water.

2.2.2 RNA concentration and quality check

The concentration of RNA was measured using a NanoDrop spectrophotometer ND-1000 (Nanodrop technologies, Wilmington, DE, USA) and quality was assessed using the Agilent RNA ScreenTape System (Agilent Technologies, Santa Clara, CA, USA) according to the manufacturer's instructions. The reagents were allowed to warm to room temperature 30 minutes before usage. 1 µl RNA was mixed with 5 µl sample buffer by vortex at 2000 rpm for 1 minute at room temperature. Afterwards, the samples were centrifuged and incubated at 72°C for 3 minutes and at 4°C for 2 minutes. The samples were then run on the TapeStation 2200 apparatus. The RIN^e value (RNA compactness value) measured the integrity of the RNA by comparing the ratio of 18S and 28S rRNA (Schroeder *et al*., 2006). Only RNA samples with values around or above 8 (out of 10) were used for CAGE sequencing.

2.2.3 cDNA synthesis

The Superscript III Reverse Transcriptase protocol was adapted from manufacturer's instructions as follows: 500 ng DNase I treated RNA was used, together with 2 µl random primers (50 ng/µl, Invitrogen, Paisley, UK) and 1 µl dNTPs (10 mM, Invitrogen), with water to top up the reaction to 13 µl. The RNA with random primers was denatured at 65°C for 5 minutes and then cooled at 4°C for at least 1 minute. Afterwards, 4 µl 5x first strand buffer (Invitrogen), 1 µl 0.1 M DTT (Invitrogen, Canada), 1 µl RNAsin Plus (Promega, Madison, WI, USA) and 1 µl Superscript III Reverse Transcriptase (RT; Invitrogen) were added to the mixture. RT negative control was prepared using RNAse-free water instead of RT. The cDNA synthesis reaction was incubated at 25°C for 5 minutes, 50°C for 60 minutes, and finally 70°C for 15 minutes to stop the reaction.

2.2.4 qRT-PCR

Before Quantitative Polymerase Chain Reaction (qPCR), the cDNA was diluted 1:1 with 20 µl water. The concentration of cDNA, if starting from 500 ng RNA, was estimated to be 12.5 ng/µl.

To establish standard curves for different primers, cDNA from THP-1 RNA was diluted three times. The first point of the standard curve was undiluted cDNA (estimated 12.5 ng/µl), then 1:1 (estimated 6.25 ng/µl), then 1:4 (estimated 3.125 ng/µl) and 1:8 (estimated 1.5625 ng/µl).

qPCR was used to assess the levels of gene expression and validate the results from CAGE. qPCR was carried out using standard Roche protocols for SYBR Green 1 Master Mix with Light Cycler 480 96-well white plates (Roche, Mannheim, Germany). For 20 µl reaction, 10 µl SYBR Green 2x (Roche) was mixed with 7 µl PCR-grade H_2O (Roche) and 1 µl forward and reverse primers (10mM each, see Table 2-1). 2 µl cDNA was then added to the well (25 ng cDNA in total). RT negative and no cDNA (just H_2O) controls were used with every run for every primer set master mix.

The settings for all qPCR analyses (both quantification and melting curves for primers) were as follows: pre-incubation was carried out at 95°C for 5 minutes (ramp rate 4.40°C/second), then amplification steps were repeated for 45 cycles. Amplification steps were as follows: 95°C for 10 seconds (ramp rate 4.40°C/second), 60°C for 15 seconds (ramp rate 2.20°C/second) and 70°C for 30 seconds (ramp rate 4.40°C/second). Afterwards, the melting curve for primers was measured by incubating at 95°C for 5 seconds (ramp rate 4.40°C/second), 65°C for 1 minute (ramp rate 2.20°C/second) and then the temperature was increased to 97°C by 0.11C°/second. At the end, the plate was cooled for 30 seconds at 40°C.

All qRT-PCR analysis was carried out using the Advanced Quantification setting of the Light Cycler 480 Roche software. ΔCt was calculated with previously established values of primer efficiencies from standard curves (calculated using the same software, by using Abs Quant/2nd Derivative Max setting).

Gene Target	Sequence 5' - 3'	Slope value
ACTB_F	ATTGCCGACAGGATGCAGAA	-3.398
ACTB_R	GCTGATCCACATCTGCTGGAA	
GAPDH	Qiagen	-3.498
MYB_ex5-6_F	ACAGATGGGCAGAAATCGCA	-3.483
MYB_ex5-6_R	GCAGGGAGTTGAGCTGTAGG	
CD14_ex2-3_F	TAAACTGTCAGAGGCAGCCG	-3.345
CD14_ex2-3_R	TCGTCCAGCTCACAAGGTTC	
BAP1_ex3-4_F	GGGTGCAAGTGGAGGAGATC	-3.316
BAP1_ex3-4_R	TAGAGACCTTTCGCCGGGA	
KDM6B_ex3-4_F	GGAGGCCACACGCTGCTAC	-3.442
KDM6B_ex3-4_R	GCCAGTATGAAAGTTCCAGAGCTG	
USP12_ex6-7_F	GCAGCAAACAGGAAGCACAC	-3.366
USP12_ex6-7_R	AGGTCGTACATTCTGTCTGGA	
USP16_ex4-5_F	TGCCAAGACTGTAAGACTGACA	-3.543
USP16_ex4-5_R	TGGCGTCAGATAGTGCTTCA	
USP16_ex15-16_F	AGTATGCACACGGAGACAGT	-3.528
USP16_ex15-16_R	AGAGTAAGAACAGGAGGAGCA	
USP16_ex17-18_F	CCTACGCAAAGTTAACAAACACA	-3.014
USP16_ex17-18_R	GTGTAATGCCCCGACCTCAT	
UTY_F	GGAACAACTGCGAGCAAATAG	-3.270
UTY_R	TGCTGCATTAAGACAAACTGACT	

Table 2-1 qPCR primers and efficiencies. The slope value for each primer set was calculated from standard curves using LightCycler 480 Roche software

2.2.5 Primer validation and optimisation

Two sets of primers for two different housekeeping genes were used in this study. The first one was designed for the human beta actin gene (*ACTN*) by Maess *et al*. (2010); the second for *GAPDH* was purchased from Qiagen (QuantiTect Primer Assay, QT0112646, Hilden, Germany). The rest of the primers were designed to span an intron, to have melting temperature (T_m) of 60°C and to generate a cDNA product of approximately 200bp using Primer3 programme (http://primer3.ut.ee/). The ideal slope value from standard curves is around -3.345 when the primer efficiency is 2, but values from -3.0 to -3.5 were considered acceptable (Table 2-1).

2.2.6 CAGE (Cap Analysis of Gene Expression)

Each CAGE library was made from 7 or 8 samples of 5 µg RNA each, following a protocol adapted from Takahashi *et al*. (2012). All the primer sequences were taken from Takahashi *et al*. (2012). A list of oligonucleotide sequences is located in Appendix 8.1 and they were all synthesized by IDT (Coralville, IA, USA). They were prepared with help of Lucas Lefèvre (The Roslin Institute) according to the Takahashi *et al* (2012) protocol.

For each sample, 5 µg RNA was concentrated to 5.3 µl using a vacuum concentrator (Speed Vac, RC 10.22, Jouan SA, Saint-Herblain, France). This was mixed with 2.2 µl 210 µM (2 µg/µl) RT-N15-EcoP primer (7.5 µl total) and incubated at 65°C for 5 minutes then cooled on ice immediately. This RNA-primer mix was reverse transcribed to produce a cDNA/RNA duplex by adding the components detailed in Table 2-2 and incubating at 25°C for 30 seconds, 42°C for 30 minutes, 50°C for 10 minutes, 56°C for 10 minutes, and 60°C for 10 minutes.

Component	**Volume (µl)**	**Final concentration**
PrimeScript buffer (TAKARA Clontech, Otsu, Japan)	7.5	1x
dNTPs (10 mM each, Invitrogen)	1.87	0.5 mM each
Sorbitol (3.3 M)/(0.66 M) trehalose mix solution (kindly provided by Lucas Lefèvre)	7.5	0.66 M/0.132 M
PrimeScript™ reverse transcriptase (200 U/µl; TAKARA Clontech)	3.75	750 U
Water	9.38	-
Total Volume	30	-

Table 2-2 CAGE: Reverse transcription of RNA

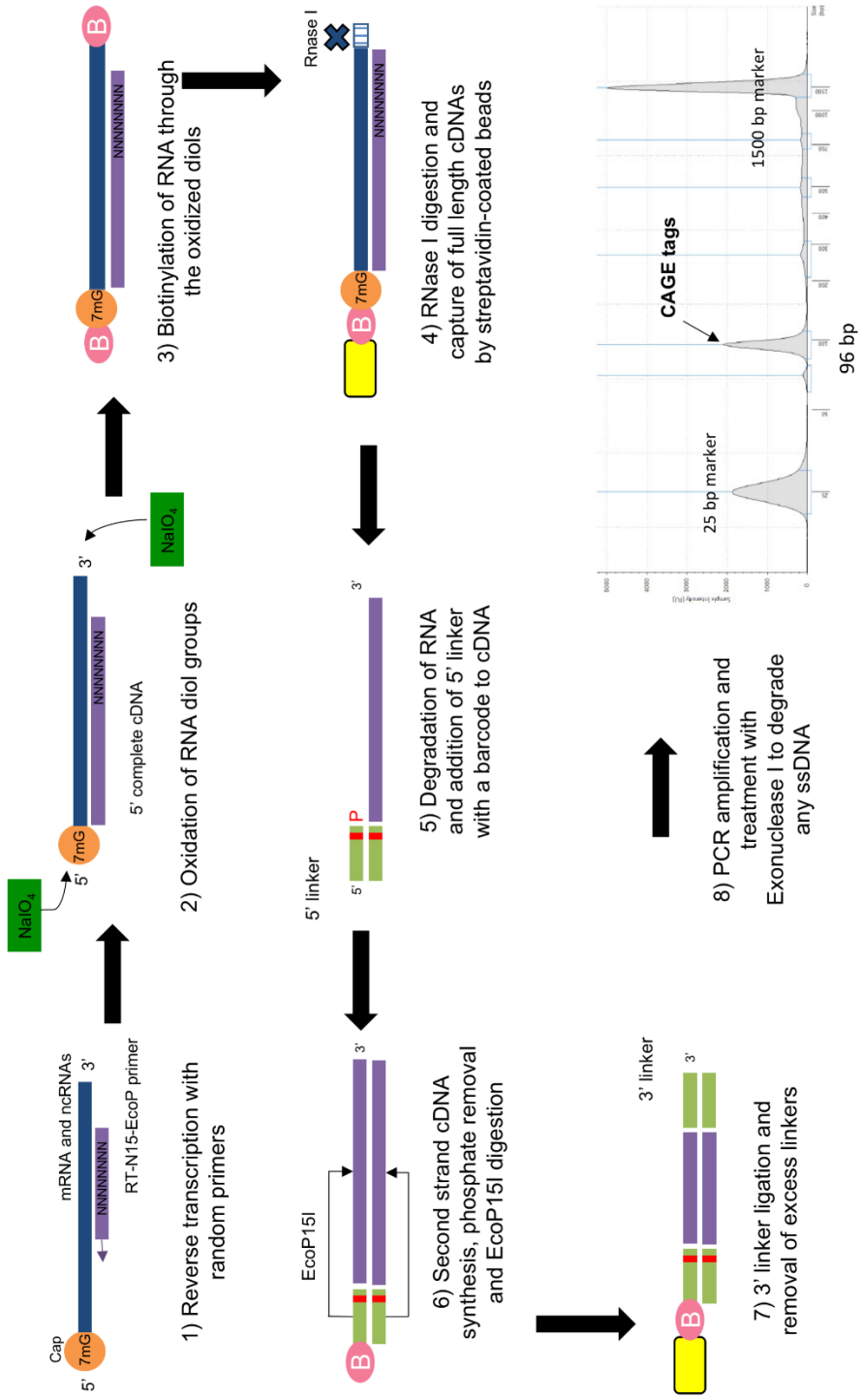

Figure 2-1 CAGE library preparation. Total RNA from cells is reverse transcribed with random primers with an EcoPI site (1). Next, NaIO4 oxidises diol groups (2) to open them so the biotin can attach to the cap (3). The ssRNA on the ends is digested and the biotin-labelled cap with cDNA/RNA hybrid is captured by streptavidin beads (4). The cap is deattached, the RNA is digested, and 5' linker with a barcode and EcoP15I site is attached (5). Second strand of cDNA synthesized and prior to endonuclease digestion, phosphates are removed by Antarctic phosphatase to prevent nonspecific binding of 3'linker in the further step (6). EcoP15I cleaves 27 nt downstream, leaving only 5'linker with 27 nt of cDNA, to which 3' end another linker is attached. Excess linkers need to be removed by vigorous washing steps (7), otherwise PCR amplification picks up high levels or linker DNA (8). Lowest possible number of PCR cycles is chosen (from 9 to 11, usually 10) to reduce PCR bias (8).

To purify these cDNA/RNA hybrids, Agencourt® RNAClean® XP beads (ratio 1.8 to sample; Beckman Coulter, Indianapolis, IN, USA) were used according to manufacturer's instructions. The mix of beads and samples were incubated at room temperature for 30 minutes, mixing 10 times every 10 minutes. A 96-well magnetic stand (DynaMag™ 96-slide, Life Technologies, USA) was used to separate the beads from the supernatant and to perform two washes with 150 µl 70% ethanol. To elute, 40 µl water was heated to 37°C and mixed thoroughly with the beads. This mix was incubated at 37°C for a minimum of 5 minutes. The magnetic stand was used to separate the beads from the eluate containing the cDNA/RNA hybrids, which were transferred to a new tube.

The 3' end of the RNA and the 7-methylguanylate cap on the 5' end were diol oxidated using 250 mM $NaIO_4$ (0.053 g $NaIO_4$ in 1 ml water; Sigma-Aldrich) according to Table 2-3. The reagents were mixed and incubated on ice for 45 minutes in darkness.

Component	**Volume (µl)**	**Final concentration**
RNA-cDNA hybrid	40	-
Sodium acetate (1 M, pH 4.5, sterile-filtered)	2	45.7 mM
$NaIO_4$ (250 mM)	2	11 mM
Total volume	44	-

Table 2-3 CAGE: Diol oxidation

Following incubation, 2 µl 40% (wt/vol) glycerol (Fischer Scientific, Waltham, MA, USA) was added to stop the oxidation reaction. 14 µl 1 M Tris-HCl (pH 8.5, sterile-filtered) was then added to bring the pH above 5.6 (total volume 60 µl). The cDNA/RNA hybrids were purified again with Agencourt® RNAClean® XP beads as above.

Biotin was added to the RNA diols using 15mM biotin hydrazide (long arm; 0.0038g in 675µl water; Vector Laboratories, Burlingame, CA, USA) by mixing the reagents in Table 2-4 and incubating at 23°C overnight.

Component	Volume (µl)	Final concentration
Purified oxidated cDNA/RNA hybrids	40	-
Sodium citrate (1M, pH 6.0, sterile-filtered)	4	70mM
Biotin hydrazide (long arm, 15mM)	13.5	3.5mM
Total volume	57.5	-

Table 2-4 CAGE: Biotinylation reaction

The next day, all single stranded RNA was degraded by RNase ONE ribonuclease (Promega), mixing the reagents in Table *2-5* and incubating at 37°C for 30 minutes followed by 65°C for 5 minutes. This process left only cDNA/RNA hybrids containing a biotinylated 5' cap which were purified with Agencourt® RNAClean® XP beads as above.

Component	Volume (µl)	Final concentration
Biotinylation reaction	57.5	-
Tris-HCl (1 M, pH 8.5, sterile-filtered)	6	86 mM
EDTA (0.5 M, pH 8.0, Ambion)	1	7.2 mM
RNase ONE ribonuclease (10 U/µl, Promega)	5	50 U
Total volume	69.5	-

Table 2-5 CAGE: RNase digestion

MPG® Streptavidin beads (Takara Clontech) were prepared by mixing 100 µl beads and 1.5 µl tRNA (20 µg/µl; prepared by Lucas Lefèvre; transfer ribonucleic acid from *Escherichia coli*, Sigma-Aldrich) per sample, incubating at 4° C for at least an hour by continuous mixing. Beads were then separated from the supernatant by magnetic stand (SureBeads™ BioRad, USA) and washed twice with 50 µl per sample wash buffer 1 (45 ml 5 M NaCl and 5 ml 0.5 M EDTA pH 8; all wash buffers were prepared with help of Lucas Lefèvre). The beads were the suspended in 80 µl wash buffer 1 for each sample.

The tRNA-coated streptavidin beads were used to trap the biotinylated RNA/cDNA hybrids. Each purified sample was mixed with 80 µl prepared beads and incubated at room temperature for 30 minutes, mixing 10 times every 5 minutes. The beads were separated from the supernatant using 96-well magnetic stand and washed 6 times total with various buffers: once with 150 µl wash buffer 1, once with 150 µl wash buffer 2 (3 ml 5 M NaCl, 100 µl 0.5 M EDTA pH 8, 46.9 ml water), twice with 150 µl wash buffer 3 (1 ml 1M Tris-HCl pH 8.5, 100 µl 0.5M EDTA pH 8, 25 ml 1 M sodium acetate pH 6.1, 2 ml 10% (w/v) SDS, 21.9 ml water), and twice again with 150 µl wash buffer 4 (500 µl 1 M Tris-HCl pH 8.5, 100 µl 0.5 M EDTA pH 8, 25 ml 1 M sodium acetate pH 6.1, 24.4 ml water).

RNA was degraded and the cDNA released from the beads by the addition of 60 µl 50 mM NaOH solution and incubation at room temperature for 10 minutes, pipetting every 2-3 minutes. The solution was then buffered to a lower pH by adding 12 µl ice cold 1 M Tris-HCl pH 7.0 (Trizma® hydrochloride, Sigma-Aldrich).

The single stranded cDNA was purified with Agencourt AMPure® XP beads (1.8x to sample ratio; Beckman Coulter) according to manufacturer's instructions. The beads with the samples were incubated at room temperature for 30 minutes, mixing 10 times every 10 minutes. The 96-well magnetic stand was used to separate the beads from the supernatant and to wash twice with 150 µl 70% ethanol. To elute, 40 µl of preheated 37°C water was mixed with the beads thoroughly and incubated at 37°C for a minimum of 5 minutes. The magnetic stand was then used to separate the beads from the eluent containing the cDNA which was transferred to a new tube.

The cDNA was concentrated to 4 µl at room temperature using a centrifugal vacuum concentrator as described previously. The cDNA was incubated at 65°C for 5 minutes and cooled on ice for 2 minutes. For each sample, 1.5 µl 5' linker was added to a separate tube and incubated at 37°C for 5 minutes then cooled on ice for a minimum of 2 minutes. 1 µl of the linker and mixed with 10 µl of DNA Ligation Kit Mighty Mix (Takara Clontech) was added to each cDNA sample for incubation at 16°C overnight.

The next day, the samples were purified twice using Agencourt AMPure® XP beads as above. In the first purification step, up to 4 samples could be pooled in 1 tube (60 µl) and 10 µl of water was added. In the second purification step, up to 8 samples total could be pooled. Two tubes (4 samples each) were incubated with beads, but at the first wash step, 150 µl 70% ethanol was used to wash both wells and the contents were mixed. The samples were eluted in 30.5 µl preheated 37°C water.

The second stand was synthesized using La Taq Polymerase, by adding the reagents listed in Table 2-6 (including a biotinylated primer) and incubating at 94°C for 3 minutes, 42°C for 5 minutes, 68°C for 20 minutes and 62°C for 2 minutes.

Component	Volume (µl)	Final concentration
10x La Taq buffer (TAKARA Clontech)	5	1x
$MgCl_2$ (25 mM, TAKARA Clontech)	5	2.5 mM
dNTPs (2.5 mM, TAKARA Clontech)	8	0.4 mM each
2nd SOL primer (200 ng/µl)	1	2.4 µM
La Taq Polymerase (5 U/µl, TAKARA Clontech)	0.5	2.5 U
Total volume	19.5	

Table 2-6 CAGE: Second strand synthesis

A phosphate group was removed from the 5' lower linker of the cDNA by adding 6 µl 10x Antarctic Phosphatase Reaction Buffer and 4 µl Antarctic Phosphatase (5 U/µl). The mixture was then incubated at 37°C for 60 minutes followed by 65°C for 5 minutes. The cDNA was then purified with Agencourt AMPure® XP beads as above.

The cDNA was digested by *Eco*P15I restriction endonuclease which binds to a recognition site in the second strand primer and cuts 27 nucleotides into the cDNA. The reagents in Table 2-7 were combined, 10 µl master mix was added to 30 µl sample. This mix was incubated for 3 hours at 37°C and then put on ice for at least 2 minutes.

Component	Master Mix (enough for 10 tubes, µl)	Final concentration for 1 sample
Water	11	
10x NEB buffer 3.1 (NEB, Hitchin, UK)	40	1x
100x BSA (NEB)	4	10x
10x ATP (10 mM, NEB)	40	1 mM
Sinefungin (10 mM, kindly provided by Lucas Lefèvre)	4	0.1 mM
***Eco*P15I** (10 U/µl, NEB)	1	1 U

Table 2-7 CAGE: EcoP15I restriction digest

After 3 hours, 1 µl $MgCl_2$ 0.4 M was added and the mix incubated at 65°C for 20 minutes then 4°C for 2 minutes. The 3' linker was ligated to the cDNA by mixing 39 µl of the mix listed in Table 2-8 and incubating at 16°C overnight (14-16 hours).

Component	Volume (µl)	Final amount
5x 3'linker ligation buffer (see **Table 2-9**)	16	
3' linker (100 ng/µl)	1	100 ng
T4 DNA ligase (400 U/µl, NEB)	3	1200 U
Water	19	
Total volume	39	

Table 2-8 CAGE: 3' linker ligation

Component	Volume (µl)
Tris-HCl (1 M, pH 7.0, Trizma® hydrochloride, Sigma-Aldrich)	50
ATP (10 mM, NEB)	100
BSA (10 mg/ml, NEB)	0.5
Water	49.5

Table 2-9 CAGE: 5x 3' ligation buffer

The next day, 15 µl MPG® Streptavidin beads (Takara Clontech) were incubated on ice with 1.5 µl tRNA (20 µg/µl) for 30 minutes, mixing them every 5 minutes. The supernatant was separated from the beads using magnetic stand, and the beads were washed twice with 75 µl of the wash buffer 1. The beads were then suspended in 37.5 µl wash buffer 1.

The samples were further purified from the 3' linker using the biotin from the second strand primer. 25 µl prepared MPG beads with tRNA were added to 80 µl sample from the previous day and incubated at room temperature for 30 minutes, mixing every 5 minutes. The beads were separated from the supernatant using 96-well magnetic stand and washed 7 times total with various buffers: once with 150 µl Wash buffer 1, once with 150 µl Wash buffer 2, twice with 150 µl Wash buffer 3, twice again with 150 µl Wash buffer 4, and once with 50 µl water. The beads were then suspended in 20 µl water and used as the PCR template below.

A PCR reaction was carried out to find out how many cycles were needed to maximise the 96 bp product and minimise the amount of linkers (around 70 bp length). One reaction mix was prepared as in Table 2-10, and 2 µl of the beads were added to the mix.

Component	**Volume (µl)**	**Final concentration**
Water	32.5	
5xHF buffer (Thermo Scientific)	10	1x
dNTPs (2.5 mM, Invitrogen)	4	0.2 mM
Forward primer (100 µM)	0.5	1 µM
Reverse primer (100 µM)	0.5	1 µM
Phusion Polymerase (2 U/µl, Thermo Scientific or NEB)	0.5	1 U
Total	48 µl	

Table 2-10 CAGE: final PCR reaction

The reaction was incubated in the PCR machine at 98°C for 30 seconds, then 20 cycles of 98°C for 10 seconds and 60°C for 10 seconds. The machine was paused after 8, 10, 12, 14 and 16 cycles to take 4 µl out of the tube quickly.

The optimum PCR cycle number was determined by D1000 ScreenTape (Agilent Technologies). The ladder was prepared mixing 3 µl D1000 Sample Buffer with 1 µl D1000 ladder, and the samples were prepared by mixing 3 µl D1000 sample buffer with 1 µl of the cDNA sample that had been removed at different cycle numbers. The samples were vortexed at 2000 rpm for 1 minute, centrifuged, and then analysed on the Agilent 2200 TapeStation system. The number of cycles chosen was from 9 to 11, usually 10.

Afterwards, the PCR with the optimal number of cycles was repeated in 6 replicates as shown in Table 2-10 to generate more product. Three reactions were pooled (150 µl in total) and 1 µl Exonuclease I (20 U/µl, NEB) was added to each of the 150 µl PCR solutions. These two samples were incubated at 37°C for 30 minutes, then the Exonuclease I- treated CAGE tags were purified using Minelute PCR purification kit (Qiagen), following the manufacturer's instructions. CAGE tags were then eluted in 12 µl EB buffer each and the two samples pooled (final volume 24 µl). Quantity and quality of the tags was measured using D1000 ScreenTape as described previously.

The libraries were sent to Edinburgh Genomics (Edinburgh, UK) for sequencing on an Illumina HiSeq 2500 machine (Illumina) in high throughput mode. One library made out of 7 or 8 pooled samples was sequenced on 1 lane, with custom sequencing primer and inline barcodes.

2.3 Perturbing gene expression using siRNA

ON-TARGETplus SMARTpool siRNAs and ON-TARGETplus Control non-targetting pool (D-001810-10-05) (GE Healthcare Dharmacon, Lafayette, CO, USA) were used. Each is a pool of 4 different siRNAs targeting the same gene (Table 2-11) to ensure maximal silencing. Non-targeting control (NT) from the same manufacturer was also used. 5 nmol were diluted in 250 µl RNase-free water (stock concentration

20 µM). Final concentration of siRNA when added to cells was 50 nM (6.25 µl 20 µM siRNA for a 6 well plate with 2.5 ml total volume).

A 6-well plate was seeded with $5x10^5$ THP-1 cells per well in 2 ml of THP-1 medium. 6.25 µl 20 µM siRNA was diluted in 250 µl OPTI-MEM I reduced serum medium (Gibco, UK), mixed gently and incubated at room temperature for 5 minutes. Meanwhile, 5 µl Lipofectamine 2000 (Invitrogen, USA) was diluted in 250 µl OPTI-MEM I reduced serum medium, mixed gently and incubated at room temperature for 5 minutes. The diluted siRNA and diluted Lipofectamine 2000 were combined (final volume 500 µl), mixed gently and incubated at room temperature for 20 minutes. The mixture was then added to 2 ml of cells before being replaced after 4-6 hours.

The efficiency of the siRNA pools was tested by a preliminary experiment, in which the siRNA pool was added at day -1, the differentiating agent (PMA) was added to final concentration of 30 ng/ml at day 0, and samples were collected at day 1, 2, 3 and 7 to examine the level of the target mRNA.

Target gene	Target sequences	Location of target sequence
BAP1	GAUGAUACGUCCGUGAUUG	Exon 4
	CAACCGUGCUGUCCGUGAU	Exon 13
	CCAUCAACGUCUUGGCUGA	Exon 13
	GAGCAAAGGAUAUGCGAUU	Exons 5+6
KDM6B	CUGUACAGACCCUCGAAAU	Exon 13
	CAACCGUGCUGUCCGUGAU	Exon 14
	AAAGGGAAGUUUCGAGAGU	Exon 12
	UGGGAGACCAUCAGCGCUU	Exon 18
USP12	UGGAUCAACUUCAUCGAUA	Exon 7
	UAGCAGAUCUCUUCCAUAG	Exon 3
	AGAAGUUCAUCACAAGAUU	Exon 3 fw
	GAAGCACACAAACGGAUGA	Exon 6+7
USP16	CGAAUAAACUGCUUUGUGA	Exon 15
	GUAAGAAUGUUGCAGAAGA	Exon 17
	GGACCAAAGGCAAAUAUAA	Exon 15
	GGCAUUAACAGAACCAUUA	Exon 8+9
UTY	GUAUUAAGAAGGCGAAUG	Exon 18
	GCUAGAAGCAAACGUUGUA	Exon 12
	CCAUGGAGAUUCACCAAAU	Exon 18
	GCUAGGCAGUAAUUGUAUA	Exon 17

Table 2-11 siRNA pools used in this study

Following the preliminary siRNA experiments, a larger experiment was set up to harvest RNA for CAGE. Each sample was collected by pooling the RNA from 3 wells of a 6-well. The siRNA was administered on day -1, PMA (30 ng/ml) was added on day 0 and samples were harvested at day 0 (24 hours after siRNA treatment, immediately prior to PMA treatment) and day 1 (48 hours after siRNA treatment, 24 hours after PMA treatment).

2.4 Cell culture techniques

2.4.1 MTT assay

A 96-well plate was seeded with $2x10^4$ THP-1 cells in 100 µl of media per well. 10 µl MTT (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide, 5 mg/ml, Sigma Aldrich) was added to each well and incubated for 3 hours at 37°C in 5% CO_2. Then 100 µl solubilisation solution (89% (v/v) Isopropanol, 10% (v/v) Triton 100x, 1% (v/v) HCl) was added and left at 37°C 5% CO_2 overnight. The following day, substrate conversion was determined via the optical densities which were measured using a plate reader at 570nm.

2.4.2 Propidium iodide staining during the cell cycle

A pellet of $1x10^6$ THP-1 cells was produced by centrifugation at 400 x *g* for 5 minutes, this was resuspended in 1 ml ice-cold 70% ethanol and left at 4°C for at least 24 hours (up to 2 months). On the day of flow cytometry, the cells were centrifuged at 16,000 x *g* for 5 minutes, washed twice with PBS (provided by the Roslin Institue core services), and 300 µl Cell Cycle Staining Solution was added to the pellet (See Table 2-12). Samples were incubated in the dark for 1-2 hours and then 300 µl of PBS was added prior to flow cytometry. Samples were analysed on a Fortessa cytometer (BD Biosciences, Franklin Lakes, NJ, USA) using a 610nm laser. Results were analysed using FlowJo software version 10.0.8r1 (Flow Jo LLC, Ashland, OR, USA) by gating single viable cells by excluding doublets. Final window showed the histogram of the area of Propidium iodide channel, which was examined using the Cell cycle option of FlowJo software by choosing the Watson (pragmatic) type of analysis (Watson *et al*., 1987).

	Final concentration	Volume for 1ml total
Sodium citrate (0.1 M, kindly provided by Dr Mark Barnett, The Roslin Institute)	38 mM	380 µl
RNase A from Pancreas (1 mg/ml, Sigma-Aldrich)	10 µg/ml	10 µl
Propidium iodide (1 mg/ml, Sigma-Aldrich)	68 µM	45 µl
Water		565 µl

Table 2-12 Cell Cycle Staining Solution for propidium iodide staining

2.5 Bioinformatic analysis

2.5.1 CAGE data quality check, trimming and genome mapping

The reads were at first split according to their barcodes using perl script split_library.pl (Appendix 8.2) which was kindly provided by Dr Lel Eory (The Roslin Institute). The quality and length of the reads was assessed using FastQC programme (Andrews, 2010) and summarized using MultiQC programme (Ewels *et al*., 2016). The reads were then trimmed from nested 5' and non-nested 3' adapters using Cutadapt (Martin, 2011). Single end reads were then aligned to hg38 (downloaded in September 2016 from UCSC: http://hgdownload.cse.ucsc.edu/goldenPath/hg38/bigZips/, release Dec 2013) with Burrows Wheeler Alignment (BWA) (Li and Durbin, 2009) using commands aln and samse. The files were then converted to bam file using Samtools (Li *et al*., 2009).

2.5.2 Normalisation of CAGE data

The expression levels (number of reads) were normalised using CAGEr package (Haberle *et al*., 2015) using getCTSS and normalizeTagCount commands with options "powerLaw" and a specific alpha value per million tags, which is provided in each chapter. The final expression value for each CTSS (CAGE transcription start site) was therefore provided as TPM (Tags per million).

2.5.3 Visualisation of CAGE data

The normalised data was then formatted into OSCtable (details found at https://zenbu-wiki.gsc.riken.jp/zenbu/wiki/index.php/OSCtable), providing chromosome, end, start and strand coordinates. A header was included as specified in the Zenbu website (see above) and uploaded into Zenbu (found at http://fantom.gsc.riken.jp/zenbu/) (Severin *et al*., 2014). Instructions to access this data are in Appendix 8.10. Normalised promoter TPM values were taken from Zenbu browser by manually highlighting the promoter region and copy pasting the data into an Excel spreadsheet.

2.5.4 Clustering CTSS

CTSS were clustered in each sample based on their distance apart using distclu option in CAGEr package (Haberle *et al*., 2015). The settings were as follows, the minimum CTSS TPM value was 1, the distance between CTSS was maximum of 20 bp and singletons (single CTSS not neighbouring any other CTSS) were not removed. To be able to compare the clustered CTSS across different samples, the CTSS range values needed to be aggregated. In this step, only the CTSS clusters with expression in at least one sample of higher than 5 TPM were kept, the others were filtered. The maximum distance between CTSS was kept at 100 bp. These commands created a single matrix file with cluster coordinates (start, end, strand) and normalised TPM values for the aggregated clusters.

2.5.5 Miru network analysis

The normalised and aggregated clusters of CTSS were annotated for their gene names by Dr Stephen Bush (The Roslin Institute). The gene names together with expression values for each sample were uploaded into Miru software (Kajeka Limited, Midlothian, UK; https://kajeka.com/download-miru/). Sample to sample correlation was created by transposing the data in pre-processing. The correlation graph was created at correlation coefficients (R) > 0.75 (R values for each analysis are provided in individual chapters) using Fast Multiple Multilevel Method (FMMM) format. Gene network analysis created a graph by visualising each node as one aggregated cluster

of CTSS. The analysis was done by clustering using a Markov Cluster Algorithm (MCL) at inflation value of 2.0, which created clusters of correlated CTSS.

2.5.6 DAVID GO analysis

The cluster gene lists compiled by Miru were individually entered into the upload tool DAVID (Database for annotation, visualisation and integrated discovery), which was used to decipher the functional GO (gene ontology) annotation (Huang da *et al*., 2009). OFFICIAL_GENE_SYMBOL was chosen as an identifier for the gene list. The background was selected as *Homo sapiens*, other settings were left as default.

2.5.7 EdgeR

Differential expression between groups of genes was analysed using EdgeR package (Robinson *et al*., 2010, McCarthy *et al*., 2012). As EdgeR normalises the data, data from CAGEr were left raw (using NormalisedTagCount with option method = "none"). CTSS were clustered and aggregated the same way as in section 2.5.4, with raw data options. Aggregated clusters matrix was exported into an R object. The two time points compared were substracted from the object, and they were uploaded into EdgeR by DGEList command. Differential expression was calculated with subsequent commands calcNormFactors(), estimateCommonDisp(), estimateTagwiseDisp(), and exactTest(). The cut off p-value was 0.01 with command summary(decideTestsDGE()). The top hits were displayed with command topTags().

2.5.8 Shifting promoters

The shifting promoters score was calculated by CAGEr package using scoreshift() command, between two groups of time points. The optional settings were to use testKS=TRUE for using Kolomogorov-Smirnov test to calculate the shift and useTpmKS=TRUE for using normalised values of the data. The data was subsequently filtered to only include promoters with score shift > 0.6 and false discovery rate (FDR) threshold lower than 0.01.

2.5.9 Statistical tests

Statistical analyses were done using two sample T-test, for a difference between two groups, using Minitab 17 software (Minitab Inc, State College, PA, USA).

Chapter 3: Characterising monocyte to macrophage differentiation in THP-1 cells

3.1 Introduction

As discussed in section 1.1.3, cells undergoing a state change have altered gene expression patterns, in part mediated by epigenetic changes including histone modifications that alter the chromatin state. In their extensive survey of cell state transition time courses, Arner *et al*. (2015) showed that alteration in enhancer expression precedes alterations in transcription factor expression which are then followed by altered expression of target genes. In particular, the time points considered included closely spaced samples in the early stages of the transition to detect changes in expression of immediate early genes. The original THP-1 time course paper (Suzuki *et al*., 2009) had a much coarser time course, and would have missed many changes that were transient, as well as potentially missing the peak of expression or repression of some genes. However, this earlier study did demonstrate differences in expression of several chromatin modifying enzymes (section 1.4.1). Therefore, more detailed exploration of the time course of THP-1 differentiation was warranted.

The original THP-1 cell line is sensitive to PMA stimulation, but only a subset of the cells undergo differentiation in any one experiment. To increase the sensitivity of the study and ensure that the majority of the cells would undergo differentiation, the line was cloned by limiting dilution and clone 5 was selected for further study (Suzuki *et al*., 2009). This clone of the THP-1 line differentiates relatively homogeneously into adherent cells resembling macrophages, and was also used in the present study, to ensure comparability with the previous work and achieve a high level of differentiation.

CAGE (cap analysis of gene expression) relies on high-throughput sequencing. This allowed a high depth of sequencing (called deepCAGE when applied to CAGE libraries) that can therefore reveal low level and rare transcripts. The FANTOM4 study used the 454 Life Sciences large-scale pyrosequencing approach (Suzuki *et al*., 2009).

Subsequently, for FANTOM5 (Forrest *et al*., 2014, Arner *et al*., 2015) the Helicos Genetic Analysis System for single molecule sequencing (Thompson and Steinmann, 2010) of CAGE libraries (HeliScopeCAGE; (Kanamori-Katayama *et al*., 2011)) was used. These two sequencing methods have been superseded by the whole genome approach implemented by Illumina, which provides high depth of coverage at a relatively low price, and was the sequencing method implemented in the present study.

As shown by the FANTOM5 study (Forrest *et al*., 2014, Andersson *et al*., 2014, Arner *et al*., 2015), deepCAGE can detect a range of changes in transcriptional activity, including enhancer activation or repression and changes in promoter usage, as well as changes in the overall level of expression of a gene. The depth of coverage in the earlier study (Suzuki *et al*., 2009) was not sufficient to identify subtle changes in promoter usage or low expression enhancer sequences. Since the THP-1 time course using the high differentiation rate clone 5 was not repeated for FANTOM5 (Arner *et al*., 2015). Repeating the study would provide valuable additional insight into this popular model of differentiation from monocyte to macrophage.

3.1.1 Aim of this chapter

The aim of this chapter was to characterise the transcriptional changes associated with the monocyte to macrophage differentiation process by stimulating the pro-monocytic THP-1 cells with phorbol ester. A focus was on chromatin modifying enzymes. Various methods were used to assess differentiation state, from morphology, and cell cycle to transcription start site usage and gene expression. A time course of THP-1 stimulation with PMA was carried out, with 18 different time points subjected to CAGE, to show the genome-wide changes in transcription start site usage. This repeat of the FANTOM4 THP-1 time course with more time points showed a detailed map of hour-to-hour expression, with deeper sequencing technology unravelling new previously unknown features of monocyte to macrophage differentiation.

3.2 Materials and methods

3.2.1 THP-1 differentiation assay

THP-1 cells were grown up at a low passage number (passage number 10-13 of clone 5, as mentioned in section 2.1). The day before the start of the differentiation (24 hours), cells were counted by haemocytometer, and around 3-5 $x10^6$ cells were pelleted and resuspended in 10 ml of fresh medium per time point. All THP-1 cells were then differentiated simultaneously by adding 30 ng/ml (48.6 nM) of Phorbol 12-myristate 13-acetate - PMA (P1585, Sigma-Aldrich) in DMSO. Medium was not changed during the experiment, to prevent a serum-specific response during the time course. The cells were plated on a 90mm Sterilin plate (without tissue culture treatment of the plastic), and the cells were scraped off by washing with a blunt-end needle syringe and resuspended in 1 ml of RNABee (AMS Biosciences).

For each experiment, one culture was used to set up two batches of cells. The first batch (annotated with .1 in the results and shown in red in Figure 3-1) was treated with PMA on the morning of the first day. Time points between 0 and 12 hours were taken from this batch. The second batch from the same culture (annotated with .2 and shown in blue in Figure 1-1) was treated with PMA in the evening of the first day, which is the 0 hour time point from this batch was taken at the same time as the 12 hour time point of the first batch. All cells were then incubated overnight. On the morning of the second day, the 24 hour time point of the first batch and the 12 hour time point of the second batch were taken, followed by time points between 12 and 24 hours taken from the second batch. Batches had several overlapping time points (0 hour, 12 hours, 24 hours and 36 hours) to ensure that cells were at similar state of differentiation. This approach was used so that all samples could be processed during working hours.

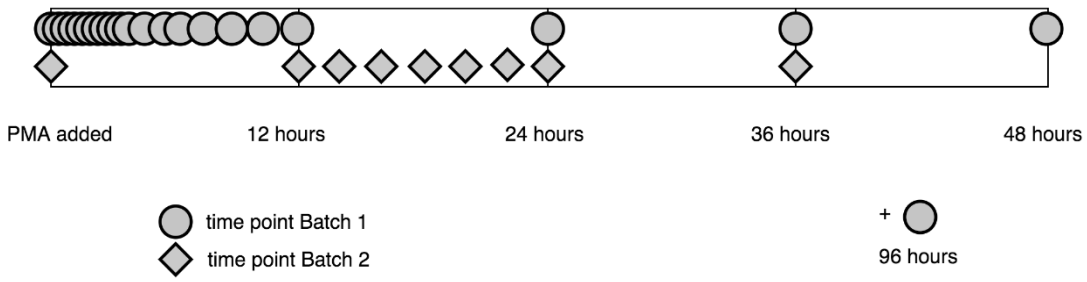

Figure 3-1 Distribution of time points included in the THP-1 time course. The experiments were done in two batches. First batch is a circle; second batch is a diamond.

3.2.2 Cell cycle and morphological analysis of the differentiation

Cell cycle analysis by propidium iodide staining was done as described in section 2.4.2 and cell morphology was analysed using a fluorescent microscope (Zeiss Vert.A.1, Carl Zeiss Limited, Cambridge, UK).

3.2.3 RNA extraction, cDNA production and qRT-PCR

RNA was extracted using the RNABee method as described in section 2.2.1. CAGE sequencing was verified by assessing gene expression levels of important differentiation markers using qRT-PCR as described in section 2.2.4.

3.2.4 CAGE

Samples for CAGE were chosen to have multiple replicates per time point, including the time points used in the FANTOM4 study (Suzuki *et al*., 2009) and covering the important time points for myeloid differentiation missed in that study. After taking into consideration RNA quality and quantity (Appendix 8.3), 18 time points with multiple replicates were chosen for the study in this chapter (Figure 3-2).

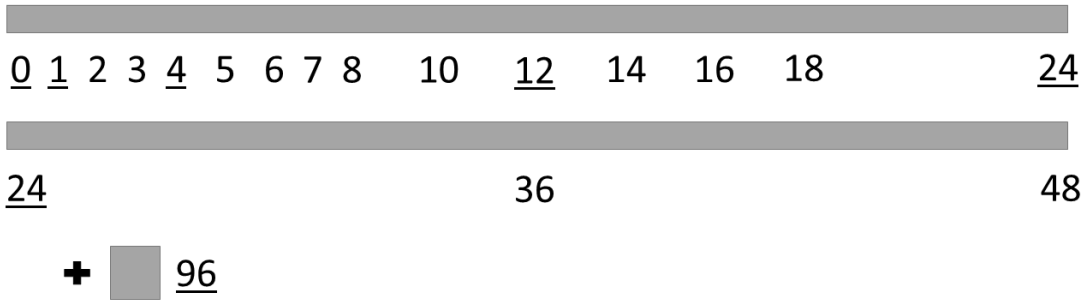

Figure 3-2 Time points chosen for CAGE. Numbers indicate hours. Underlined time points were also in the original FANTOM4 study.

CAGE was performed as described in Chapter 2. Results for 66 samples were uploaded into the CAGEr programme (Haberle *et al*., 2015) (Appendix 8.3). All 66 samples were normalised together using CAGEr using the power law with an alpha of 1.17, with results being in Tags per million (TPM).

3.3 Results

3.3.1 Characterisation of the THP-1 differentiation time course

The differentiation of THP-1 cells (section 3.2.1) was carried out 5 different times to harvest RNA for CAGE sequencing. To ensure that differentiation from monocyte to macrophage phenotype had occurred, changes in morphology and attachment to the culture vessel surface were examined for each time course.

3.3.1.1 Morphological changes

Undifferentiated THP-1 monocyte-like cells grew in suspension, with most cells not touching the bottom surface of the vessel (as seen in the first picture in Figure 3-3). Stimulation with 30 ng/ml of PMA induced the THP-1 cells to start to adhere around 6 to 8 hours, with all cells definitely attached to the bottom of the vessel at 12 hours. After this, the cells started to change their morphology to more macrophage-like cells, with the most dramatic difference observed at 48 or 96 hours. Their shape changed from being round to more variable and the cytoplasm contained more phagocytic vacuoles. There were a few cells where the morphology indicated they were more mature macrophages than the majority at that time point (for example as shown by the white arrow in the 18 hour time point in Figure 3-3). More than 99% cells were adherent and had similar morphology by the end of the time course.

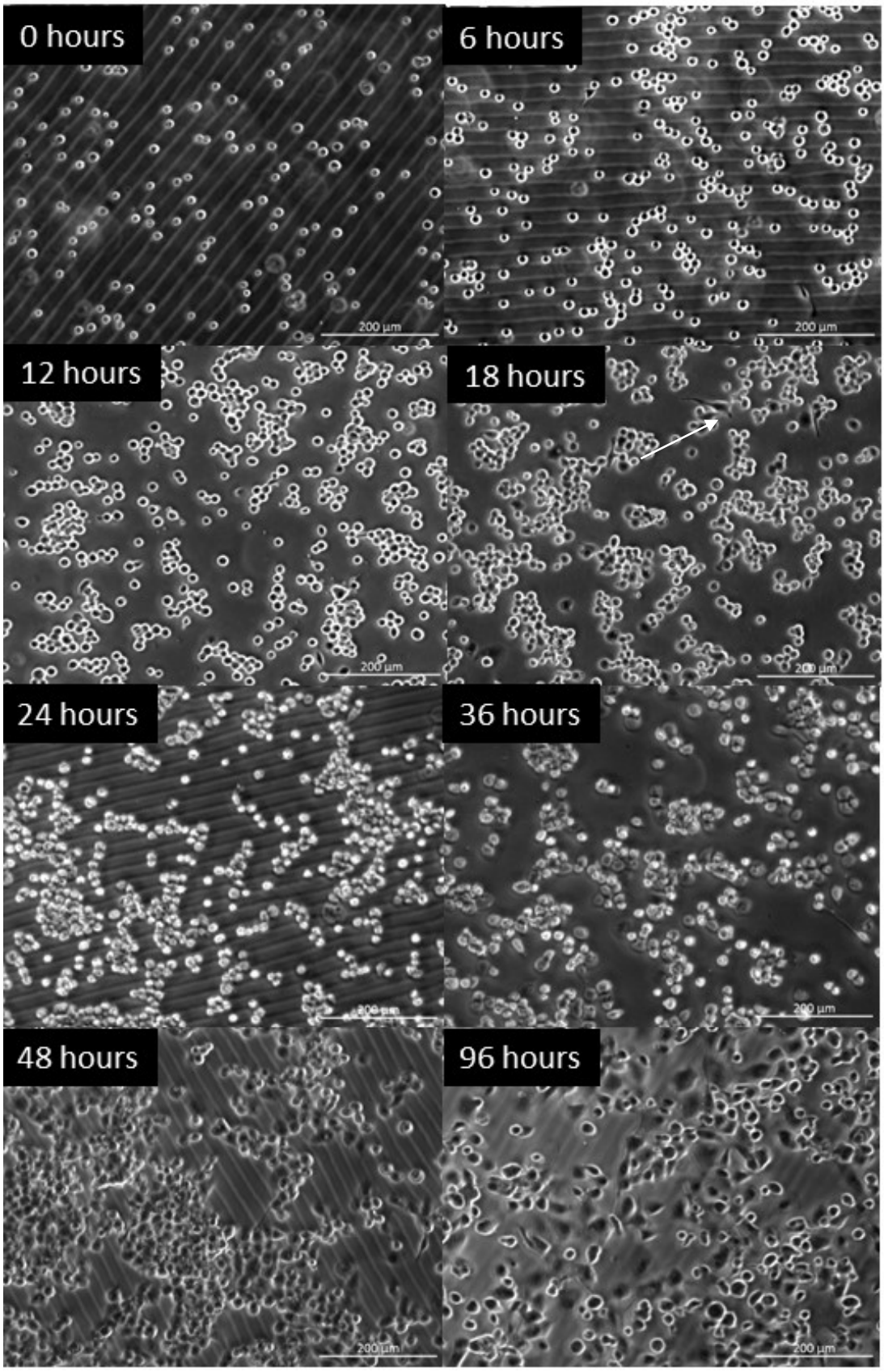

Figure 3-3 Differentiation of THP-1 Clone 5 cells over time after PMA stimulation. Scale bar is 200 µm. White arrow shows macrophage of more mature morphology than the others at the same time point.

3.3.1.2 Cell cycle analysis

As described in Chapter 1, THP-1 cells should cease proliferation upon stimulation with PMA to differentiate into macrophages. To validate that this was the case in these experiments of THP-1 differentiation with PMA, the cell cycle of the THP-1 cells was analysed using propidium iodide staining.

Firstly, the cell cycle of non-synchronised THP-1 monocytes was assessed (Figure 3-4). THP-1 cells were found in all three stages of the cell cycle, with the majority of the cells in G1 (growth) phase, followed by DNA synthesis phase (S-phase) and fewest cells were found in G2 phase, preparing for mitosis.

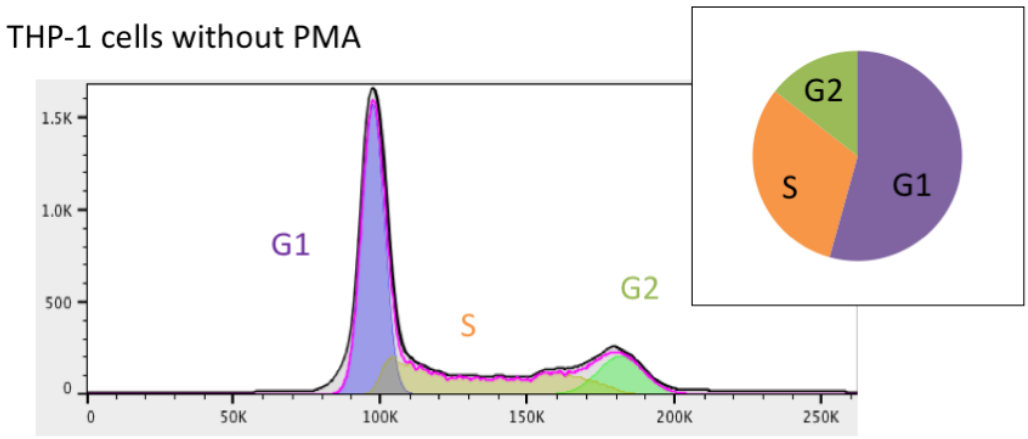

Figure 3-4 Cell cycle analysis of THP-1 cells without PMA stimulation. Highest number of cells were in G1 phase (purple), followed by S-phase (orange) and least amount of cells were in G2 phase (green). Data was analysed using FlowJo, Watson model (pragmatic) (section 2.4.2). Pie chart shows the proportion of cells in each phase. Histogram of number of cells in each phase is taken from FlowJo software, with X axis showing histogram parameter of propidium iodide fluorescence-A (Area), Y-axis shows cell count.

The cell cycle profile of THP-1 cells drastically changed after PMA stimulation (Figure 3-5). After 10 hours, the cell count found in the beginning of S-phase started to diminish (red arrow), cells stopping their cell cycle at G1/S checkpoint. At 24 hours, there were almost no cells left in S-phase, with majority of cells in G1 phase. Cells did not enter the S-phase for the remainder of the experiment and their proportion stayed the same. As shown by the pie charts in Figure 3-5, the proportion of cells in G2 phase decreased from 24 hours to 48 hours, with an increasing proportion found in G1 phase.

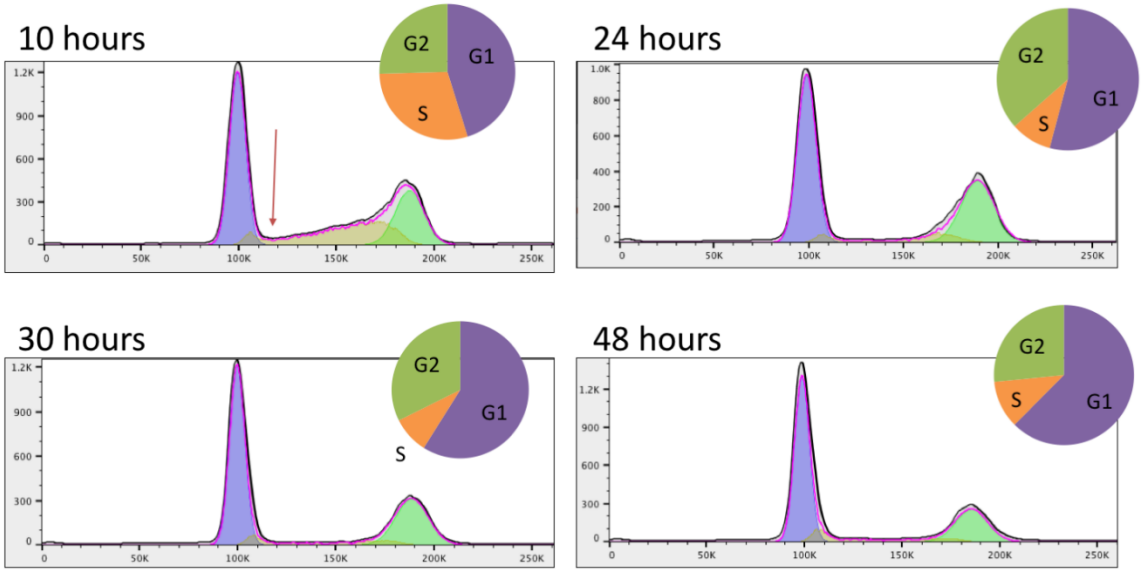

Figure 3-5 Cell cycle analysis of THP-1 cells, 10, 24, 30 and 48 hours post stimulation with PMA. Data was analysed using FlowJo, Watson model (pragmatic). Pie chart shows the proportion of cells in each phase. Red arrow shows the G1/S stage where the population of cells diminished after 10 hours of PMA stimulation. Histogram of number of cells in each phase is taken from FlowJo software, with X axis showing histogram parameter of propidium iodide fluorescence-A (Area), Y-axis shows cell count.

3.3.2 CAGE sequencing

3.3.2.1 RNA extractions and choosing samples for CAGE

RNA samples from the time courses were assessed for quality (RINe score) and quantity (concentration). Taking together all the information, 66 samples were selected for CAGE library preparation (highlighted green in Figure 3-6). Samples which are highlighted red did not have sufficient RNA quality or quantity. RINe scores of each sequenced sample is available in Appendix 8.3.

		Diff 1		Diff 2		Diff 3		Diff 4		Diff 5	
Timepoint		amount of RNA (µg)	Reads mapped	amount of RNA (µg)	Reads mapped	amount of RNA (µg)	Reads mapped	amount of RNA (µg)	Reads mapped	amount of RNA (µg)	Reads mapped
0	0	14.16	5,699,840	45	7,024,112	4.455	X	11.77	17,223,015	37.075	5,396,899
00:15	1	11.15		21.3		18.9425	X	16.0325		34.33	
00:30	2	17.1825		16.9		12.5825		10.2425		38.72	
00:45	3	12.94		14.5		14.0925				29.615	
01:00	4	12.1475	8,777,120	11.35	6,525,778	22.2825	9,598,514	11.935	25,612,992	67.67	
01:20	5	15.6325		18.4		18.795				42.045	
01:40	6	15.055		12.75		16.5375				64.685	
02:00	7	20.72	17,219,397	10.35	12,024,107	24.49	4,483,846	5.4475	5,417,069	108.625	
02:30	8	18.845		3.755	X	0.33575	X			31.03	
03:00	9	19.3425	12,987,295	12.75	11,780,232	19.91	4,942,942	7.0825	9,090,923	20.875	
03:30	10	18.97		10.35	X	21.44	X	17.7675		16.625	
04:00	11	13.67	2,735,692	18.45	15,212,394	29.365	9,463,074	9.61	6,252,772	28.62	
05:00	12	15.54	21,475,093	16.25	9,007,998	27.735	8,443,081			40.895	
06:00	13	17.1575	5,716,087	12.65	5,499,199	23.6025	16,342,678	21.6625	10,382,061	35.925	
07:00	14	16.545	8,077,012	14	9,514,908	20.7025	X	11.97	10,862,435		
08:00	15	26.3775	6,229,274	15.3	13,163,671	10.87	7,797,280	14.9175	10,758,835		
10:00	16	19.11	6,958,588	13.6	10,001,366	31.8675	6,245,739				
12:00	17.1	19.2675	5,526,165	21	11,619,664	19.7225					
24:00:00	23.1	3.0175	X	11.5	10,790,367	17.4075	7,331,012			22.505	
36:00:00	24.1	7.4825		10	5,297,212	6.8875	X	7.1675	3,239,350	13.255	
48:00:00	25.1	13.225	4,568,103	6.8	6,426,732	12.93	11,832,855				
96:00:00	26.1	14.275	5,867,382	14.85	9,094,756	17.955	4,074,529				
0	0.2			19.95	10,082,415	21.1775	12,304,712				
12:00	17.2	29.6875		29.25	6,030,920	35.64	13,206,798				
14:00	18.2	29.245	8,993,278	21.85	6,476,608	36.025	7,895,333				
16:00	19.2	16.88	11,003,763	18.3	10,453,298	27.08	8,864,592				
18:00	20.2	15.395	13,614,839	23	18,834,987	34.6225	15,444,884				
20:00	21.2	29.0625		12.05		20.1625					
22:00	22.2	22.87		9.8		25.22					
24:00:00	23.2	16.6	X	8.8	12,407,024	26.31	8,663,794				
36:00:00	24.2	8.3925	5,437,721	11.45		19.3225	6,788,935				

Figure 3-6 CAGE data points together with RINe number and recognised reads in CAGEr. Red X data entries denote insufficient RNA quality or quantity. Samples in green cells were used for CAGE. Black cells indicate that there was no sample for that time point.

3.3.2.1 CAGE sequencing quality assessment

RNA samples were successfully made into CAGE libraries. The CAGE product was amplified using PCR with the lowest number of cycles to give an adequate amount of product. The amount of final PCR product (at around 97 bp) was assessed using the Agilent TapeStation (section 3.2.4), as shown in a typical graph in Figure 3-7. All libraries had a smaller peak at around 75 bp, which was deemed to be primer dimer contamination.

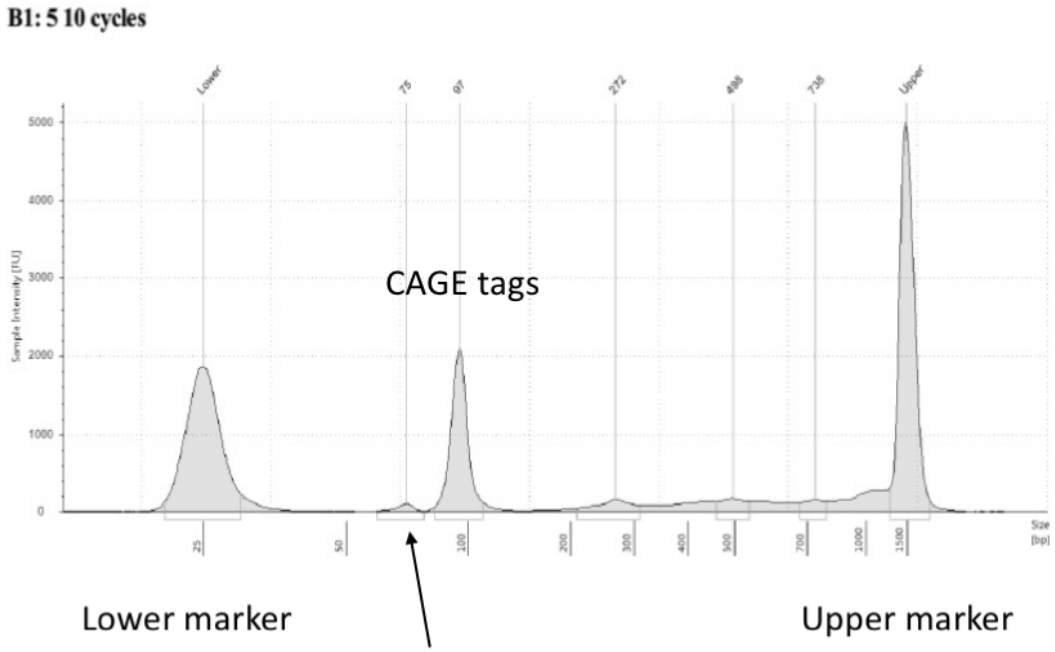

Figure 3-7 Tapestation results of fifth CAGE library after 10 PCR cycles. Lower and upper markers used by the system to estimate fragment lengths are shown. CAGE tags are the highest product peak, at around 97 bp long. Arrow points to the smaller peak at 75 bp.

The resulting sequences were sorted according to their barcode for each library, and their 5' and 3' barcodes trimmed off bioinformatically. Unfortunately, one barcode (CTT) seemed to have a large amount of sequences that had to be discarded due to their short length (see Appendix 8.3). FastQC analysis revealed that samples with this barcode had a high number of overrepresented sequences (sequences which are found in high numbers, indicating potential library contamination), which mostly belonged to an Illumina barcode (Figure 3-8).

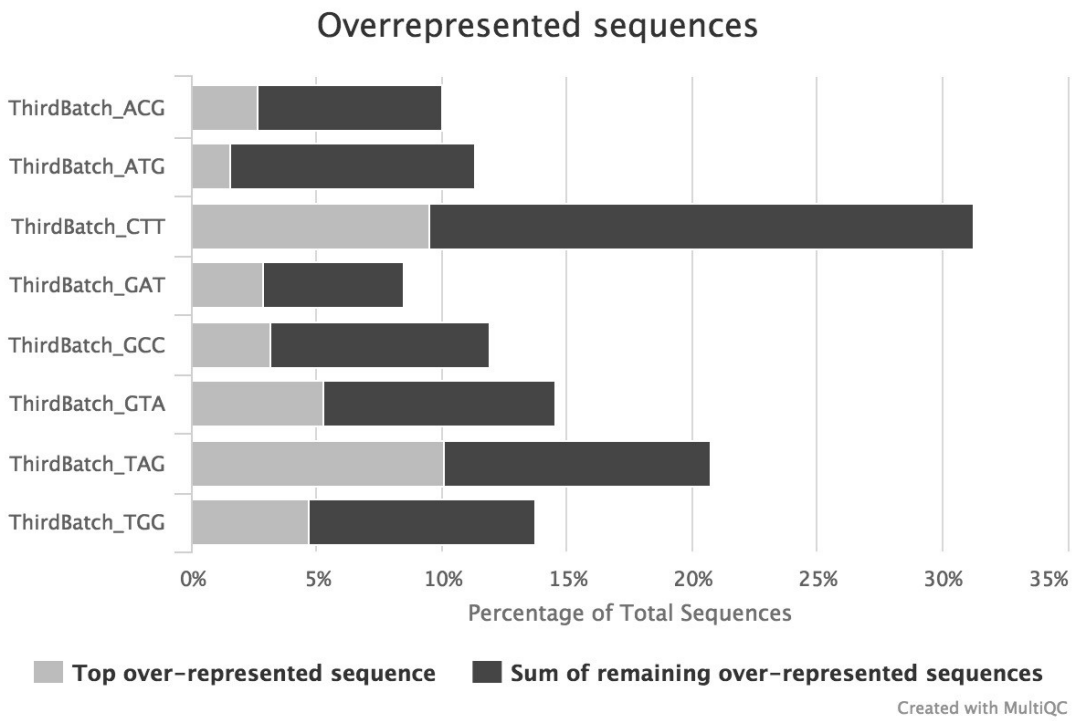

Figure 3-8 Overrepresented sequences in 8 barcodes in CAGE library number 3. Image created with MultiQC using FastQC data.

By the time this problem became apparent, 6 libraries had been sequenced or already prepared. The contamination was determined to be due to a faulty oligonucleotide containing CTT barcode (Dr Lel Eory, personal correspondence, 2017), and the CTT barcode oligonucleotide was remade for the last few libraries (ninth and tenth). This seemed to reduce the amount of overrepresented sequences (Figure 3-9). The CTT barcode was not used for CAGE libraries 11 and 12.

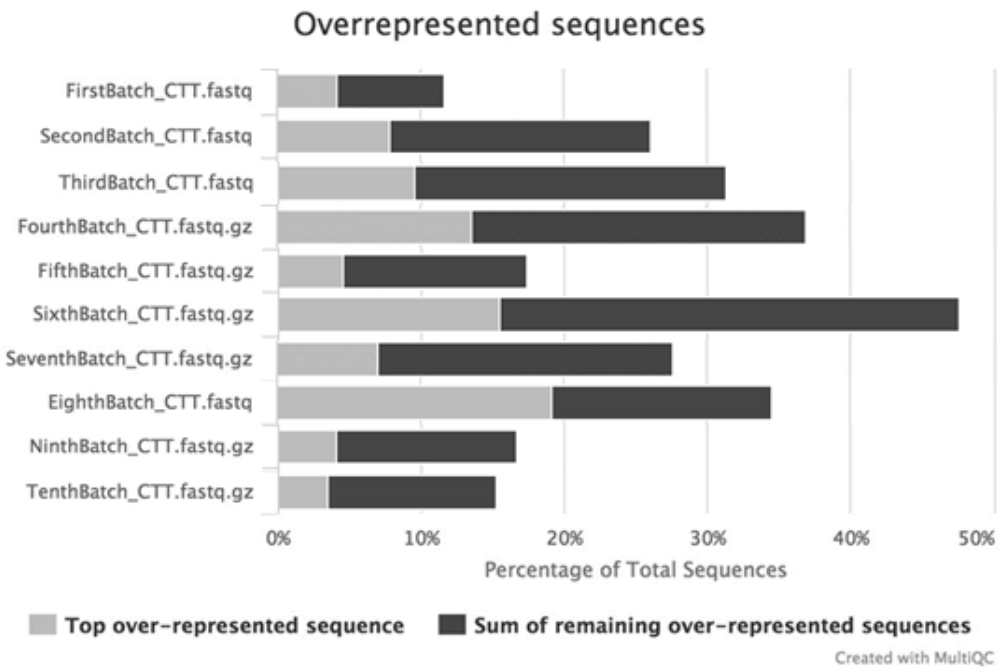

Figure 3-9 Percentage of overrepresented sequences in CTT barcoded samples in ten CAGE libraries. Image created with MultiQC using FastQC data.

Trimmed sequences were mapped to the human genome version hg38 (Yates *et al*., 2016). The number of mapped reads in 94 samples (from both Chapter 3 and Chapter 6) was very variable. The number could be as high as 33 million reads, or as low as around 3 million (Figure 3-10). The median value was around 9 million mapped reads.

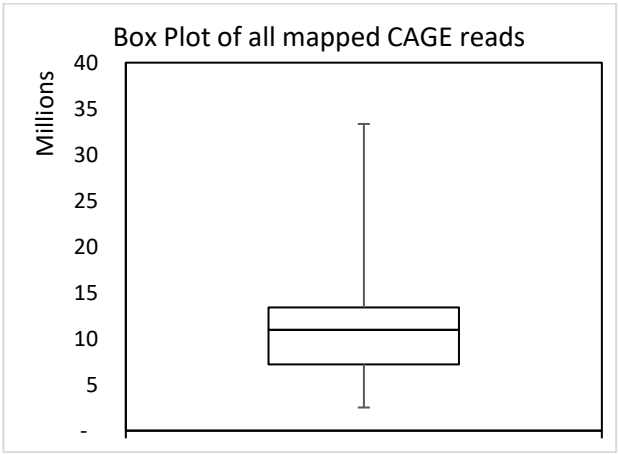

Figure 3-10 Box plot of all mapped reads from CAGE analysis showing maximum, interquartile ranges, median and minimum figures.

CAGE reads from the 66 samples analysed in this chapter were imported into the CAGEr programme and normalised against each other with the resulting value being in Tags per million (TPM), i.e. how many of the specific tags (reads) are found in one million reads. These CAGE tags indicated CAGE transcription start sites (CTSS) which were aggregated together according to their TPM values and distance between each other, into individual group clusters. These aggregated CTSS were annotated with the name of the gene closest to the middle of the CTSS.

3.3.2.2 THP-1 differentiation markers and qPCR validation of key enzymes

To ensure the differentiation of THP-1 cells from pro-monocytic to macrophage-like phenotype was confirmed also at the transcriptional level, CAGE sequencing data was checked for two relevant markers, *MYB* which encodes a transcription factor associated with proliferating haematopoietic cells (reviewed by Zhou and Ness (2011)) and monocyte differentiation antigen gene CD14 (Bazil *et al*., 1982, Wright *et al*., 1990). CAGE results were also simultaneously validated by qRT-PCR of the same genes by testing the RNA the CAGE libraries were made from.

MYB was downregulated during THP-1 differentiation in the earlier study (Suzuki *et al*., 2009). In the present study, *MYB* expression was found to decrease after PMA stimulation in all biological replicates, with the lowest point being at 3 hours (Figure 3-11). qRT-PCR experiments showed a similar pattern, where *MYB* was also downregulated, with the lowest level at between 2:30 and 3 hours (Figure 3-12). Note that qRT-PCR was carried out on all samples taken from the THP-1 differentiation experiment, not just the 18 time points selected for CAGE sequencing.

Expression of *CD14*, which was not expressed at all in undifferentiated THP-1 clone 5 cells (Bosshart and Heinzelmann, 2016), was found to be very highly upregulated in the last few points of the time course (Figure 3-13) according to CAGE sequencing. The same pattern was documented by qRT-PCR experiments (Figure 3-14).

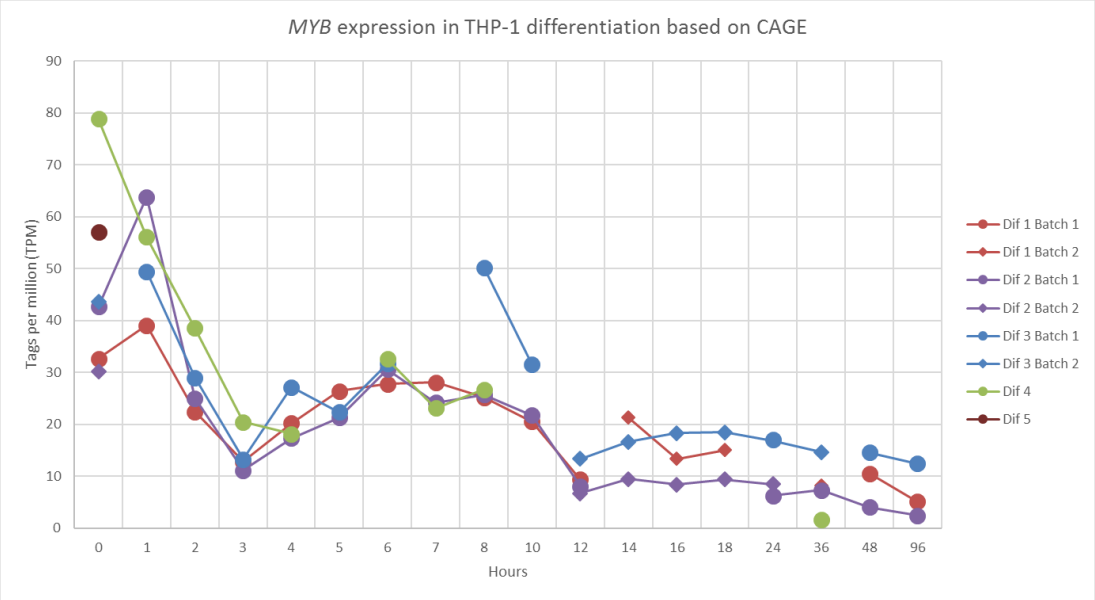

Figure 3-11 MYB expression in THP-1 differentiation based on CAGE sequencing showing all biological replicates. Data is normalised to Tags per million (TPM).

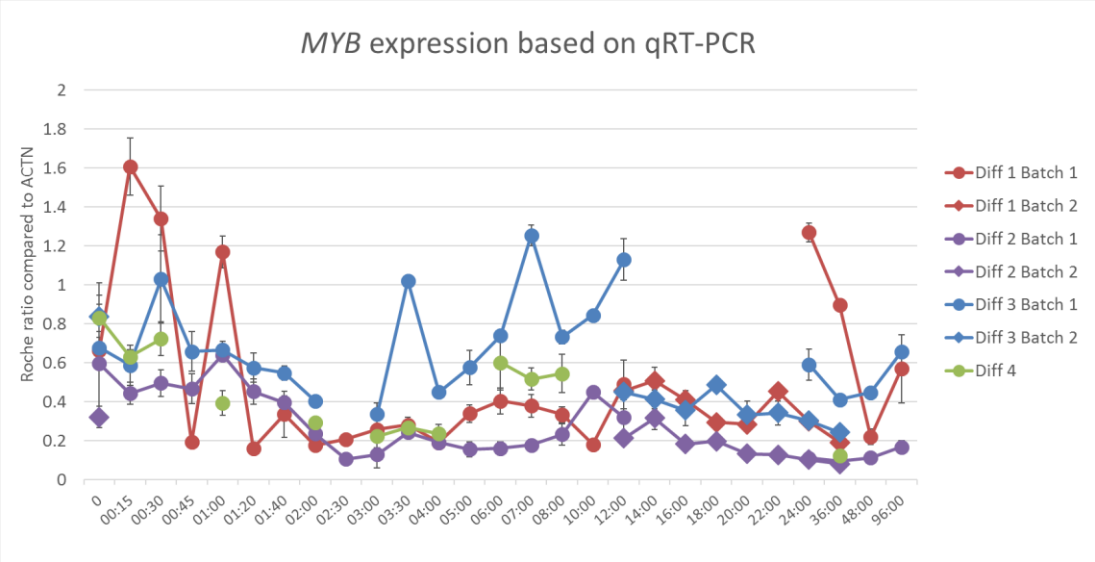

Figure 3-12 MYB expression in the whole THP-1 differentiation time course (all time points taken) based on qRT-PCR. Pictured here is the MYB expression normalised to ACTN expression and error bars showing standard error (based on two technical replicates and calculated by the by Roche LightCycler 480 software).

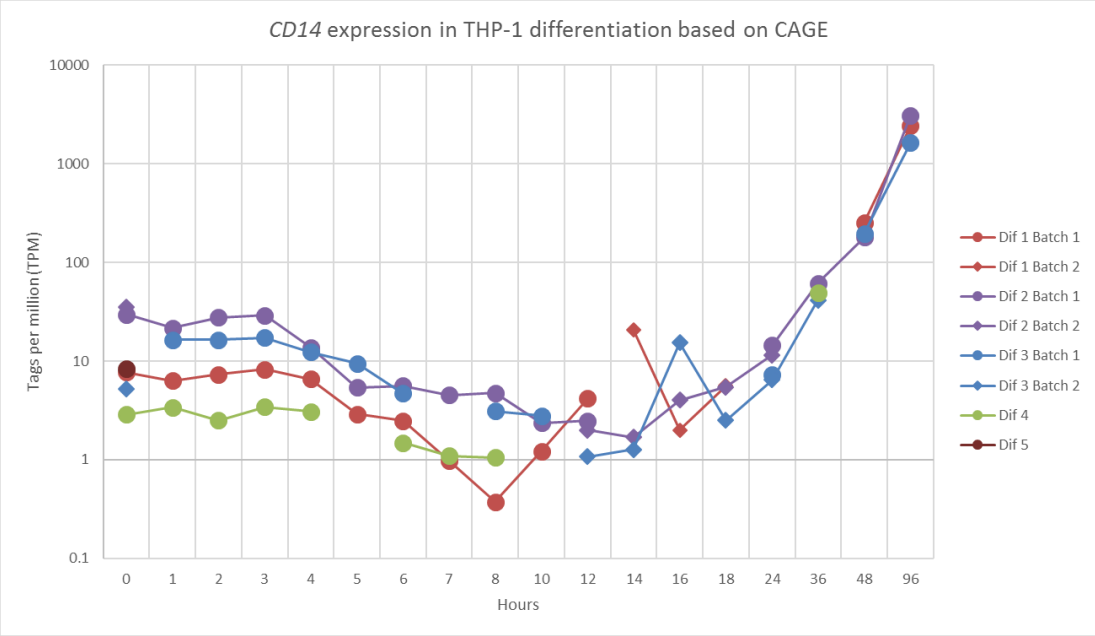

Figure 3-13 CD14 expression in THP-1 differentiation based on CAGE sequencing showing all biological replicates. Data is normalised to Tags per million (TPM). Note that the Y axis is in logarithmic scale.

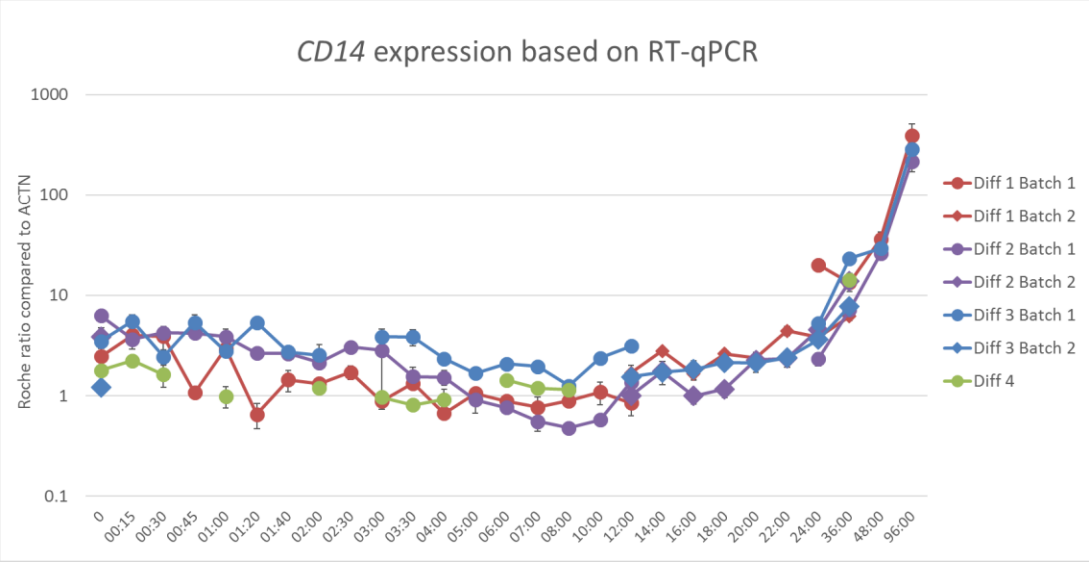

Figure 3-14 CD14 expression in the whole THP-1 differentiation (all time points taken) based on qRT-PCR. Note that the Y axis is in logarithmic scale. Pictured here is the CD14 expression normalised to ACTN expression and error bars showing standard error (based on two technical replicates and calculated by the by Roche LightCycler 480 software).

CAGE sequencing data (validated by qRT-PCR) indicated that the THP-1 cells had differentiated, and the quality and quantity of reads were deemed satisfactory. Thus, the next few sections of this chapter deal with analysis of this data, using various bioinformatic tools, looking for waves of gene activation and repression across the large number of tight time points in the time course and any additional information that could be found due to deep sequencing of the samples.

3.3.3 Clusters of CTSS throughout differentiation

CAGE sequencing data values were loaded into the network analysis tool Miru (section 2.5.5) to find patterns of gene activation and repression across the time course. Firstly, a sample-to-sample network of all 66 samples with 18 different time points was created using a correlation coefficient of 0.92 or higher. (Figure 3-15). The samples at the beginning of the differentiation (0 hours, deep purple) are at the other end of the network from the fully differentiated THP-1 cells at 96 hours (deep red). The graph shows that the samples followed the differentiation pattern by hour, as for example samples of 0 hour are mostly correlated to 1 hour sample and so on. There are few exceptions, most notably 2 points of 4 hours (blue) samples (Dif1_11 and Dif4_11) which are connected to 36 hours and 8 hours (red and green). Most biological repeats of each time point clustered together, which shows that they were highly correlated to each other. Again, there were exceptions, especially the two 4 hour samples. These samples were also the least correlated to any other samples, as they disappeared from the graph when the correlation coefficient was raised to 0.94. Thus, it was decided to exclude these two samples from further analysis. One of the samples (Dif1_11) had the lowest number of mapped reads of all samples (2.7 million). The second sample (Dif4_11) had 6.2 million mapped reads, which was one of the lower numbers of reads.

When these two samples were removed, and correlation coefficient was raised to 0.94, the sample-to-sample correlation graph shows the differentiation stages more clearly, with 0 hours the furthest away from 96 hours (Figure 3-16).

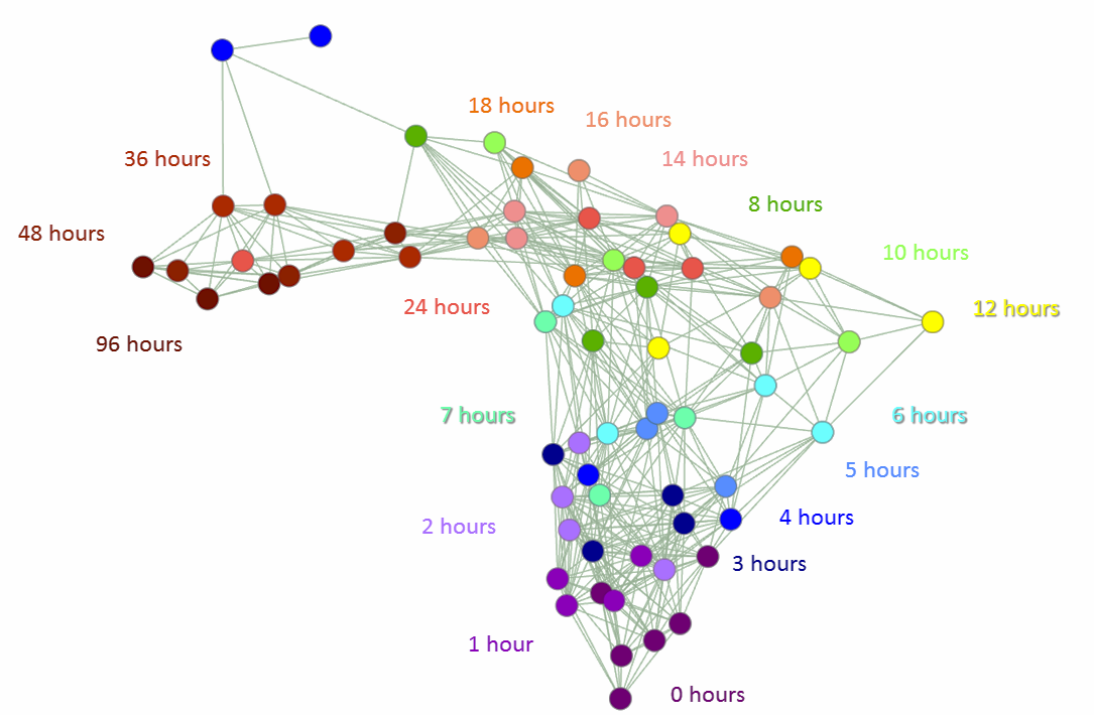

Figure 3-15 Sample-sample network at correlation coefficient of ≥ 0.92 derived from Miru programme (all 66 samples). Samples are nodes and edges are connections between them with a correlation coefficient of 0.92 or greater, i.e. how closely correlated the expression patterns are to each other across all samples. Nodes which are close are highly correlated. Each time point is different colour, as pictured in the figure.

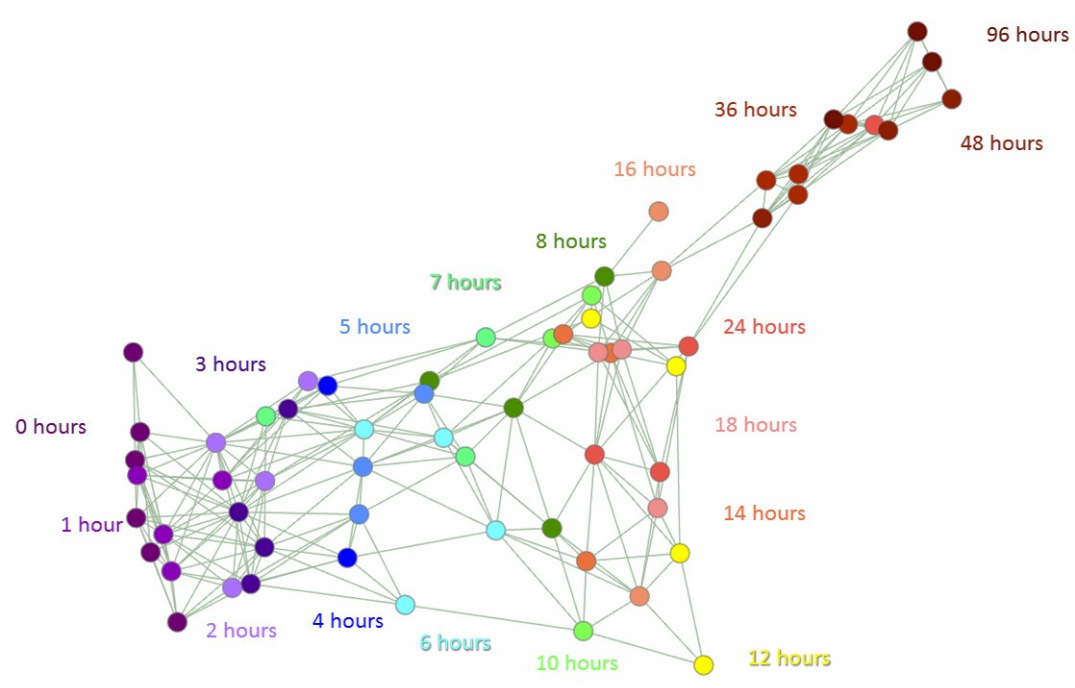

Figure 3-16 Sample-sample network at correlation coefficient of ≥ 0.94 derived from Miru programme of 64 samples, excluding two samples at 4 hours. Samples are nodes and edges are connections between them with a correlation coefficient of 0.94 or greater, i.e. how closely correlated the expression patterns are to each other across all samples. Nodes which are close are highly correlated. Each time point is different colour, as pictured in the figure.

Next, aggregated CTSSs were correlated with each other based on the expression from 64 samples (to create a gene-to-gene network). The analysis was done with a correlation coefficient threshold of 0.75, which included 12,275 nodes out of 15,341 possible aggregated CTSS (Figure 3-17). Expression patterns of excluded CTSSs were not correlated with any other CTSS at a correlation coefficient of 0.75.

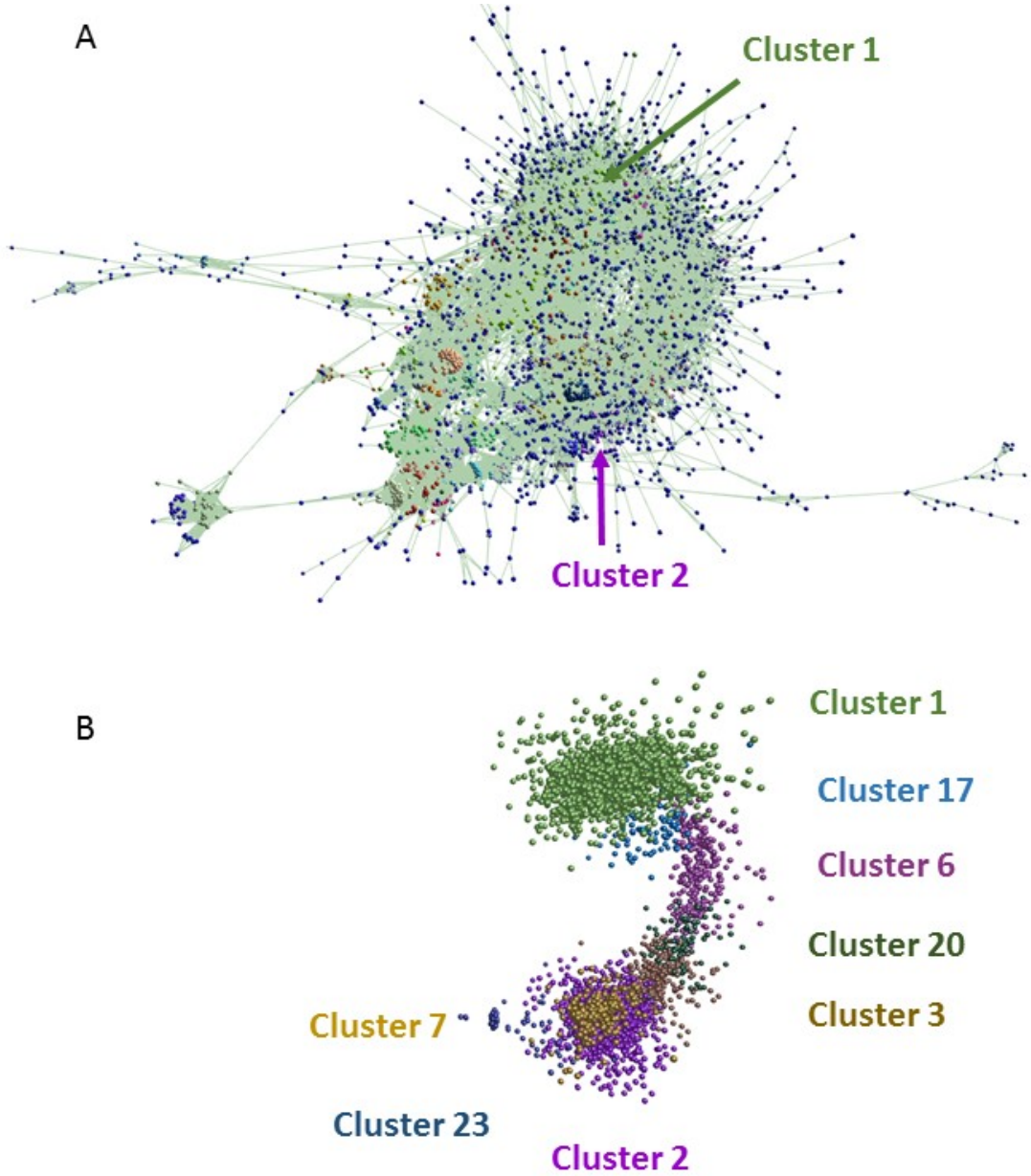

Figure 3-17 Network analysis of CTSS expression patterns. A. Image from Miru shows the network layout based on correlations between CTSSs (nodes = CTSS; edges = correlations of ≥ 0.75). 64 THP-1 differentiation samples were analysed using a Pearson's correlation coefficient ≥ 0.75. Location of important clusters is annotated on the main element of the network correlation. B. Image showing the eight clusters that demonstrate the transition in expression patterns. Edges have been removed for ease of visualisation. Orientation is the same as for A.

Eight of the interesting clusters were selected for detailed analysis and their gene lists were screened for GO term enrichment using DAVID programme. Expression profiles of genes in these clusters are shown in Figure 3-18, ordered by the time point with the highest expression, from most monocyte-like (Cluster 1) to most macrophage-like cells (Cluster 7). These profiles show that there were waves of gene expression as differentiation progressed, with the functions of the group of genes changing as the phenotype of the cells changed. A list of genes associated with CTSSs in each cluster and the average expression profiles of the clusters are presented in Appendix 8.4.

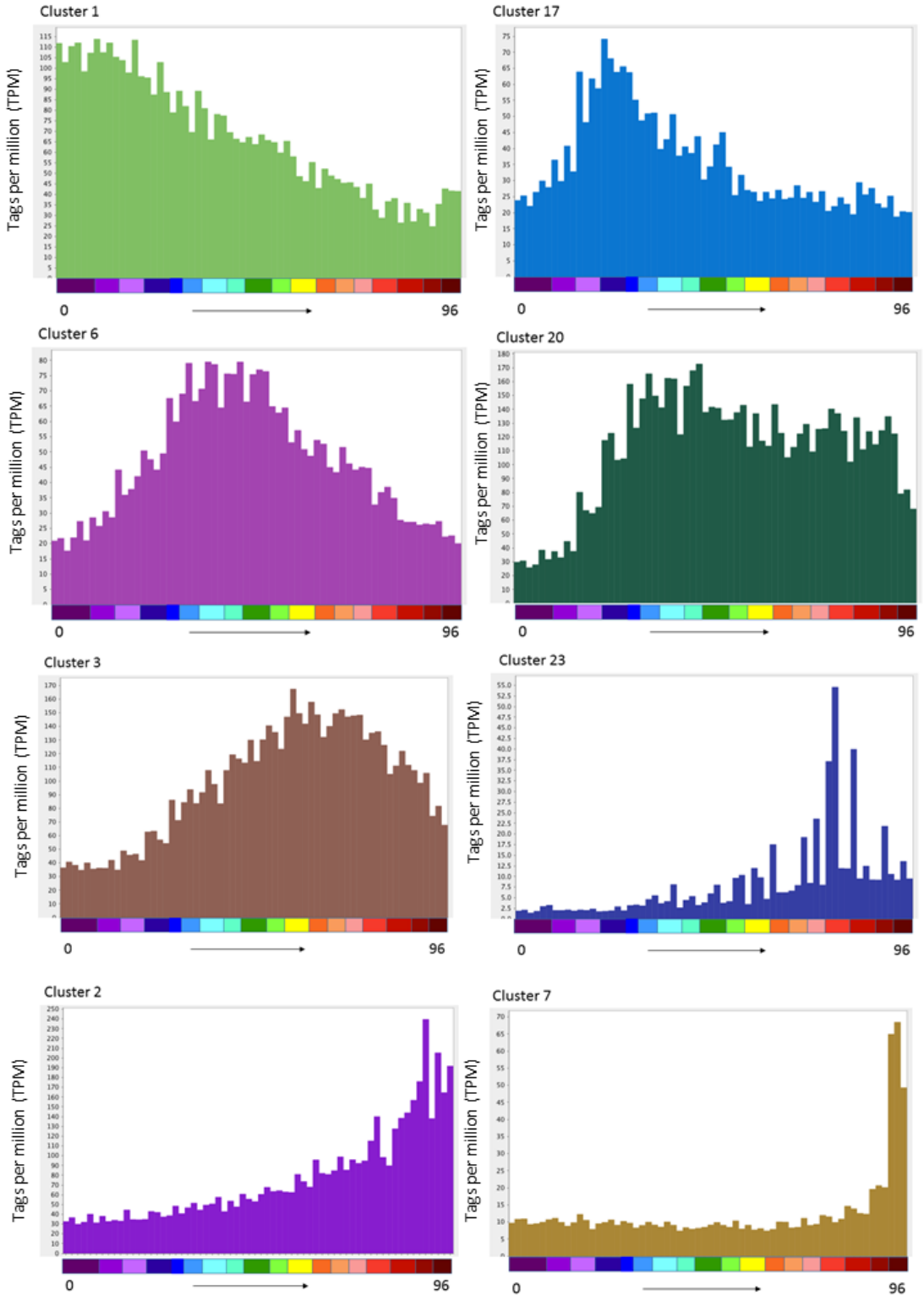

Figure 3-18 Expression profiles of different clusters found with Miru programme at correlation coefficient ≥ 0.75 during THP-1 differentiation. Average expression for all CTSS in the cluster is shown. Clusters are ordered based on the highest level of expression, from monocyte-like expression (Cluster 1) to most macrophage-like expression (Cluster 7) of the THP-1 differentiation. Y axis is in Tags per million (TPM), X axis shows the individual colour-coded time points, from 0 hours to 96 hours. Colours are the same as in Figure 3-16.

Cluster 1		**GO terms**	**Group score**
	Group 1	Cell cycle, cell division	37.68
	Group 2	Nucleosome, histone core	25.09
	Group 3	DNA damage, DNA repair	17.01
	Group 4	Kinetochore, centromere	16.31
	Group 5	Telomere, nucleosome core, epigenetic regulation	9.19
	Group 6	Histone H2A	7.09
	Group 7	Mitochondria	6.51
	Group 8	Fanconi anemia pathway	5.8
	Group 9	ATP binding	4.84
	Group 10	Histone H2B	4.77
Cluster 17		**GO terms**	**Group score**
	Group 1	Transcriptional regulation, DNA binding	2.35
	Group 2	Zinc finger	2.1
Cluster 6		**GO terms**	**Group score**
	Group 1	Nucleotide (ATP) binding, protein kinase	3.42
	Group 2	Transcriptional regulation, DNA binding	2.06
Cluster 20		**GO terms**	**Group score**
	Group 1	Cell-cell adhesion, cadherin binding	2.34
	Group 2	Signal transducer activity	2.19
	Group 3	Endosomal transport, autophagy	2.07
Cluster 3		**GO terms**	**Group score**
	Group 1	Lysosome	4.59
	Group 2	Endoplasmic reticulum	3.3
	Group 3	Phosphotyrosine interaction domain	2.41
Cluster 23		**GO terms**	**Group score**
	Group 1	Antiviral defence, innate immunity	8.81
	Group 2	Double stranded RNA binding, 2'-5' oligoadenylate synthetase activity	4.61
	Group 3	TPR repeat	1.94
Cluster 2		**GO terms**	**Group score**
	Group 1	Lysosome	11.55
	Group 2	Pleckstrin homology-like	7.29
	Group 3	Membrane, transmembrane	5.01
	Group 4	Phosphatidylinositol binding	3.89
	Group 5	Integrin complex	3.48
	Group 6	WD repeat domain	3.22
Cluster 7		**GO terms**	**Group score**
	Group 1	Lysosome, Glycoside hydrolase	2.11
	Group 2	Lipoprotein metabolic process, phospholipid efflux, cholesterol	2.04

Table 3-1 DAVID GO term annotation for different clusters found at a correlation coefficient ≥ 0.75 using Miru programme.

GO term analysis for most monocyte-like Cluster 1 (including cell cycle genes such as histone-coding genes (e.g. *HIST1H1B*) and *BUB1* and mitochondrial genes such as *NDUFA5*) includes high scores for cell cycle, mitochondria and DNA repair (Albig *et al*., 1997, Tang *et al*., 2004, Marui *et al*., 2011). These CTSS were associated with high expression in THP-1 cells at their undifferentiated state (Table 3-1). As differentiation progressed, early response genes with functions such as transcriptional activation (Clusters 17 and 6, including genes such as *SP3* for Cluster 17 and *GTF2I* for Cluster 6) were turned on at around 2 and 4 hours respectively, with concurrent upregulation of genes, such as *EFR3A* and *HRH1*, which code for proteins involved in signal transduction and cell adhesion at 4 hours (Cluster 20) (Essafi-Benkhadir *et al*., 2009, Roy *et al*., 1993, Bojjireddy *et al*., 2015, Ayuso *et al*., 2013). Expression of genes (e.g. *CD63* and *CD109*) in Cluster 3 slowly increased from the 2 hour time point, peaking at 12 hours, and this cluster was enriched for GO terms such as innate immunity and antiviral defence (Anzai *et al*., 2002, Lin *et al*., 2002). The second largest cluster, Cluster 2, contained genes (like *ATP6* family members, *CD163*, *CSF1R* and integrin family members such as *ITGAM*) which code for proteins associated with lysosome activity, transmembrane location and integrin complex (Rojo *et al*., 2017, Hume and Freeman, 2014, Hume *et al*., 2016). This cluster contained genes with expression that steadily increased throughout differentiation, peaking at the end of the time course. Cluster 7 contained genes (for example *CD14* and *DNASE2*) which were upregulated only very late in the time course, at 48 hours and 96 hours (Wright *et al*., 1990, Kawane *et al*., 2001). This cluster was enriched for GO terms such as lysosome and lipoprotein metabolic processes.

3.3.4 Novel enhancers found in THP-1 differentiation assay

One of the features of CAGE sequencing is that this technology is sensitive enough to discern RNA derived from enhancers, which are characterised by balanced bidirectional transcription (Andersson *et al*., 2014). Thanks to the deep sequencing of the CAGE samples and the many closely spaced time points, it was possible to identify enhancer activity in this dataset.

MYB gene expression, as mentioned in section 3.3.2.2, decreases during THP-1 differentiation with PMA. FANTOM5 identified one *MYB* enhancer, close to its promoter (orange arrow in Figure 3-19). However, in the THP-1 differentiation data, there was also a strong bidirectional peak further from the promoter (blue arrow in Figure 3-19), for which expression also declined throughout the THP-1 differentiation (Figure 3-20 and Figure 3-21). This putative enhancer has a similar pattern of expression to the confirmed FANTOM5 enhancer. There are two annotated transcripts, one in each direction, for which the TSS is at the same position as the putative enhancer (ENST00000455534.1 and ENST00000534736.1). Similar to the THP-1 samples prior to PMA treatment, the negative strand was not widely expressed in any FANTOM5 samples, but the positive strand was most highly expressed in leukaemia samples (Table 3-2).

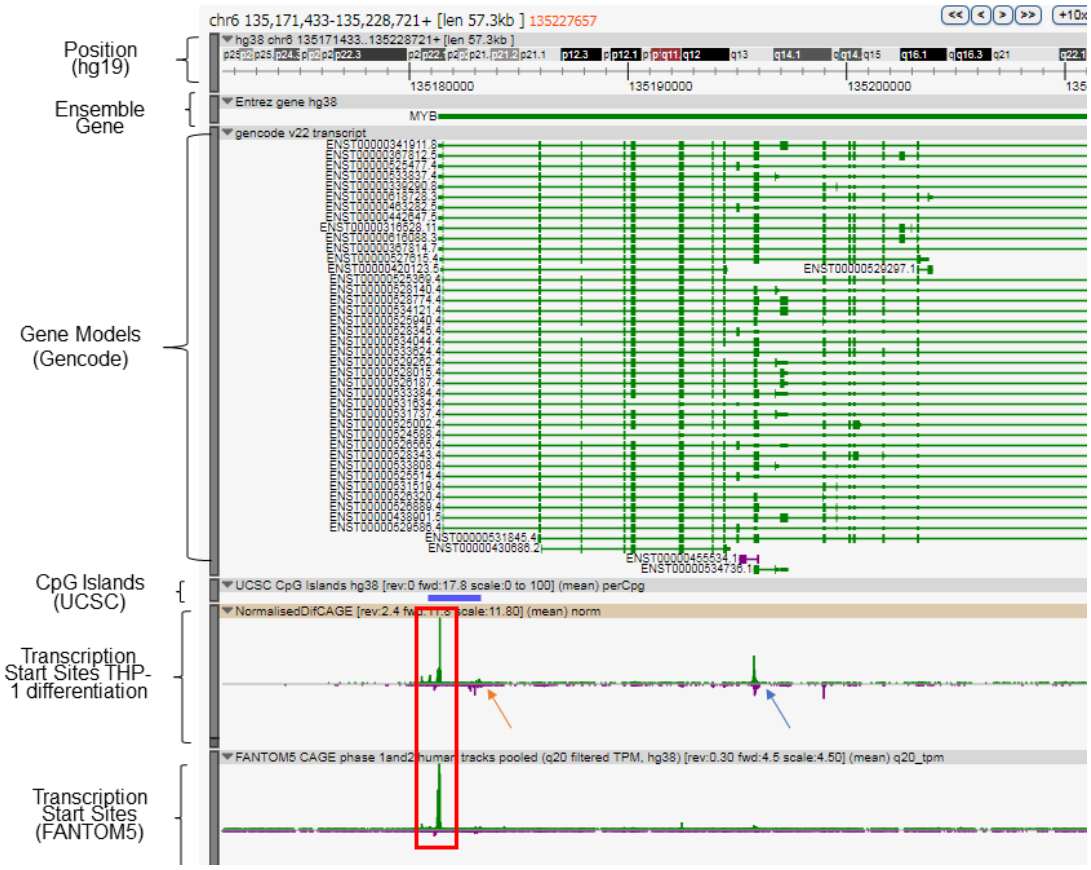

Figure 3-19 MYB gene structure and its TSS. Image taken from genome viewer Zenbu. Upper tracks show the position, extent of the locus as determined by Ensembl, gene models from Gencode data, and identified CpG islands from UCSC data. Lower tracks show the number of tags at the TSS detected in this study (all 66 samples) and TSS detected in the FANTOM5 study (Forrest et al., 2014) re-mapped to hg38. Green indicates transcription from the forward strand; purple indicates transcription from the reverse strand. Red box shows MYB promoter, orange arrow shows known MYB enhancer, blue arrow shows MYB bidirectional TSS (putative enhancer).

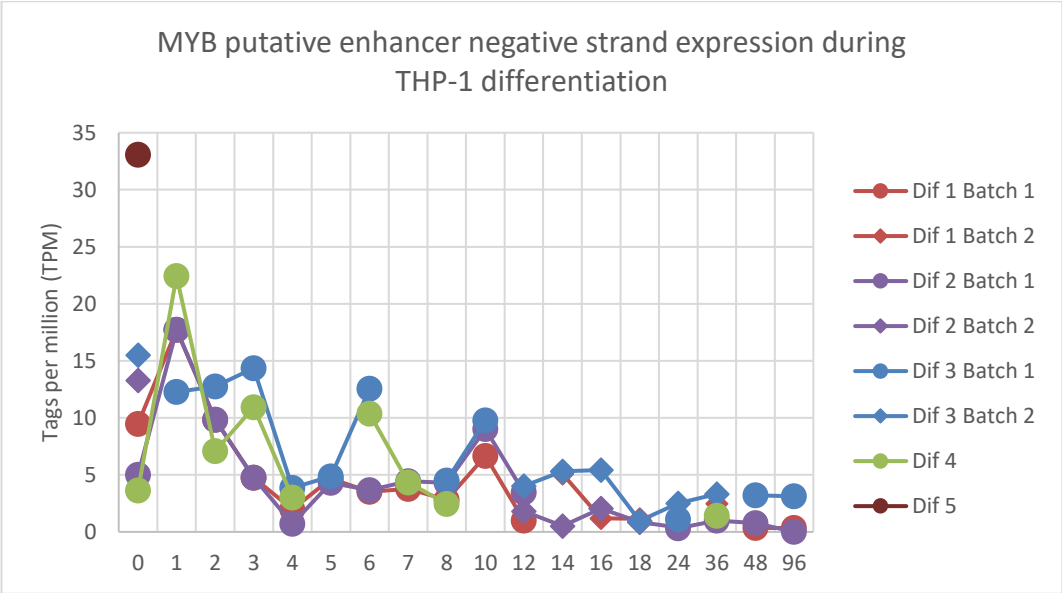

Figure 3-20 MYB putative enhancer negative strand expression during THP-1 differentiation based on CAGE sequencing data. Y axis shows normalised values in Tags per million (TPM); X axis shows time points after PMA treatment.

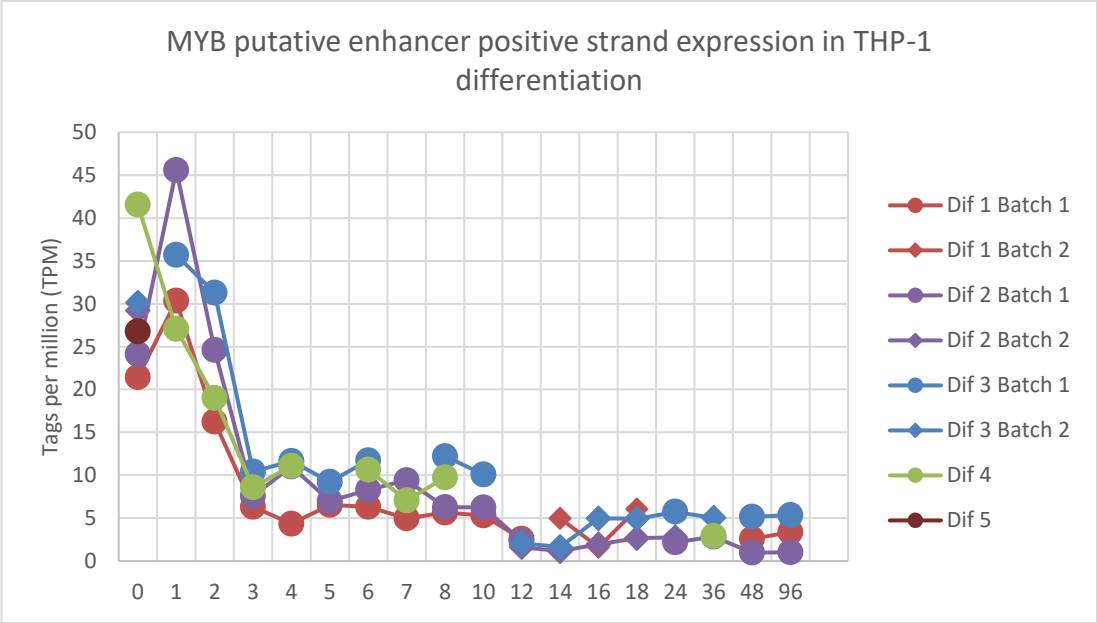

Figure 3-21 MYB putative enhancer positive strand expression during THP-1 differentiation based on CAGE sequencing data. . Y axis shows normalised values in Tags per million (TPM); X axis shows time points after PMA treatment.

FANTOM5 sample	Negative strand TPM	Positive strand TPM
biphenotypic B myelomonocytic leukemia cell line:MV-4-11	2.628	109.853
acute myeloid leukemia (FAB M5) cell line:THP-1 (cytoplasmic fraction)	1.187	91.758
acute myeloid leukemia (FAB M5) cell line:THP-1 (fresh)	1.959	85.813
acute myeloid leukemia (FAB M5) cell line:THP-1 (thawed)	2.298	84.189
acute myeloid leukemia (FAB M5) cell line:THP-1 (revived)	1.456	64.367
acute myeloid leukemia (FAB M5) cell line:U-937 DE-4	3.296	31.64
acute myeloid leukemia (FAB M1) cell line:HYT-1	0.823	29.233
acute myeloid leukemia (FAB M4eo) cell line:EoL-3	2.808	27.276
acute myeloid leukemia (FAB M5) cell line:P31/FUJ	1.312	27.223
myelodysplastic syndrome cell line:SKM-1	1.502	24.275
acute myeloid leukemia (FAB M3) cell line:HL60	1.186	22.202
acute myeloid leukemia (FAB M2) cell line:Kasumi-1	0.662	20.514
acute myeloid leukemia (FAB M4eo) cell line:EoL-1	1.949	20.017
non T non B acute lymphoblastic leukemia (ALL) cell line:P30/OHK	2.554	19.579
chronic myelogenous leukemia (CML) cell line:MEG-A2	3.709	19.102

Table 3-2 Expression of putative MYB enhancer in FANTOM5 samples mapped to hg38. The 15 samples with highest positive strand expression are shown. Values are in Tags per million (TPM).

3.3.5 Promoter switching found during THP-1 differentiation

One of the advantages of CAGE technology is that it is able to identify specific promoter usage and thus whether the use of different promoters shifts (called promoter switching (Forrest *et al*., 2014)) during a cell state transition such as the differentiation of THP-1 monocytic cells. The programme CAGEr was used to identify whether there was a shift between two different promoters in the same aggregated CTSS between two time points using the scoreShift command.

Most of the shifting promoters identified were false positives. Most false positives had low overall TPM value or the distance between the TSS was too small to count as a different promoter start. Three different genes were identified initially to have a potential shift in promoter usage in THP-1 differentiation (Table 3-3).

Shifting promoters between:

0 to 96 hours

chr	start	end	strand	shifting.score	pvalue.KS	fdr.KS	AssociatedGene
chr1	161069633	161069976	-	0.609055	0.00E+00	0.00E+00	*ARHGAP30*

0 to 6 hours

chr	start	end	strand	shifting.score	pvalue.KS	fdr.KS	Associated Gene
chr8	73976187	73976354	+	0.6043697	1.11E-15	9.05E-14	*TMEM70*

0 to 8 hours

0 to 10 hours

chr	start	end	strand	shifting.score	pvalue.KS	fdr.KS	AssociatedGene
chr6	144064566	144064695	-	0.6536973	5.35E-11	2.28E-09	*PLAGL1*

Table 3-3 Promoter switching identified using CAGEr programme. The time points where differential promoter usage was detected are given above each part of the table. PLAGL1 was flagged in two searches, 0 vs 8 hours and 0 vs 10 hours. Each part of the table shows chromosome number (chr), start, end and strand orientation of the region. Shifting score, p-value and FDR were calculated by CAGEr. Shifting score is from –Infinity to 1, 1 being complete shift of expression. P-value and False Discovery Rate (FDR) were calculated using Kolmogorov-Smirnov (KS) test.

One of the shifting promoters identified was associated with *ARHGAP30*. *ARHGAP30* has two main transcripts, one is longer (ENST00000368013 coding for 1101 amino acid long peptide) and the other is shorter (ENST00000368016 coding for 890 amino acids) (Figure 3-22). According to the Ensembl database, the only protein coding transcripts are these two and ENST00000368015 coding for 924 amino acids, for which the promoter is near the longer transcript (Figure 3-22).

The area shown as P1 (Figure 3-22) had ubiquitous expression in THP-1 differentiation, with little variation throughout (Figure 3-23). This promoter was associated with ENST00000368013 and ENST00000368015. On the other hand, the area shown as P2 had a drastic change in its expression (Figure 3-24), as it sharply increased from around 5 TPM at 0 hour to 120 TPM (similar levels as P1) at 96 hours of the time course (both graphs are at the same Y axis scale). This promoter was associated with ENST00000368016.

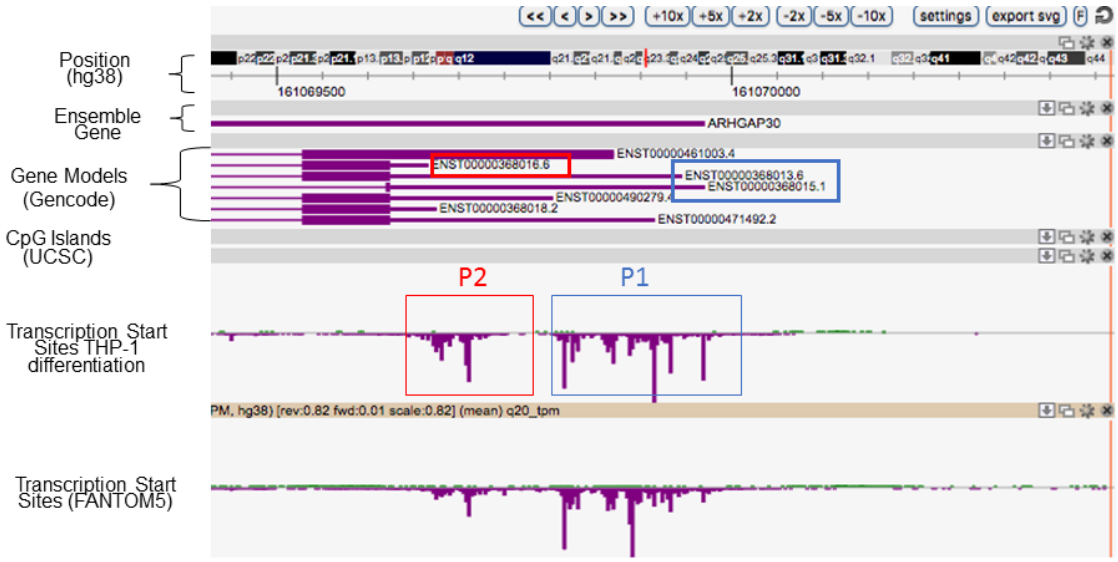

Figure 3-22 ARHGAP30 has two promoters P1 (in blue box) and P2 (in red box) with differential expression during THP-1 differentiation. This picture is taken from the Zenbu browser, with the data from this chapter (labelled Transcription Start Sites THP-1 differentiation) and FANTOM5 data remapped to hg38 (labelled Transcription Start Sites (FANTOM5)). Protein-coding transcripts associated with these promoters are highlighted in blue (P1) or red (P2) box in the Gene Models track. Transcription is from the reverse strand, from right to left of the image.

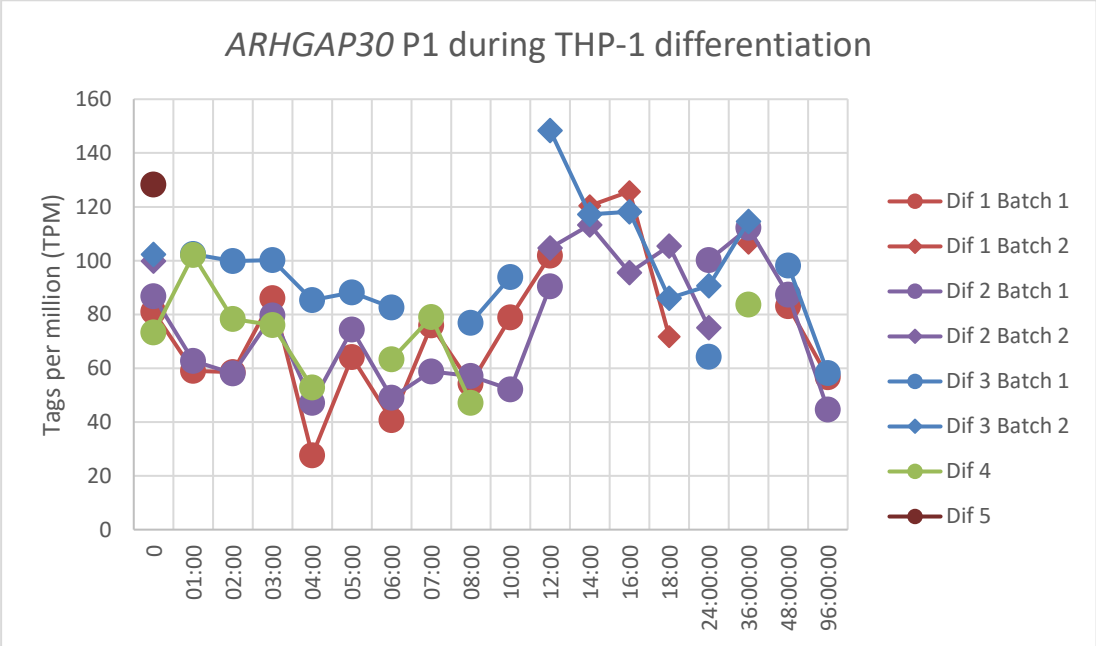

Figure 3-23 Expression of ARHGAP30 in area P1 during THP-1 differentiation according to CAGE sequencing. Expression is shown as normalised Tags per million (TPM) value.

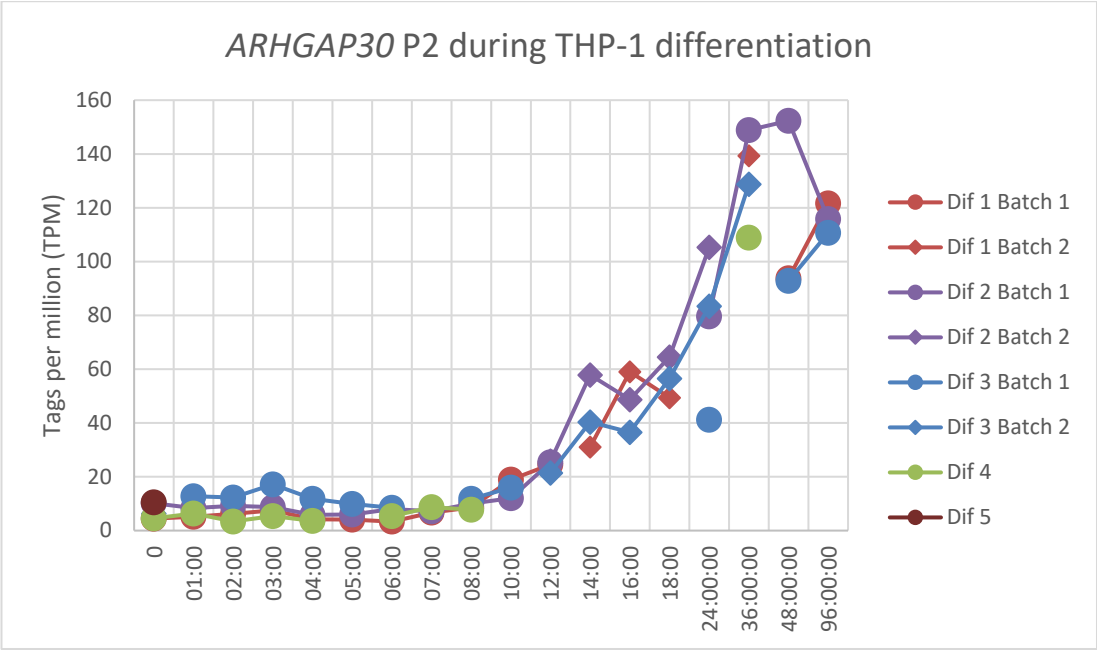

Figure 3-24 Expression of ARHGAP30 in area P2 during THP-1 differentiation according to CAGE sequencing. Expression is shown as normalised Tags per million (TPM) value.

Looking at samples from FANTOM5 with the highest TPM for the same P1 and P2 areas, P1 was mostly expressed in primary hematopoietic cells, especially T cells (Table 3-4). P2 did not have high TPM values for these samples. However, when looking at samples with the highest P2 expression, they were also primarily hematopoietic samples, with most hits being in the time course of monocyte-derived macrophages treated with LPS (highlighted grey in Table 3-4). Samples which had high TPM for P2 also had similar TPM levels for P1.

When the amino acid sequences of the two most important transcripts according to Ensembl database (ENST00000368013 and ENST00000368016) were compared to each other using MEG AlignPro with Clustal Omega, the main difference in these two peptides is that the shorter transcript is missing a glutamic acid rich stretch at the C-terminal end (Figure 3-25).

	RNA samples with highest expression of P1 *ARHGAP30*	P1	P2
1	CD4+CD25-CD45RA- memory conventional T cells expanded, donor3	149.733	1.156
2	gamma delta positive T cells, donor1	142.28	1.426
3	CD4+CD25+CD45RA+ naive regulatory T cells expanded, donor3	142.19	1.063
4	CD8+ T Cells (pluriselect), donor090309, donation1	134.62	1.022
5	CD8+ T Cells (pluriselect), donor090325, donation1	132.98	5.91
6	CD8+ T Cells (pluriselect), donor090612, donation1	132.344	2.877
7	CD4+CD25+CD45RA+ naive regulatory T cells expanded, donor2	130.724	0.945
8	CD8+ T Cells (pluriselect), donor090612, donation2	124.836	6.323
9	CD8+ T Cells (pluriselect), donor090325, donation2	123.533	3.148
10	CD4+CD25+CD45RA- memory regulatory T cells expanded, donor2	119.32	0.38

	RNA samples with highest expression of P2 *ARHGAP30*	P1	P2
1	Monocyte-derived macrophages response to LPS, 01hr00min, donor1	14.968	67.357
2	Monocyte-derived macrophages response to LPS, 10hr, donor1	22.489	54.617
3	Monocyte-derived macrophages response to LPS, 36hr, donor1	32.783	50.267
4	Eosinophils, donor2 : CNhs12548 ctss	61.493	45.249
5	Neutrophils, donor1 : CNhs10862 ctss	49.175	40.069
6	Monocyte-derived macrophages response to LPS, 16hr, donor1	19.892	39.785
7	Neutrophils, donor3 : CNhs11905 ctss	65.25	37.689
8	CD14-CD16+ Monocytes, donor2 : CNhs13207 ctss	34.454	34.454
9	CD14-CD16+ Monocytes, donor1 : CNhs13229 ctss	46.338	33.098
10	Monocyte-derived macrophages response to LPS, 24hr, donor1	29.129	32.928

Table 3-4 Highest TPM values for P1 and P2 regions of ARHGAP30 taken from FANTOM5.

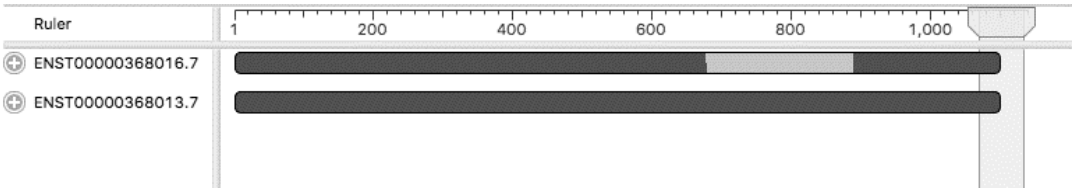

Figure 3-25 Two most important ARHGAP30 transcripts coding for polypeptides. This image is taken from MEGAlignPro software, which aligned the two amino acid sequences using Clustal Omega. The light grey area in ENST00000368016 is missing compared to the longer transcript.

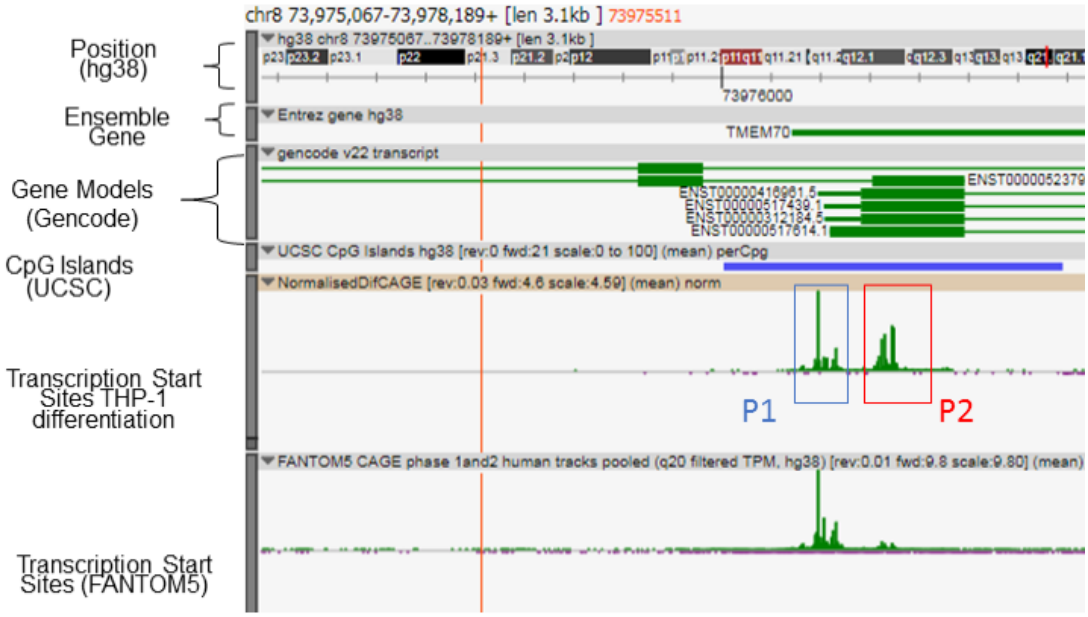

Figure 3-26 TMEM70 has two promoters P1 (in blue box) and P2 (in red box) with differential expression during THP-1 differentiation. This picture is taken from Zenbu browser, with the data from this chapter (labelled Transcription Start Sites THP-1 differentiation) and FANTOM5 data remapped to hg38 (Transcription Start Sites (FANTOM5)). Transcription is from the forward strand, from left to right of the image.

Another gene with two promoter regions with varying expression during THP-1 differentiation was *TMEM70* (Figure 3-26). P1was located at the beginning of multiple coding transcripts, whereas P2 was at the beginning of the translated region of this gene, where none of the known transcripts' start sites were located. P1 expression did not change very much throughout the THP-1 differentiation (Figure 3-27), but P2 was associated with increase of expression after the start of PMA stimulation, with decrease at 14 and 16 hours, and additional decrease as the cells become macrophage-like (48 and 96 hours) (Figure 3-28). The P1 region of *TMEM70* was most highly expressed in tumours and cell lines among the FANTOM5 samples (with almost no expression of P2) (Table 3-5), whereas the P2 region's top 10 samples were mostly primary cells such as artery, skeletal muscle and monocyte-derived macrophages during the response to influenza virus. Samples with high P2 expression also had a smaller P1 expression.

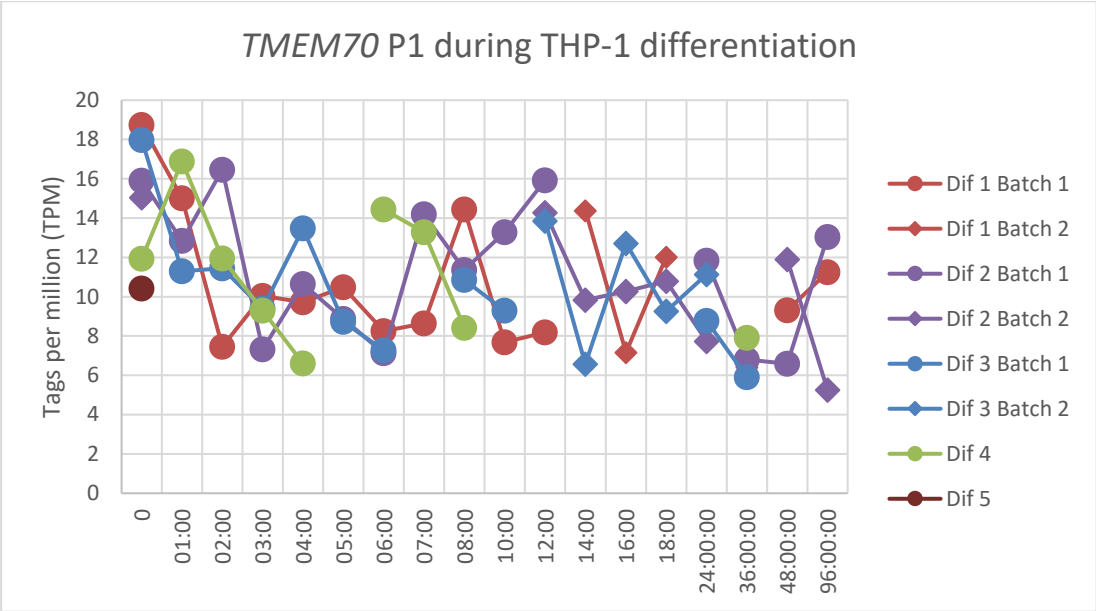

Figure 3-27 Expression of TMEM70 in area P1 in THP-1 differentiation according to CAGE sequencing. Expression is shown as normalised Tags per million (TPM) value.

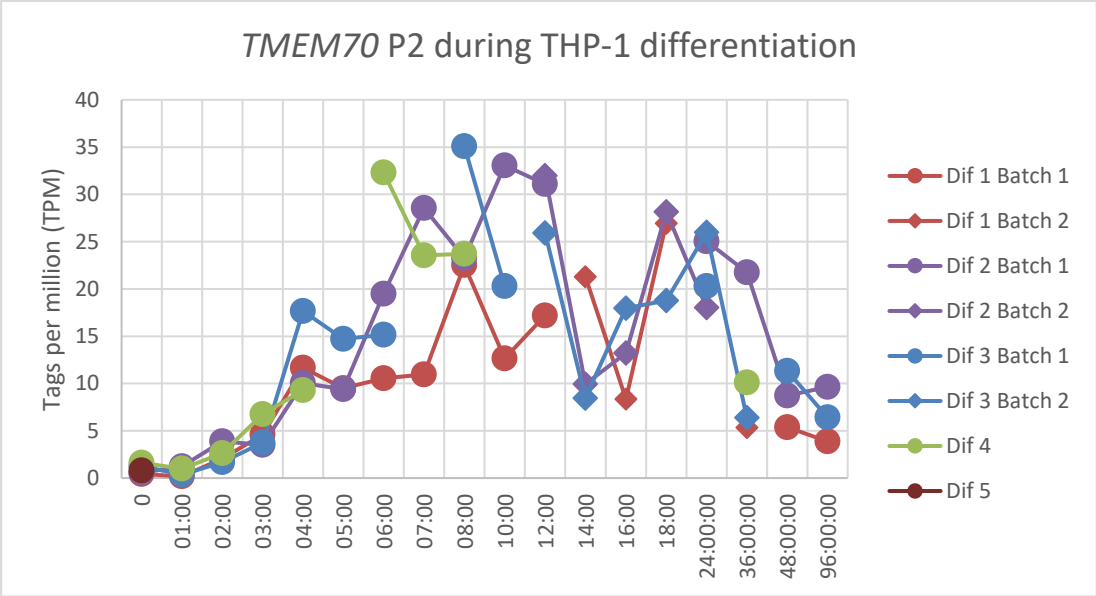

Figure 3-28 Expression of TMEM70 in area P2 in THP-1 differentiation according to CAGE sequencing. Expression is shown as normalised Tags per million (TPM) value.

	RNA samples with highest expression of P1 TMEM70	P1	P2
1	neuroepithelioma cell line:SK-N-MC	507.374	13.73
2	peripheral neuroectodermal tumor cell line:KU-SN	233.5	1.02
3	carcinoid cell line:SK-PN-DW	164.388	4.443
4	293SLAM rinderpest infection, 12hr, biol rep2	160.383	1.302
5	acute myeloid leukemia (FAB M0) cell line:KG-1	157.732	0.655
6	NK T cell leukemia cell line:KHYG-1	152.036	2.438
7	293SLAM rinderpest infection, 12hr, biol rep3	150.495	2.595
8	293SLAM rinderpest infection, 00hr, biol rep2	132.649	3.007
9	293SLAM rinderpest infection, 00hr, biol rep1	131.225	3.196
10	293SLAM rinderpest infection, 06hr, biol rep1	129.222	4.561

	RNA samples with highest expression of P2 TMEM70	P1	P2
1	artery, adult	103.031	212.053
2	skeletal muscle, adult, pool1	99.905	152.516
3	Monocyte-derived macrophages response to mock influenza infection, 24hr00min, donor2 (150 120:MI 24h)	42.208	105.05
4	Monocyte-derived macrophages response to udorn influenza infection, 02hr00min, donor1 (868 121:Ud 2h)	57.986	99.813
5	Monocyte-derived macrophages response to mock influenza infection, 24hr00min, donor1 (868 121:MI 24h)	53.296	91.497
6	Monocyte-derived macrophages response to mock influenza infection, 24hr00min, donor4 (227 121:MI 24h)	43.665	88.246
7	skeletal muscle - soleus muscle, donor1	37.249	87.377
8	Monocyte-derived macrophages response to mock influenza infection, 00hr00min, donor2 (150 120:MI 0h)	44.215	83.125
9	Monocyte-derived macrophages response to LPS, 01hr00min, donor3 (t5 Subject3)	38.682	79.739
10	Monocyte-derived macrophages response to udorn influenza infection, 02hr00min, donor4 (227 121:Ud 2h)	30.891	78.551

Table 3-5 Highest TPM values for P1 and P2 regions of TMEM70 taken from FANTOM5.

PLAGL1 promoter was also flagged by the CAGEr programme to have a shift in promoter usage between 0 and 8 hours and also between 0 and 10 hours. Further analysis revealed that *PLAGL1* has multiple transcription start sites with generally two different sized transcripts, but the one recognized by CAGEr was found at the 5' end of the gene (and thus codes for the longer transcript). This promoter could be divided into two areas, P1 (which is the one mostly used in FANTOM5 samples) and P2 (Figure 3-29). The P1 region was highly expressed in undifferentiated THP-1 samples, its activity declining as differentiation progressed (Figure 3-30). On the other hand, P2 had very low expression at 0 hour of differentiation, with its expression peaking (and being higher than P1) around hours 6 or 7 (Figure 3-31).

Figure 3-29 PLAGL1 has two promoters P1 (in blue box) and P2 (in red box) with differential expression during THP-1 differentiation. This picture is taken from Zenbu browser, with the data from this chapter (labelled Transcription Start Sites THP-1 differentiation) and FANTOM5 data remapped to hg38 (Transcription Start Sites (FANTOM5)).Transcription is from the reverse strand, from right to left of the image.

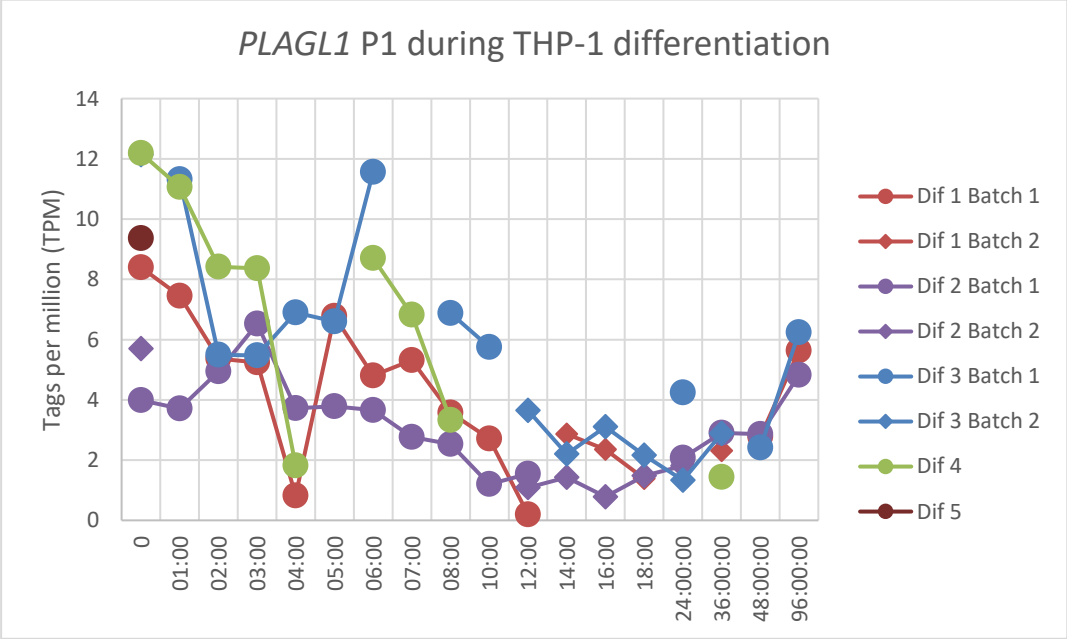

Figure 3-30 Expression of PLAGL1 in area P1 in THP-1 differentiation according to CAGE sequencing. Expression is shown as normalised Tags per million (TPM) value.

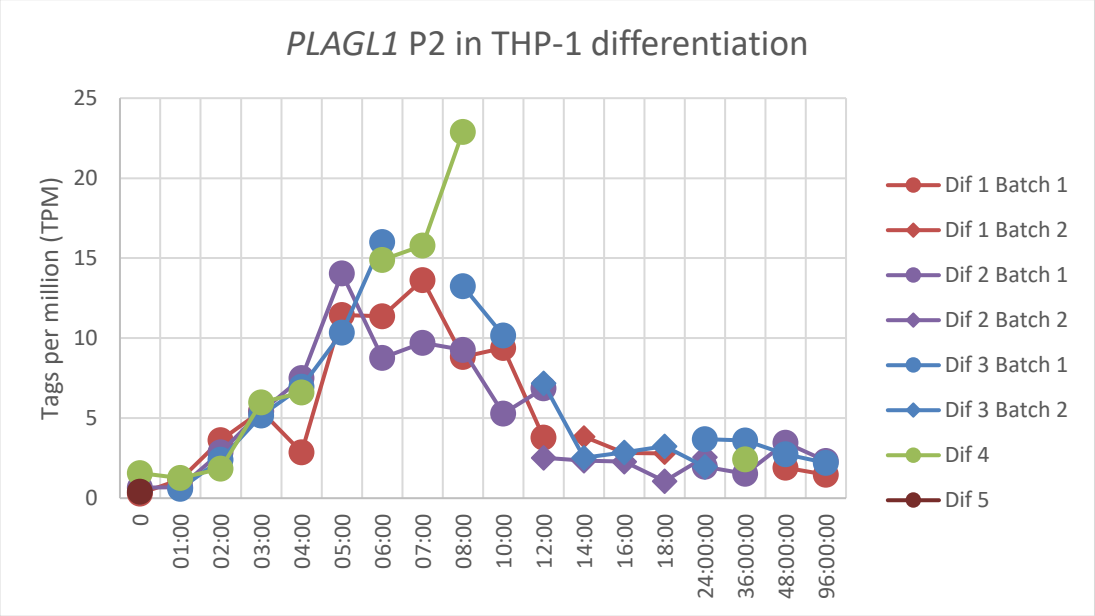

Figure 3-31 Expression of PLAGL10 in area P2 in THP-1 differentiation according to CAGE sequencing. Expression is shown as normalised Tags per million (TPM) value.

	RNA samples with highest expression of P1 PLAGL1	**P1**	**P2**
1	Neutrophils, donor1	69.21	0.304
2	common myeloid progenitor CMP, donor1	63.591	0
3	promyelocytes/myelocytes PMC, donor1	63.187	0
4	Neutrophils, donor3	56.617	0
5	granulocyte macrophage progenitor, donor1	55.263	0
6	Eosinophils, donor2	51.824	0
7	Eosinophils, donor3	47.367	0
8	b cell line:RPMI1788	41.132	2.13
9	CD133+ stem cells - adult bone marrow derived, pool1	38.932	0
10	CD14+CD16- Monocytes, donor3	37.981	0.136
	RNA samples with highest expression of P2 PLAGL1	**P1**	**P2**
1	Monocyte-derived macrophages response to LPS, 16hr, donor1 (t20 Subject1)	9.946	9.946
2	COBL-a rinderpest infection, 12hr, biol rep3	10.244	5.885
3	COBL-a rinderpest(-C) infection, 12hr, biol rep1	12.688	5.767
4	COBL-a rinderpest infection, 48hr, biol rep1	20.771	5.36
5	COBL-a rinderpest(-C) infection, 06hr, biol rep1	9.355	5.346
6	cord blood derived cell line:COBL-a untreated	17.753	5.222
7	COBL-a rinderpest(-C) infection, 06hr, biol rep2	7.715	5.143
8	COBL-a rinderpest infection, 06hr, biol rep3	10.991	5.088
9	COBL-a rinderpest(-C) infection, 24hr, biol rep1	11.355	5.047
10	COBL-a rinderpest(-C) infection, 24hr, biol rep2	10.336	5.011

Table 3-6 Hightest TPM values samples for P1 and P2 regions of PLAGL1 taken from FANTOM5.

The FANTOM5 samples with highest expression in P1 *PLAGL1* area were mostly primary hematopoietic cells, which simultaneously had almost no P2 expression. On the other hand, P2 was not really expressed highly except for monocyte-derived macrophages treated with LPS and the COBL-a cell line, untreated and treated with viral infection. COBL-a cell line is a relatively new cell line, established from human umbilical cord blood specifically to study measles virus (Kobune *et al*., 2007).

3.3.6 Chromatin modifying enzymes gene expression in THP-1 differentiation

As stated in the introduction (section 1.2), chromatin modifying enzymes play a regulatory role in cell differentiation, allowing genes to be switched on and off by opening or closing chromatin. During the exploration of FANTOM4 time course data, two genes encoding chromatin modifying enzymes (*KDM6B* and *USP12*) were found to be upregulated during the THP-1 stimulation with PMA (section 1.4.1). In this new tighter time course, this was validated in multiple ways.

The FANTOM4 data showed that deubiquitinases gene *USP12* was upregulated around hour 4 of the THP-1 time course. The new data from the present study show that the highest level of expression was actually at around the 5th hour of the differentiation (Figure 3-32). This result was also validated by qRT-PCR (Figure 3-33). There was considerable batch-to-batch variability, with Diff2 Batch 1 having an expression level less than half that of Diff3 Batch 1. However, by following each batch separately it can be seen that *USP12* expression was elevated between 4 and 12 hours in all cases.

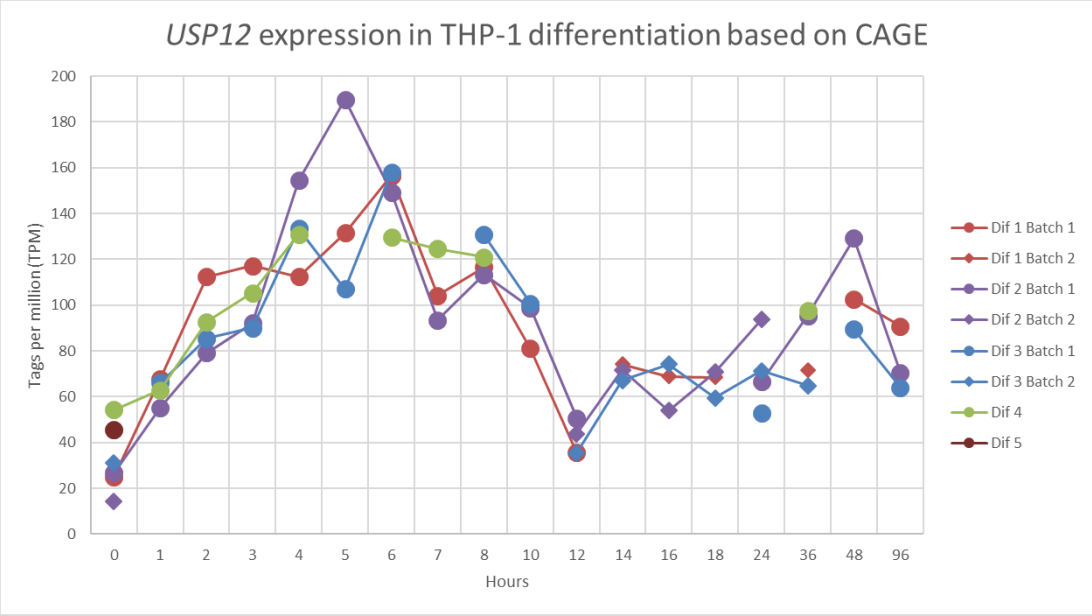

Figure 3-32 Expression of USP12 in THP-1 differentiation according to CAGE sequencing. Expression is shown as normalised Tags per million (TPM) value.

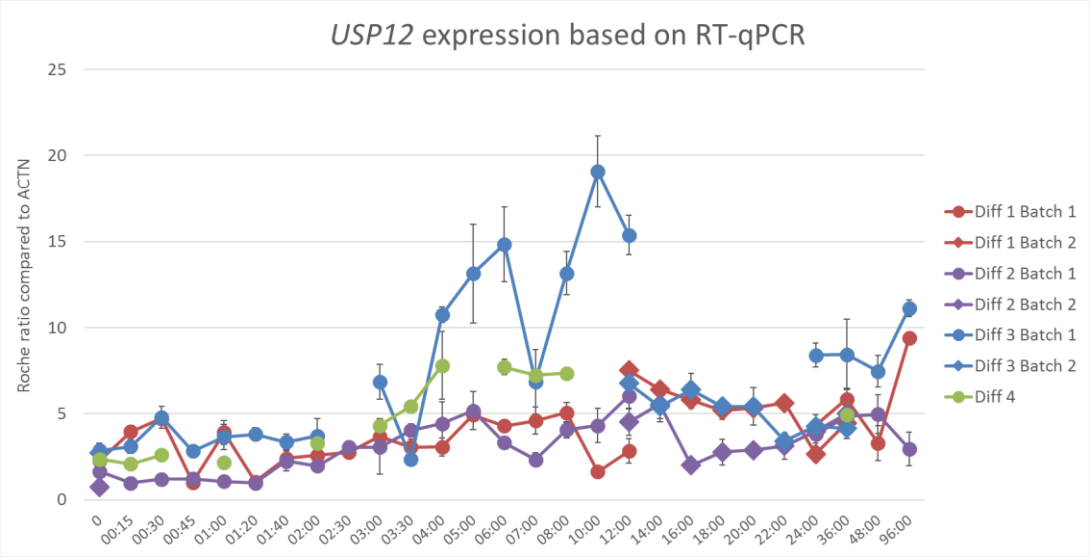

Figure 3-33 USP12 expression in the whole THP-1 differentiation (all time points taken) based on qRT-PCR. Pictured here is the USP12 expression normalised to ACTN expression and error bars showing standard error (calculated by the by Roche LightCycler 480 software).

CAGE sequencing results were also used to look at other H2A deubiquitinases, *BAP1* and *USP16* and histone demethylases *KDM6B* and *UTY*. The detailed time course allowed detection of small fluctuations in gene expression, and thus it was discovered that *BAP1* has a cyclic expression pattern, peaking three times in the first 12 hours of

the time course (Figure 3-34), which was not previously detected in FANTOM4 data (Figure 1-3).

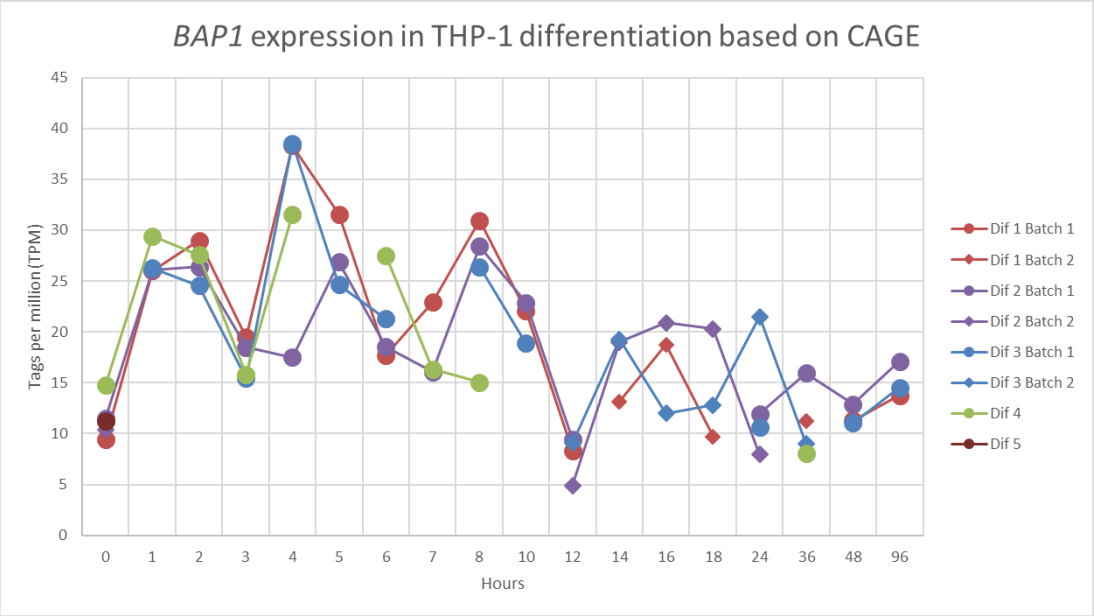

Figure 3-34 Expression of BAP1 in THP-1 differentiation determined by CAGE sequencing. Expression is shown as normalised Tags per million (TPM) value.

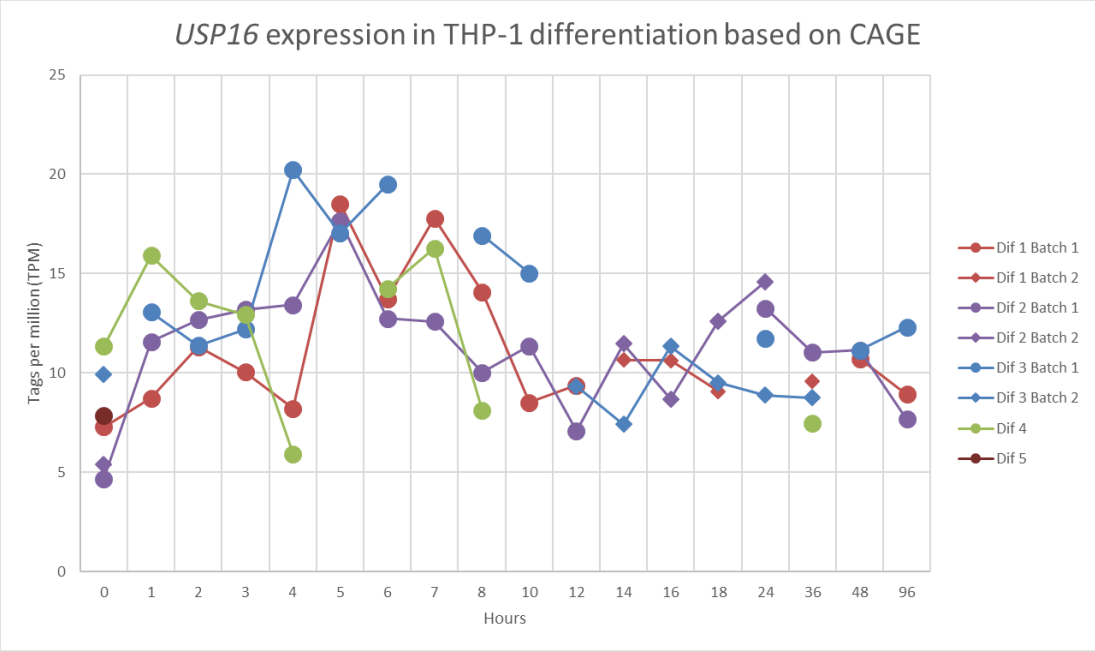

Figure 3-35 Expression of USP16 in THP-1 differentiation according to CAGE sequencing. Expression is shown as normalised Tags per million (TPM) value.

The level of *USP16* did not change much throughout the time course (Figure 3-35), but *KDM6B* was found to be upregulated, just as in FANTOM4, with a steep increase

during the first few hours (Figure 3-36). The *KDM6B* data was also supported by qRT-PCR analysis (Figure 3-37). *UTY* expression was found to be upregulated during the time course, but only towards the end, as the cells were becoming more macrophage-like (Figure 3-38).

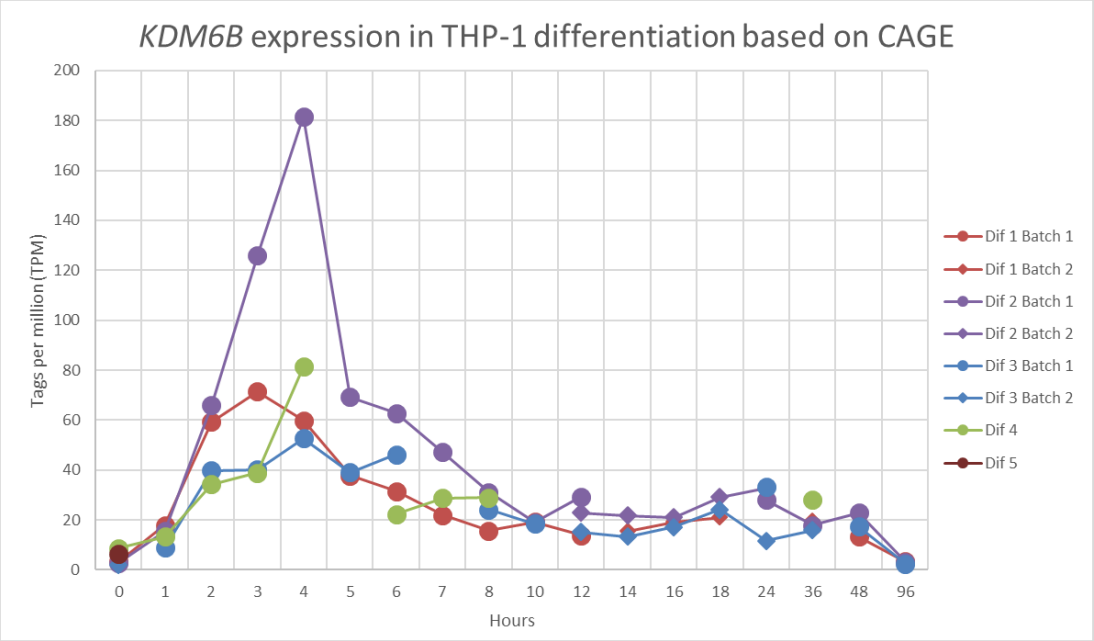

Figure 3-36 Expression of KDM6B in THP-1 differentiation according to CAGE sequencing. Expression is shown as normalised Tags per million (TPM) value.

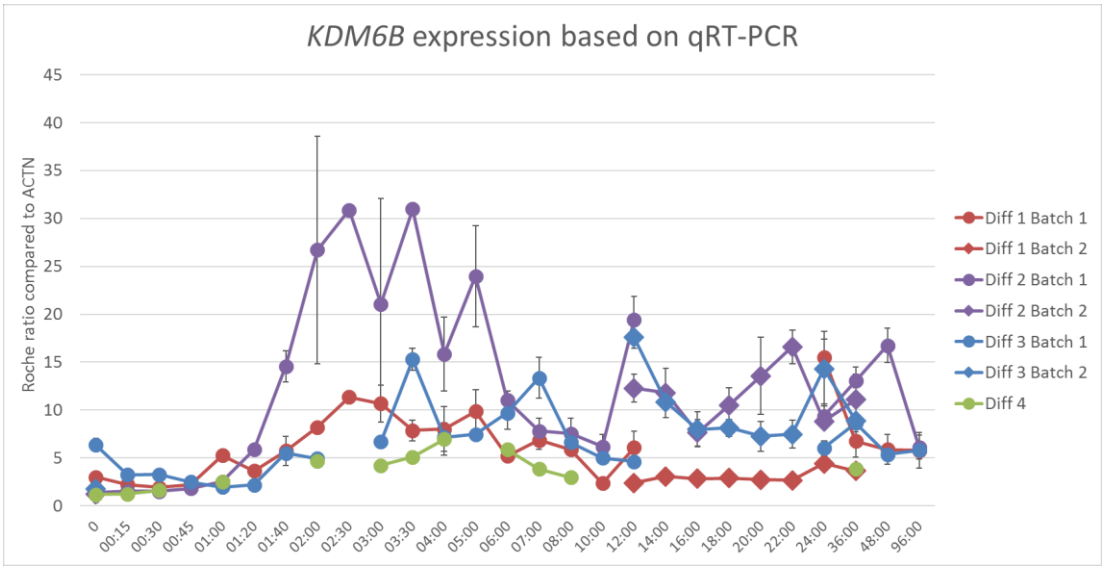

Figure 3-37 KDM6B expression in the whole THP-1 differentiation (all time points taken) based on qRT-PCR. Pictured here is the KDM6B expression normalised to ACTN expression and error bars showing standard error (calculated by the by Roche LightCycler 480 software).

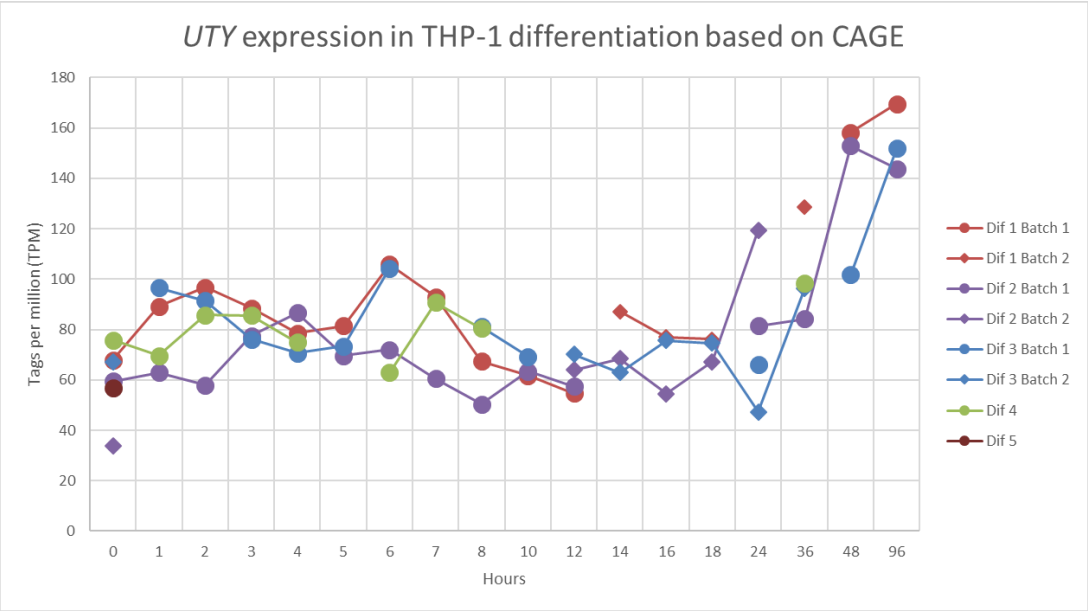

Figure 3-38 Expression of UTY in THP-1 differentiation according to CAGE sequencing. Expression is shown as normalised Tags per million (TPM) value.

These five deubiquitinases and demethylases were analysed in the Miru network created in section 3.3.3 using a Pearson correlation coefficient threshold of 0.75, and four of them were found to be in different clusters. *USP16* was not included in any cluster. *BAP1* was found in Cluster 15 (100 nodes; Figure 3-39), with other genes coding for proteins with K homology domain, and other ubiquitin-associated protein genes (Table 3-7). Interestingly the averaged expression profile for this cluster (Figure 3-39) also seems to oscillate in the same way as *BAP1* (Figure 3-34) suggesting that this cluster contains genes with a cyclical regulation during differentiation. *USP12* was found in Cluster 36 (42 nodes), which was also enriched for ubiquitin pathway proteins (Figure 3-39, Table 3-7). *KDM6B* was in the small Cluster 122 (11 nodes), which was only enriched in cell junction GO terms (Figure 3-39, Table 3-7). *UTY* was grouped into Cluster 2, as analysed in section 3.3.3 (Figure 3-39).

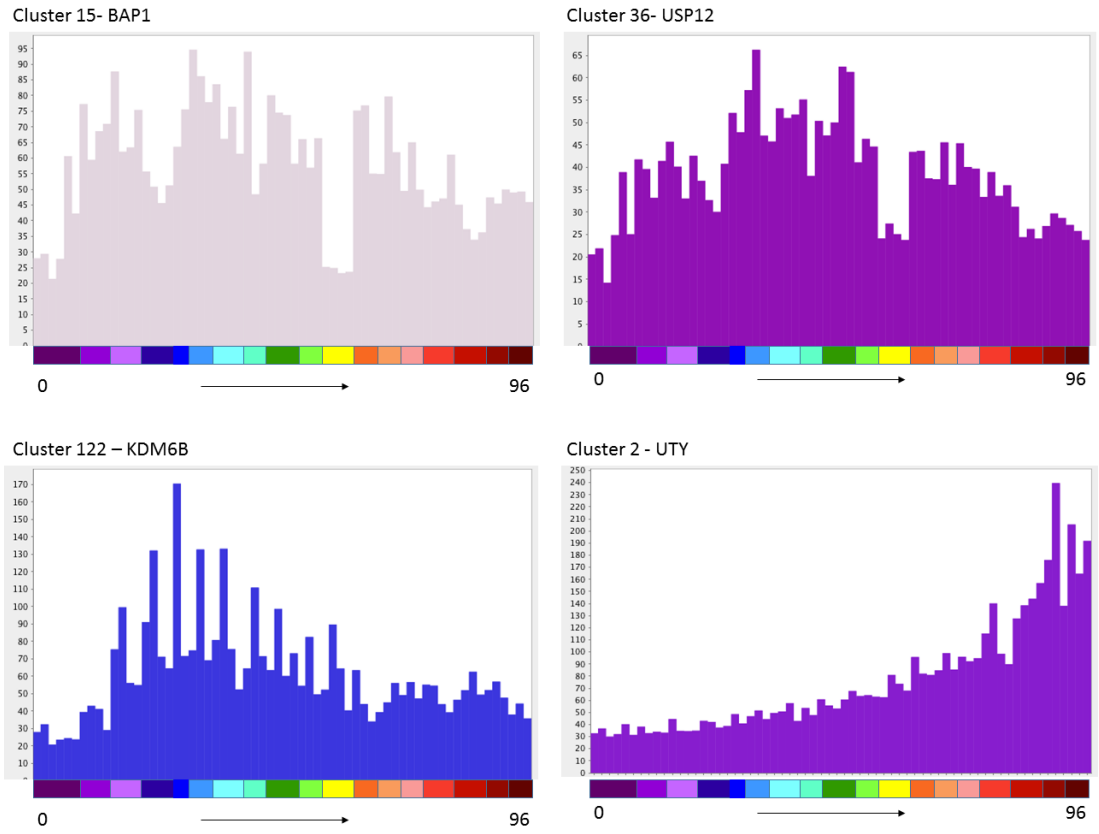

Figure 3-39 Averaged expression profiles of different clusters containing expression of chromatin modifying enzymes found with Miru programme at correlation coefficient ≥ 0.75. Y axis is in Tags per million (TPM), X axis shows the individual colour-coded sample types, from 0 hours to 96 hours. Colours are the same as in Figure 3-16.

Cluster 15		**GO terms**	**Group score**
	Group 1	K Homology domain	1.81
	Group 2	Ubl conjugation pathway	1.67
	Group 3	Protein import into nucleus	1.58
Cluster 36		**GO terms**	**Group score**
	Group 1	Ubiquitin-dependent protein catabolic process	1.74
	Group 2	Ubiquitin protein ligase binding	1.62
Cluster 122		**GO terms**	**Group score**
	Group 1	Cell junction	1.1

Table 3-7 DAVID GO term annotation for clusters 15, 36 and 155 found at a correlation coefficient ≥ 0.75 using Miru programme.

3.4 Discussion

3.4.1 Chapter overview

The main point this chapter demonstrates is the wide range of information that can be ascertained from a detailed CAGE time course study. This chapter includes data from a time course of THP-1 differentiation similar to the FANTOM4 study (Suzuki *et al* 2009), but with an increased number of time points (18) especially at the early stages, as well as deeper sequencing of the samples. Firstly, THP-1 differentiating into macrophage-like cells were confirmed based on morphological changes, and the cells were found to become adherent around hour 8 to 10, developing a macrophage-like phenotype. Next, cell cycle analysis was performed, showing that the differentiating THP-1 cells arrest their cell cycle in G1/S phase at around 10 hours of the time course, consistent with the cessation of proliferation as differentiation proceeds. The transcriptomic analysis of this monocyte to macrophage differentiation time course was carried out with at least three biological replicates per time point, using CAGE technology. Eight samples were multiplexed on one lane of the flow cell used in the sequencing machine, which resulted in around 9 million mapped reads per sample. The CAGE results were subsequently used to confirm macrophage differentiation of the samples, which were also validated by qRT-PCR on the RNA used for CAGE library preparation. The difference between FANTOM4 (Suzuki *et al* 2009) and this study was mainly in the number of time points and depth of sequencing, which revealed the continuous waves of gene activation hour to hour post PMA stimulation. This was achieved using Miru programme by analysis of correlation between the samples and the CTSS. Further information could also be extracted looking at enhancer transcriptional activity and analysis of shifting promoters using CAGEr programme. This work also looked at gene expression changes of epigenetic modifiers in the THP-1 cells during differentiation.

3.4.2 THP-1 cells as a model for monocyte to macrophage differentiation

Since their isolation nearly 40 years ago, THP- 1 cells have been widely used as a model cell line for monocyte and macrophage cells (Tsuchiya *et al*., 1980). As

mentioned in section 1.3.2, THP-1 cells can be stimulated with various compounds, with the phorbol ester phorbol-12-myristate-13-acetate (PMA) as the most potent one. An important question is whether THP-1 cells are a good model for studying monocytes and macrophages. Firstly, they are of pro-monocytic origin, so they are not considered to be mature monocytes (Bosshart and Heinzelmann, 2016). Differentiated THP-1 cells also did not show comparable macrophage differentiation to primary bone marrow derived macrophages (Daigneault *et al*., 2010). They were derived from a monocytic leukaemia and have therefore undergone changes associated with the escape from growth controls typical of cancer cells. However, they do have an almost normal human karyotype, with few deletions and one trisomy of chromosome 8 (Adati *et al*., 2009). The THP-1 cells used throughout this study were of clonal lineage, in which a clone (clone number 5) was chosen by FANTOM4 collaboration as the cells were most uniform in differentiating and becoming adherent (Suzuki *et al*., 2009). This is in contrast with cells originally isolated (and the ones available at ATCC; http://www.atcc.org, product ATCC TIB-202), which are heterogeneous in response to PMA (Suzuki *et al* 2009). However, even clone 5 THP-1 cells were found to have different cell morphology at certain time points (as seen in Figure 3-3). Since the majority of the cells (up to 98%) follow a similar pattern during differentiation, the gene expression results collected from the RNA of around 5 million cells per time point should not be adversely affected by 2% of more or less differentiated cells. Having an extensive set of time points during THP-1 differentiation, it is now possible to take FANTOM5 data from primary monocytes and primary macrophages, and determine which stage of THP-1 differentiation is the most transcriptionally similar to that of primary cells.

3.4.3 Stimulation of THP-1 cells and their effect on cell cycle

As mentioned in the Introduction section 1.3.2, THP-1 cells can be stimulated in multiple ways to induce differentiation. In this repeated study of FANTOM4, the same conditions were used to allow for comparison. PMA (a phorbol ester) was added to the cell medium at concentration of 30 ng/ml (48.6 nM). Other studies have used concentrations from 5 to 100 ng/ml (Park *et al*., 2007). PMA at high concentrations (> 50 nM) had been shown to stimulate the THP-1 cells in a way similar to the response

of monocyte derived macrophages responding to lipopolysaccharide (LPS), and their subsequent response to LPS was stronger than in THP-1 cells stimulated with lower PMA concentrations (Lund *et al*., 2016). It is possible that the lower concentration of PMA used in the present study was suboptimal for the full development of the macrophage phenotype. However, analysis of *MYB*, which is associated with progenitor cells and declined during the time course, and *CD14*, a macrophage marker which increased during the time course, was consistent with the transition from monocyte to macrophage (see section 3.4.4).

Based on the cell cycle data, stimulation with PMA of THP-1 causes the cells to stop proliferating, with cells arrested at the G1/S checkpoint at around 8-10 hours. Cells at later stages of the cycle completed mitosis (Figure 3-5). The proportion of G2 phase cells fell between 24 hours and 48 hours – presumably as cells moved from G2 back to G1 phase after cell division. Even after 48 hours of starting the THP-1 differentiation, there was a proportion of cells in G2 phase. Given the fact that no cells stayed in S phase after 24 hours, it could mean that the cells stayed in G2 phase (taking them longer to exit than usual). It could also indicate the imprecision of the cell cycle analysis, as just divided macrophages might not have completely finished telophase and cytokinesis so that single cells would appear to have four copies of the genome. Every possible care was taken to eliminate doublet cells from flow cytometry analysis (as cells tend to stick together), but the small proportion of cells in G2 could simply show cells with two nuclei.

The cell cycle arrest is also visible in data from section 3.3.3, where genes for histone proteins, DNA replication and mitochondria were found in Cluster 1, for which the overall profile expression is downregulated as the differentiation progresses.

Detailed study of this PMA-associated cell cycle arrest in THP-1 cells could also determine novel targets for oncological drugs against proliferation in leukaemia.

3.4.4 Validation of progression of THP-1 differentiation

The THP-1 differentiation time course for each biological replicate was confirmed initially by morphological changes in the cells (section 3.3.1.1), and then by testing the level of expression of relevant markers by CAGE sequencing and qRT-PCR.

The two markers (*MYB* and *CD14*) were chosen to represent two different patterns of gene expression, as *MYB* levels are affected by PMA within first few hours (Suzuki *et al*., 2009) and *CD14* levels rise only after few days' of PMA stimulation as a marker of macrophages (Daigneault *et al*., 2010). *MYB* (found in Cluster 1 in section 3.3.3) was downregulated within the first hour of differentiation in all biological replicates. *CD14* (found in Cluster 7 in section 3.3.3) was very highly upregulated in the last two time points according to CAGE sequencing in all replicates. The CAGE data was validated by qRT-PCR, which contained all the original time points (including time points every 15 minutes in the first hour) and provided higher resolution for gene expression in the first hours of the time course. *MYB* expression in qRT-PCR was less uniform within biological replicates than in CAGE sequencing data, which could be due to technical errors. All qRT-PCR results also include samples which were excluded from CAGE study due to their low quality of RNA, which could affect the result.

3.4.5 CAGE technology limitations

Initially, three biological replicates of the time course had been planned for CAGE analysis. However, since some RNA samples exhibited bad quality (low RINe score) or quantity, two more replicate experiments were conducted. 12 CAGE libraries were successfully created, but some libraries had a large peak at 78 bp suggesting primer dimer. The occurrence of the primer dimers decreased the number of final reads that were mapped to the genome, so the higher the ratio of CAGE peak vs primer peak, the better was the outcome of the number of final reads used in the analysis. The reason for the difference in the amount of primer contamination in each run is unknown, as all libraries used the same quantity of RNA per sample. However, another CAGE preparation in the laboratory showed that using 1 μg of RNA per sample produced a

large amount of primer contamination (Dr Sara Clohisey, personal correspondence, 2017). Even though the protocol is optimised for 1-5 μg of RNA per sample (Takahashi *et al*., 2012), the use of primers in samples in the lower range of RNA might have to be optimised in future studies. There were also larger peaks (> 100 bp) present in CAGE libraries, which were of unknown origin (Figure 3-7). After several libraries were sequenced, all enzyme stocks (particularly restriction endonuclease *Eco*P15I) were replaced to rule out enzyme degradation, but the presence of the larger peaks was not significantly reduced.

Another hindrance in the analysis of the CAGE sequences was the low number of mapped reads on average in certain samples with the CTT barcode (Figure 3-9). As the problem was discovered during the process of making the later libraries, new barcodes were produced from a frozen master stock of oligonucleotides. This improved the number of mapped reads, but it is unclear why the CTT barcode was so unstable. The first four barcodes (ACG, GAT, CTT and ATG) were created about a year before the start of the CAGE library production for this project. The rest were produced at the start of the project. All barcodes were frozen and aliquoted to prevent excessive amount of freeze-thaw cycles, but that did not prevent the malfunction. It might be that CTT sequence (the only difference among the barcodes) is particularly unstable or the oligonucleotides were incorrectly manufactured. Future work should be focused on testing the performance of all barcodes before any other large-scale project is initiated.

It was found that there was a large variability in the number of reads between different samples, from 33 million of mapped reads to just below 3 million (Figure 3-10). Barcode occurrence, quality and quantity in samples with small or large number of reads were checked, but no obvious differences were detected. This difference might be due to chance, which could be tested by running the same CAGE library on two lanes of the Illumina sequencing flow cell.

Analysis of CAGE data was made easier by using the CAGEr package, which also removes the G-nucleotide bias introduced at the beginning of reads by CAGE primers

(Haberle *et al*., 2015). The normalised CTSS reads were then uploaded to Zenbu for convenient visualisation of individual genes. For further analysis, it was decided to use this package to group the individual CTSS into 'clusters of CTSS' based on CTSS distance (max distance 20 bp). The number of CTSS clusters was then reduced to about half (from around 30,000 CTSS groups in each sample to 15,341 overall) when aggregating all CTSS groups for each of the 66 samples into a single file of non-overlapping CTSS clusters. This was due to the more stringent conditions, when only CTSS groups in which at least one sample contained TPM of at least 5 were included. This ensured that only the most significant results were considered, and the amount of possible transcriptional noise was kept to minimum. FANTOM4 identified 29,857 promoters and 14,607 wider promoter regions substantially expressed at least once (Suzuki *et al*., 2009); which is around the same number as the number of aggregated clusters in this study. Further comparison with FANTOM4 promoter region is needed to complete the analysis.

CAGE technology was chosen, as opposed to RNAseq or microarray analysis, to allow for precise comparison with FANTOM4 data and primary cell data from FANTOM5. CAGE deep sequencing is also sensitive enough to look for novel transcriptomic features, such as enhancers and shifting promoters, which technology such as microarray is unable to do. Analysis for CAGE data is also easier due to the fact that each capped mRNA only produces one CAGE read, as opposed to RNAseq, where the number of reads is dependent on the length of the transcript. On the other hand, RNAseq is able to look for new splicing variants. The cost of making the CAGE library and sequencing 8 samples on one lane of the Illumina 2500 flow cell was lower than that of RNAseq or expression microarray (Prof Kim Summers, personal correspondence, 2016), which allowed a substantial number of samples to be analysed with a limited budget.

3.4.6 Waves of gene repression and activation during the THP-1 time course

Sample to sample correlation data showed that there were at least two samples which did not correlate at all with their biological replicates (Figure 3-15). These two samples, both for the 4 hour time point, were shown to be most similar to much later

time points in the differentiation. When these two samples were removed, the resulting sample-to-sample correlation showed clearer progression of the time point samples from 0 hour to 96 hours (Figure 3-16). Thus, these two samples were excluded from further cluster analysis. All 66 samples were used in other analyses, but the results were investigated with caution. FANTOM4 expression measurements were stated to be 'noisy' (Suzuki *et al*., 2009), which could be seen by expression analysis of few genes (e.g. section 1.4.1). The samples in this study were correlated comparably, but there were still certain inconsistent results for the set of 64 samples, such as the 8 hour time point being most correlated to the 16 hour time point. It is not clear why these samples have such a poor outcome in the correlation graphs, as not all of them had low number of mapped reads (as stated in section 3.3.3) and the quality of starting RNA was similarly good for all reads used in CAGE sequencing.

Network analysis of CTSS expression patterns was performed with Pearson's correlation coefficient of 0.75. Higher correlation coefficients were examined, as genes within the clusters were more highly correlated but they excluded many CTSS (nodes). For example, some chromatin modifying enzymes analysed in section 3.3.6 were omitted from the analysis at higher coefficients. Clusters formed with a higher correlation coefficient were similar in various expression patterns to the clusters presented in this study (section 3.3.3 and Appendix 8.4), but subsequent GO term annotation by DAVID resulted in higher p-values due to their small size. The network analysis at 0.75 also showed a main element with the two largest clusters (Clusters 1 and 2) being apart.

Eight different clusters were chosen for further analysis as they represented the transition in gene expression during THP-1 differentiation (Figure 3-17 and Figure 3-18). This analysis showed a clear progression from Cluster 1 (least differentiated) to Cluster 7 (most differentiated), consistent with the notion that clusters were at the opposite ends of the differentiation spectrum (Figure 3-17). As the pattern of activation progressed through the time course, it can be clearly seen that firstly early response genes such as various transcription factors were activated, which in turn would stimulate the transcription of macrophage-specific genes responsible for functions

such as phagocytosis and lysosome activity. These detailed transcriptional waves could only be detected by a tight time course. The next steps in analysis would be to study these transcription factors, compare them with already known networks from FANTOM4 and match them with possible gene targets.

Further investigation is needed for other interesting clusters not currently included in the analysis, for example clusters with genes that are downregulated by the PMA stimulation, but returning back to the basal pre-stimulation level of expression by the end of the 4 day period. Other clusters recognised by Miru usually included sample-specific gene expression, which was not replicated in the other samples of the same time point. Cluster 23, which was included in the analysis due to its high GO enrichment, has this problem, as the replicates show different levels of expression. Investigation into the reasons for this high GO enrichment for antiviral defence could help to understand why a well-established cell line such as THP-1 can produce such variable results.

3.4.7 Novel features of THP-1 differentiation found by CAGE

3.4.7.1 Enhancers

As detailed in Andersson *et al*. (2014), deep CAGE sequencing is able to detect enhancer activity. The example presented in this study shows the enhancers for the transcription factor gene *MYB*. One of the enhancers near the *MYB* promoter was annotated in FANTOM5, but there were bidirectional transcripts located further along the gene, for which expression was correlated to the main *MYB* promoter. The positive strand transcript was annotated in the Ensembl database, as a *MYB* transcript targeted by nonsense mediated decay. When checking the expression for these transcripts, the positive strand transcript was most highly upregulated in various cancers and leukaemia. This transcript could therefore be cancer-specific and could potentially be a therapeutic target for RNAi use or as a diagnostic marker.

Further analysis of all enhancers in the time course will be continued, firstly by finding a way to correctly annotate them as enhancers (based on the study of Andersson *et al* (2014)), and then by comparing the expression with the target in the time course. These

potential enhancers can only be detected because of the tight time course and deep sequencing, as their expression is low and may be quite transient during the differentiation process, as demonstrated for other time course data by Arner *et al.* (2015). Enhancers could be further validated by intersection with publicly available data sets of transcription start sites and chromatin marks, such as FANTOM5 and ENCODE (ENCODE, 2012, Arner *et al*., 2015, Forrest *et al*., 2014).

3.4.7.2 Shifting promoters

As mentioned above, one of the advantages of using CAGE sequencing is that it very precisely determines where transcription starts. As documented in Haberle *et al*. (2014), promoters have various structures and can use different regions of the DNA upstream of the coding sequence at different times. Using CAGE technology in a tight time course such as this one, allows the discovery of finer differences in gene expression and promoter usage, than, for example, comparing static samples of widely different developmental origin (Arner *et al.*, 2015, Forrest *et al*., 2014). Information about promoter switching could be potentially useful in determining a drug target, which might give specificity to the cancer cells without impacting normal expression in primary cells. Using the CAGEr programme to look at different start sites of the same promoter area was quite effective, as multiple sites were flagged by the software. In this study, three interesting areas were chosen.

There was a difference in the promoter usage by the *ARHGAP30* gene between undifferentiated and differentiated THP-1. Further analysis showed that the main area of the promoter, which codes for the longer *ARHGAP30* transcript, was ubiquitously expressed throughout the differentiation, but there was a spike of expression of the *ARHGAP30* shorter transcript, which was only expressed at the end of the time course, in macrophage-like THP-1 cells. Little is known about this gene (there were only 5 articles on Pubmed as of November 2017), but it codes for a Rho GTPase, a family of genes which are important in signal transduction pathways that control cell migration, proliferation adhesion and other functions (Naji *et al*., 2011). ARHGAP30 also promotes p53 acetylation and thus is a possible prognostic marker in colorectal cancer (Wang *et al*., 2014). This acetylation of p53 is highly specific for the full length

isoform, not the shorter one without the glutamate-rich region. This seems to be the only functional difference between these two transcripts. However, THP-1 cells do not express p53 (Sugimoto *et al*., 1992), so it is not currently clear whether the differentiation-dependent expression of the shorter transcript is due to the p53 interaction. Interestingly, the *ARHGAP30* longer transcript was mostly expressed in immune cells (such as T-cells), but these cells with high expression levels do not produce the shorter transcript at the same levels. However, the expression of the shorter transcript was abundant in another time course in FANTOM5, monocyte-derived macrophages and their response to LPS. This could imply that this shorter transcript has a similar role in LPS treatment as in THP-1 differentiation. Further research is needed to unravel the difference between these two transcripts and why their expression is tightly regulated in the differentiation study.

Two other examples of shifting promoters were in the *TMEM70* and *PLAGL1* genes. *TMEM70* is mutated in mitochondrial complex V (ATP synthase) deficiency (OMIM 614052; (Cizkova *et al*., 2008)), associated with lactic acidosis, cardiovascular abnormalities, dysmorphic features and death in the neonatal period. TMEM70 may be important for cells' survival, although there is no clear link with macrophage function. The *TMEM70* promoter had two areas with differential expression – P1 which was at the beginning of the transcripts, and did not change throughout differentiation, and P2 which was found at the beginning of the translated region, and was upregulated few hours after PMA stimulation following which its expression decreased to almost pre-differentiation levels. No alternative transcripts based on the genomic region position of P2 area have been identified.

For *PLAGL1* the P2 area, where expression was also differentiation-specific, was outside the annotated transcript region so there were no known transcripts initiating from this promoter. *PLAGL1* is an imprinted gene (Kamiya *et al*., 2000), which codes for a transcription factor that acts as a tumour suppressor (Spengler *et al*., 1997, Varrault *et al*., 1998, Bilanges *et al*., 2001), and has at least 5 other promoters found to date (Smith *et al*., 2017). There are alternately spliced transcripts which use different promoters and vary in the 5 untranslated region and whether coding exon 7 is included

in the transcript (Valleley *et al*., 2007). This plasticity of transcript structure suggests that P2 identified here could be a real variant.

In both cases, *TMEM70* and *PLAGL1*, the differentiation-specific area of the promoter region was also expressed in other FANTOM5 samples, so the presence of two promoter regions is unlikely to be due to the incorrect mapping of the sequences. In fact, both of the regions which are specific for differentiating THP-1 cells (plus the P2 area of *ARHGAP30*), were upregulated in time course samples of immune cells, both primary cells and cell lines. It might be that these new transcripts are highly immune-cells specific and thus had not been annotated at the time of writing. The P2 area was of *TMEM70* was identified as a second promoter in FANTOM5 database, but *PLAGL1* P2 region was not annotated as a promoter. Further analysis of these transcripts and TSS could uncover specific monocyte/macrophage transcriptional activity controlled by differential promoter usage, which could not be easily found by RNAseq. Since PLAGL1 has been identified as a tumour suppressor further investigation of this gene and the alternative promoter could lead to a possible tumour specific drug target.

One of the disadvantages of this analysis is that the programme only compares CTSS usage within already defined clusters. These clusters of CTSS were aggregated with the distance apart set at a maximum of 100 bp (see section 2.5.4). Thus, if there is a different CTSS of the same gene which is further away than 100 bp, it is flagged as a different CTSS cluster and its expression is not compared with other promoters for that gene. For example, the distance between p1@BAP1 and p2@BAP1 of *BAP1* as shown by FANTOM5 database is over 4000 bp, therefore the p2@BAP1 would not be identified as a shifting promoter. Further analysis could include CTSS clustering in a different way or with longer distance between the CTSS, but this could produce more false positives. The other method is to annotate the promoters the same way as FANTOM5 did, with assigning promoter numbers to all genes, as the analysis in this study did not use this approach to look at shifting promoters. The next step would be to compare the usage of promoters between individual time points, although this process is more time consuming than using the CAGEr method.

3.4.8 Role of chromatin modifying enzymes – epigenetic regulation

The expression of five genes, three encoding H2A deubiquitinases *USP12, USP16* and *BAP1*; and two encoding H3K27me3 demethylases *KDM6B* and (potentially) *UTY* was analysed in the tighter time course. More time points have shown higher resolution of gene expression patterns, for example *BAP1* transcription exhibits cyclical behaviour not previously noticed. qRT-PCR was also used to validate the results of the two upregulated genes from FANTOM4, *KDM6B* and *USP12*, but *USP12* qRT-PCR did not show the same trend as in CAGE, where only a very small upregulation after few hours was seen. qRT-PCR of *KDM6B* also showed high variability between replicates. All qPCR results were also normalised against a housekeeping gene, *ACTN*, so the results should be comparable to normalised CAGE data. It could be due to the fact that CAGE picks up all 5' ends of the transcripts, but qRT-PCR primers were selected for the middle of the gene. As mentioned in section 3.4.4, the qRT-PCR graphs contain all samples, even those for the time points not chosen for CAGE and samples with low RNA quality, which could affect the results.

USP16 expression did not change throughout the differentiation, and thus it was not correlated with other genes and the Miru programme did not include the gene in any cluster. Other genes, however, were included in various clusters. *UTY* was found to be upregulated at the end of the time course, and thus it belonged to Cluster 2 (Figure 3-18). *BAP1* was included in Cluster 15, which shows an oscillating expression pattern but has low expression in the 12th hour, which looks like a sequencing artefact. The other genes' clusters (*KDM6B* and *USP12*) also seem to be influenced more by high expression in a few samples, which may be due to technical problems rather than real biological issues.

Further investigation of genes co-expressed with these genes in the clusters could provide clues to whether the difference in biological replicates were technical errors. Differential expression of chromatin modifying enzymes could rapidly change the transcription of other genes and thus increase the variability in biological replicates. Expanding knowledge about how these processes can occur can help in battling cancer

heterogeneity and metastasis formation. The analysis of expression of these chromatin modifying enzymes inspired further studies of their functional roles in later chapters.

3.4.9 Conclusions and further work

It is not a simple and straightforward process for a cell to change cell type during differentiation to perform different functions. The transition from primary monocyte to macrophage takes multiple days and, the same is true for the THP-1 model. However, as is shown by the data in this chapter, and supported by multiple publications as discussed above, there is an almost immediate change of transcriptomic expression after stimulation. The FANTOM4 consortium made a major contribution characterising a transcription factor network and regulatory site prediction given the limited time points used in the study (Suzuki *et al* 2009). However, further work characterising the transcriptomic changes in detail is needed for precise understanding how the activation and repression of gene expression is regulated. The role of chromatin modifying enzymes in this change should be further emphasized and explored. This study lays the ground work for further bioinformatic analysis of the deepCAGE sequencing, which could help also help find new targets for leukaemia therapy. Due to the fact that CAGE sequencing picks up all transcripts which contain capped mRNA, this data is also suitable for further investigation of the role of long non-coding RNAs and miRNAs in monocyte to macrophage differentiation, as has already been done for the FANTOM5 data (Hon *et al*., 2017, de Rie *et al*., 2017).

Chapter 4: Deubiquitinases BAP1, USP12 and USP16 and their functional role in macrophage differentiation and monocytic leukaemia

4.1 Introduction

The previous chapter described the genome-wide changes that occur during THP-1 differentiation. This chapter will focus on particular chromatin-modifying enzymes, H2A deubiquitinases, which remove the PRC1 repressive mark and pave the way to activation of gene expression. As mentioned in Chapter 1 section 1.2.2, there are multiple deubiquitinases, but this study concentrated on three – BAP1, which has been implicated as a tumour suppressor, USP12, which appeared to be regulated in the earlier FANTOM4 study of THP-1 cells, and USP16, with a role in haematopoiesis. These are discussed in more detail below.

As discussed in the Introduction (section 1.2.2.1), *BAP1* (BRCA1-associated protein 1) is a tumour suppressor gene mutated in various cancers, most relevantly to the project in chronic myeloid leukaemia (Dkhissi *et al*., 2015). BAP1, as noted from the name, interacts with BRCA1, a tumour suppressor most famously known for its involvement in human breast cancer. BAP1 has various other roles, ranging from regulating cell cycle progression, to promoting DNA double strand break repair (Yu *et al*., 2014). The *Bap1* knockout allele in mice associates with lethality around embryonic days E9.5/E8.5 (Dey *et al*., 2012). Conditional knockouts in haematopoietic lineages exhibit features of human myelodysplastic syndrome, a group of cancers where bone marrow cells do not mature properly, plus there is an increase in number of monocytes (Dey *et al*., 2012). BAP1 only contains one recognised protein domain, peptidase C12, a ubiquitin carboxyl-terminal hydrolase which is the catalytic domain for removing ubiquitin (Figure 4-1).

Unlike BAP1, USP12 and USP16 are not well studied proteins, with only 30 and 31 published publications cited on PubMed to date (accessed October 2017, https://www.ncbi.nlm.nih.gov/pubmed/).

As mentioned in section 1.2.2.2, USP12 was found to have various roles, including activating androgen receptors (Burska *et al*. 2013), as a negative regulator of NOTCH signalling (Moretti *et al*. 2012), as a cell cycle regulator (Tang *et al*., 2016). Most importantly, it was also found to deubiquinate H2A and H2B *in vivo* and its deficiency was found to have detrimental effects on the developing *Xenopus* embryo (Joo *et al*., 2011). The MGI (Mouse Genome Informatics) database reports that homozygote knockouts of *Usp12* (via insertion of the neo cassette into the gene) are viable but have reduced growth, some immune deficiencies and cardiovascular abnormalities (accessed on 21st October 2017, http://www.informatics.jax.org/marker/phenotypes/MGI:1270128). Until recently, there was no study looking at USP12 in macrophages. Consistent with human macrophages stimulated with LPS in the FANTOM5 results described in Chapter 1, a study by Nayak *et al*. (2017) found *Usp12* to be upregulated in a mouse macrophage cell line in response to LPS stimulation, and its knock down showed that Usp12 might be needed for the activation of the NF-κB pathway. USP12 is a small protein (Figure 4-1), with only one domain identified by bioinformatic tools, the same C19 ubiquitin carboxy-terminal domain found in USP16. USP12 needs several other proteins to perform its deubiquitinating function to the full extent (Joo *et al*., 2011, McClurg *et al*., 2015). Thus, USP12 may be a catalytic protein, which requires other compounds and co-factors to guide it to specific targets.

USP16 has so far been found to have three independent functions (see section 1.2.2.3): cell cycle progression (Xu *et al*., 2013), DNA damage repair (Zhang *et al*., 2014) and gene activation by removal of H2A ubiquitins (Yang *et al*., 2014). *Usp16* knockout mice are embryonically lethal (Yang *et al*., 2014), and in conditional bone marrow knockout mice, the haematopoietic stem cells were reduced in maturity and lineage commitments and there were less mature cells in peripheral blood (Gu *et al*., 2016). No studies in macrophages or leukaemic cells have been reported to date. USP16 protein contains two domains; a zinc finger, UBP type domain (also called BUZ domain), which is also found in histone deacetylase 6 (HDAC6); and a catalytic site in the C19 ubiquitin carboxy-terminal domain (Figure 4-1). The UBP type domain

contains three zinc-binding sites consisting of 12 residues (Pai *et al*., 2007), which facilitate protein-protein and DNA-protein interactions.

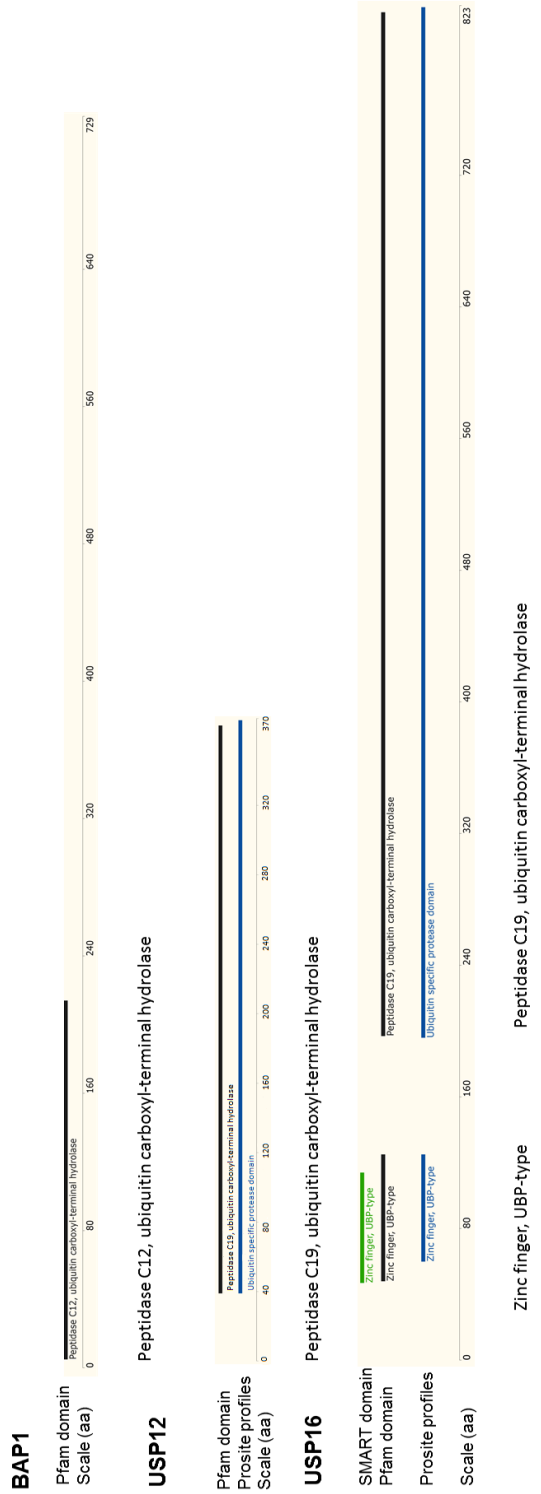

Figure 4-1 Protein domains of three H2A deubiquitinases studied in this chapter. The main transcripts and their translated protein domains as identified by SMART, Pfam and Prosite tools are shown. Information downloaded from Ensembl database. All three proteins have a catalytic ubiquitin hydrolase domain, USP16 also has a Zinc finger domain. Scale shows the length of the proteins in amino acid residues.

4.1.1 Aims of this chapter

This chapter deals with the roles of three H2A deubiquitinases in the human acute myelocytic leukaemia cell line THP-1 and how depletion of these enzymes affects differentiation into macrophages. siRNA for *BAP1, USP12* and *USP16* was used to reduce the levels of mRNA for these genes. Genome-wide analysis of gene expression of the siRNA treated cells was analysed using CAGE sequencing (described in Chapter 3). Overall, this chapter provides insights into the role of these three deubiquitinases in leukaemic cells and differentiation into macrophages.

4.2 Materials and methods

4.2.1 Gene and protein structures and phylogenic analysis of deubiquitinase family

Information about protein and gene structures was taken from the Ensembl database: https://www.ensembl.org (Yates *et al*., 2016). Information was correct as of September 2017. Transcription Start Site (TSS) information was taken from the FANTOM 5 database (Forrest *et al*., 2014) visualised on the Zenbu hg19 genome viewer (http://fantom.gsc.riken.jp/zenbu/gLyphs/) (Severin *et al*., 2014).

4.2.2 RNA interference

siRNA pools (GE Dharmacon) against *BAP1, USP12* and *USP16* were used as described in Chapter 2 section 2.3. All three siRNA pools were used in the initial experiment to test the efficacy of the siRNA, as well as the 6-well experiments set up for CAGE.

4.2.3 CAGE library preparation and bioinformatics analysis

Libraries were prepared and sequenced as specified in Chapter 2 section 2.2.6 together with processing the sequencing reads, and mapping to hg38 genome (section 2.5.1). Results for 38 samples (together with samples for Chapter 6) were uploaded into the CAGEr programme (Haberle *et al*., 2015) (Table 4-1, Appendix 8.3). All samples were normalised with CAGEr using the power law with alpha of 1.19, with results being in Tags per million (TPM). Data was analysed as specified in Chapter 2 section 2.5 using Miru software and the EdgeR package in R.

Sample	Time point	Number of biological replicates
Wildtype (from Chapter 3)	0 hours	6
	24 hours	4
Non-targeting control	0 hours	2
	24 hours	2
BAP1 siRNA	0 hours	2
	24 hours	2
USP12 siRNA	0 hours	2
	24 hours	2
USP16 siRNA	0 hours	2
	24 hours	2

Table 4-1 CAGE samples used in the analysis in this chapter. All samples had two time points, 0 and 24 hours.

4.3 Results

4.3.1 Bioinformatics analysis of H2A deubiquitinases

Human *BAP1, USP12* and *USP16* gene structures were compared based on the Ensembl database to gain more knowledge about these genes, which would be needed for genetic editing at the later stage of the project (Chapter 6).

BAP1 gene has six different protein-coding transcripts, one product predicted to undergo nonsense mediated decay and four products with retained introns not producing any protein according to Ensembl database. The BAP1 protein contains peptidase C12, a ubiquitin carboxyl-terminal hydrolase domain, which catalyses the ubiquitin removal. Only four genes in the human genome have been found to contain this domain; *BAP1*, *UCHL5*, *UCHL1* and *UCHL3*. The principal transcript isoform used for this is ENST00000460680.5, which is 3937 bp long and codes for 729 amino acids. As seen from FANTOM5 data (Figure 4-2), the *BAP1* primary promoter p1@BAP1, which initiates the principal transcript, is expressed ubiquitously throughout various tissues. The second promoter p2@BAP1 is located in the middle of the gene, producing smaller transcripts and expressed mainly in neural tissues, and not at all in monocytes. Thus, the principal isoform ENST00000460680.5 was used for further research in the THP-1 cell line.

Figure 4-2 BAP1 transcription start sites as identified in FANTOM5. Image taken from Zenbu viewer. The primary promoter (p1@BAP1) for BAP1 is identified by red arrow. The second, tissue-specific promoter p2, is identified by black arrow. BAP1 is transcribed from the reverse strand, reading from right to left across the image.

USP12 is a gene with large introns, which has 3 different protein coding transcripts. The principal transcript of *USP12* (ENST00000282344.10) is also the longest at 4511

bp and codes for a peptide only 370 amino acids long. USP12 protein contains a single catalytic domain, which is very common among ubiquitin specific peptidases, called the ubiquitin specific protease domain (according to Prosite database) and peptidase C19, another ubiquitin carboxyl-terminal hydrolase (according to Pfam database). FANTOM5 results show a very simple promoter architecture (Figure 4-3) with a single TSS region.

Figure 4-3 USP12 transcription start sites as identified in FANTOM5. Image taken from Zenbu viewer. The primary promoter (p1@USP12) for USP12 is identified by a red arrowhead. USP12 is transcribed from the reverse strand, reading from right to left across the image.

The *USP16* gene produces for 4 different protein-coding transcripts, and 2 retained intron, non-protein coding transcripts. USP16 protein contains the same deubiquitinase catalytic domain as USP12, but it also contains a zinc finger UBP type domain for DNA binding. This type of zinc finger domain has only been found in 14 human genes at present. Three of the 4 *USP16* protein-coding transcripts code for peptides with very similar length (two for 823 amino acids and one for 822 amino acids). The main difference between these three transcripts are the different lengths of the 5' untranslated regions (UTR), due to variable non-coding exon usage; they all appear to use the same promoter region based on FANTOM5 and Ensembl data. The small protein-coding transcript (ENST00000399973.1) shares the same TSS as other transcripts based on FANTOM5 data (Figure 4-4), but codes for the zinc finger domain only. Thus, the longest transcript (ENST00000399973.1) of 3004 bp coding for 823 amino acids was further used in the THP-1 study.

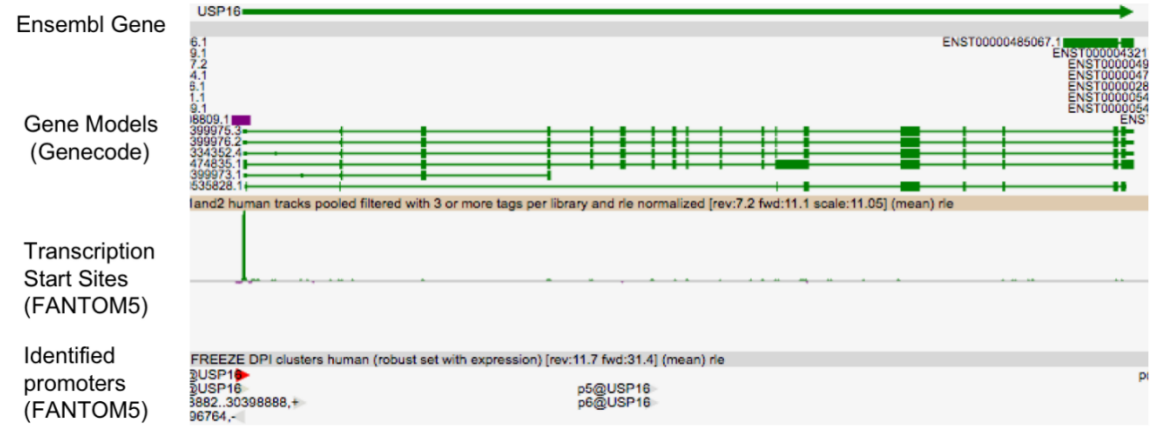

Figure 4-4 USP16 transcription start sites as identified in FANTOM5. Image taken from Zenbu viewer. Primary promoter (p1@USP16) for USP16 is identified by a red arrowhead. USP16 is transcribed from the forward strand, reading from left to right across the image.

4.3.2 Using siRNA technology to knock down the expression of *BAP1*, *USP12* and *USP16* in THP-1 cells

4.3.2.1 Preliminary experiments assessing the siRNA efficiency shows sufficient knock down for all three genes

siRNA was used to knock down the deubiquitinase transcripts and to determine what happens to these THP-1 cells during macrophage differentiation.

A pool of 4 different siRNAs was ordered (Chapter 2 section 2.3) for each gene (*BAP1, USP12,* and *USP16*) and their efficiency was tested using an 8-day time course (Figure 4-5). Lipofection of the THP-1 cell line with siRNA was carried out on day -1, and cells were stimulated by PMA to differentiate into macrophages on day 0 (section 2.3). RNA was taken on day -1, 0, 1, 2, 3 and 7.

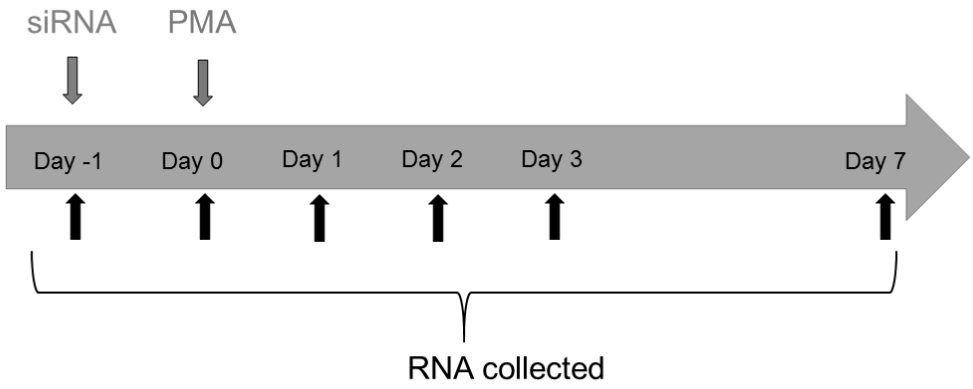

Figure 4-5 Diagram of the preliminary siRNA experiment assessing siRNA efficiency. siRNA was added at day -1, differentiation of THP-1 was started by PMA at day 0. RNA was collected at day -1, day 0, day 1, day 2, day 3 and day 7.

The *BAP1* siRNA pool of 4 oligonucleotides knocked down the expression of *BAP1* to the lowest levels among all samples (Figure 4-6). Some level of BAP1 knock down lasted till at least day 3, compared to all other controls in this experiment – Non-targeting siRNA pool control, Lipofectamine only control and wildtype THP-1 differentiation control.

The Non-targeting control (NT siRNA) and Lipofectamine only control (Lipofectamine) showed almost identical levels of *BAP1* expression. The wildtype THP-1 (Control) closely followed their expression levels until day 7, where its *BAP1* expression was lower, and similar to day -1 and 0 (before differentiation). Unlike the Control, the NT siRNA and Lipofectamine controls had the highest levels of *BAP1* at day 7. *BAP1* siRNA knock down did not affect the THP-1 macrophage morphology compared to the condition without treatment. *BAP1* siRNA pool was deemed suitable to proceed for CAGE sequencing.

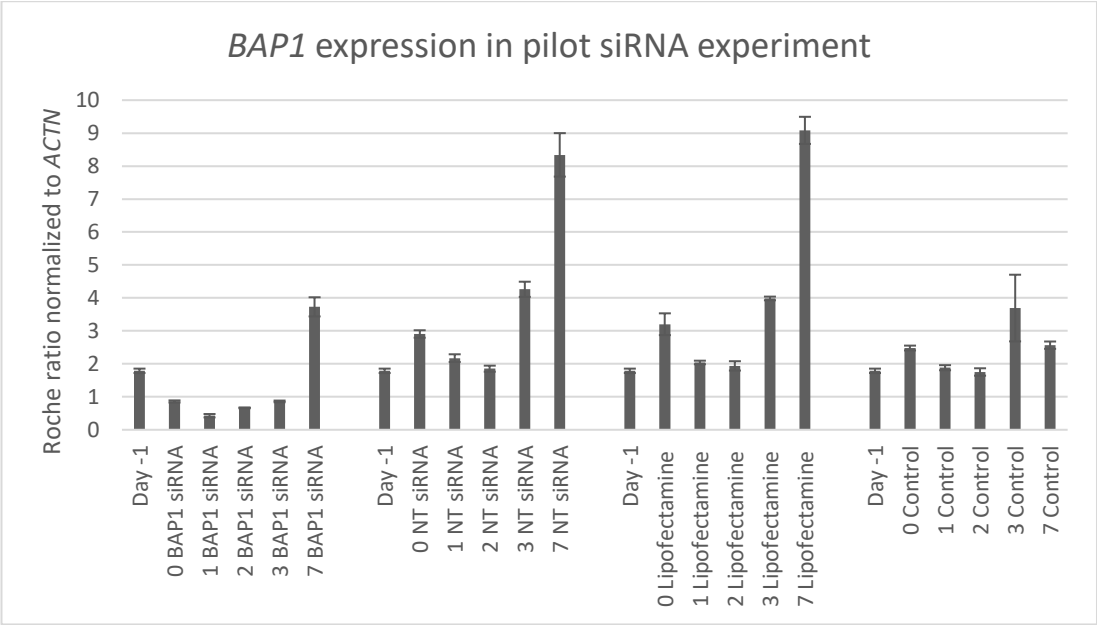

Figure 4-6 qRT-PCR results of an initial siRNA experiment. Pictured here is the BAP1 expression normalised to ACTN expression and error bars showing standard error (calculated by the by Roche LightCycler 480 software). Day -1 is the same for all treatments. 0 means day 0, 1 is day 1 and etc. 0 BAP1 siRNA stands for day 0 sample with BAP1 siRNA, 1 NT siRNA stands for day 1 with Non-targeting control siRNA, 2 Lipofectamine stands for day 2 cells with Lipofectamine only, and 3 Control stands for day 3 wildtype THP-1 differentiation.

A *USP12* siRNA pilot experiment was carried out in the same way as *BAP1* pool, except there was no Lipofectamine control. *USP12* levels were knocked down by siRNA to less than half of controls (Non-targeting siRNA pool and wildtype THP-1 differentiation) at day 0 and day 1 (Figure 4-7). The *USP12* siRNA pool seemed to considerably downregulate the *USP12* transcript levels longer than for *BAP1*, until day 3 or even day 7. The Non-targeting siRNA control pool caused the USP12 levels to rise to higher levels than in the wildtype THP-1 differentiation control (probably due to the lipofection process itself (Fiszer-Kierzkowska *et al*., 2011)). *USP12* siRNA knock down did not affect the THP-1 macrophage morphology compared to without treatment. Knock down of *USP12* by *USP12* siRNA pool was also deemed sufficient for subsequent CAGE sequencing.

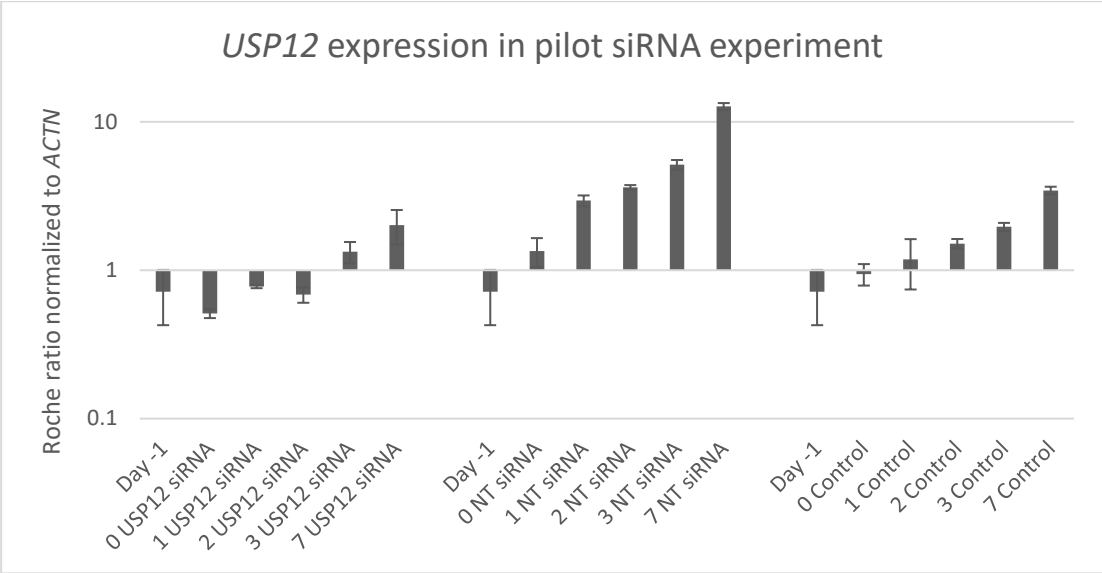

Figure 4-7 qRT-PCR results of an initial siRNA experiment. Pictured here is the USP12 expression normalised to ACTN expression and error bars showing standard error (calculated by the by Roche LightCycler 480 software). Day -1 is the same for all treatments.0 means day 0, 1 is day 1 and etc. 0 USP12 siRNA stands for day 0 sample with USP12 siRNA, 1 NT siRNA stands for day 1 Non-targeting control siRNA, and 2 Control stands for day 2 wildtype THP-1 differentiation. Y axis is in logarithmic scale (base 10).

USP16 siRNA pool efficiency was also tested the same way as the *BAP1* siRNA pool (Figure 4-8). *USP16* expression was knocked down by siRNA to less than half of the controls in day 0, and even more in day 1. The mRNA levels of USP16 were considerably lower in the siRNA treated samples than the controls for the duration of the experiment. The experiment controls (Non-targeting siRNA pool, Lipofectamine only and wildtype THP-1 differentiation) all exhibited similar levels of *USP16* mRNA for the duration of the time course. *USP16* siRNA knock down did not affect the THP-1 macrophage morphology compared to without treatment. Knock down of *USP16* by siRNA was also deemed sufficient for subsequent CAGE sequencing.

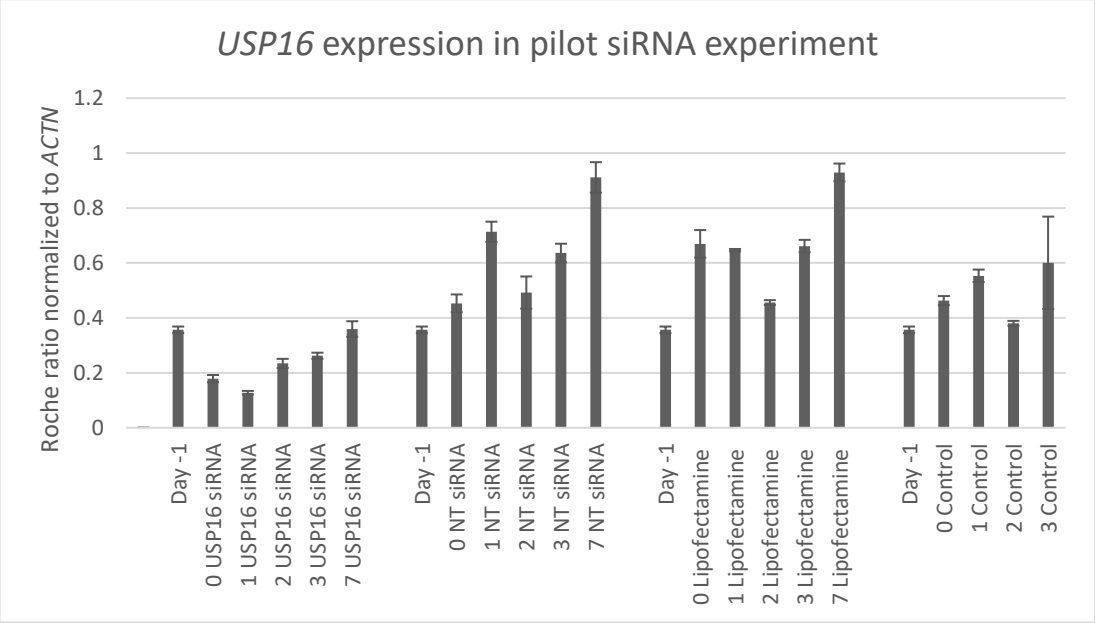

Figure 4-8 qRT-PCR results of an initial siRNA experiment. Pictured here is the USP16 expression normalised to ACTN expression and error bars (calculated by the by Roche LightCycler 480 software). Error bars show standard error. Day -1 is the same for all treatments.0 means day 0, 1 is day 1 and etc. 0 UTY siRNA stands for day 0 sample with UTY siRNA, 1 NT siRNA stands for day 1 Non-targeting control siRNA, 2 Lipofectamine stands for day 2 cells with Lipofectamine only, and 3 Control stands for day 3 wildtype THP-1 differentiation.

4.3.2.2 siRNA experiments for CAGE sequencing

The purpose of these experiments was to harvest enough good quality RNA from THP-1 cells treated with siRNA pools to perform CAGE sequencing. It was decided that only day 0 (24 hours after adding siRNA pools) and day 1 (48 hours after adding siRNA pools and 24 hours after adding PMA and starting macrophage differentiation) would be used in CAGE sequencing studies. All experiments were carried out four times to select the best 2 biological replicates with good quality and quantity of RNA, and also highest knock down levels by siRNA pools.

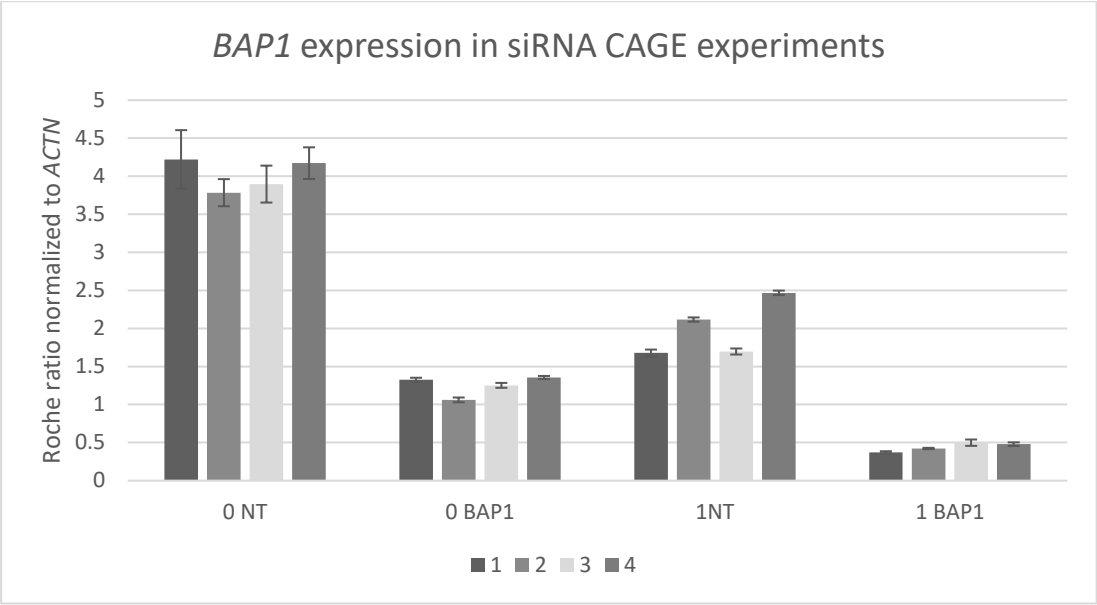

Figure 4-9. qRT-PCR results of BAP1 expression in four siRNA experiments for CAGE sequencing. Pictured here is the BAP1 expression normalised to ACTN expression and error bars showing standard error (calculated by the by Roche LightCycler 480 software). 0 NT is Non-targeting siRNA pool at day 0, 1 BAP1 is BAP1 siRNA pool at day 1.

	% of *BAP1* expression after knock down	
Experiment	day 0	day 1
1	31%	22%
2	28%	20%
3	32%	29%
4	32%	19%

Table 4-2 BAP1 expression after knock down by siRNA pools compared to NT control levels (% of control level). Calculated with qRT-PCR levels normalized to ACTN by Roche LightCycler 480.

As seen in Figure 4-9, *BAP1* downregulation by siRNA compared to the Non-targeting control (NT) was consistent between all four replicates. Calculating efficiency of the knock down (Table 4-2) shows that most replicates had BAP1 levels downregulated to about 30% of the control at day 0 and around 20% for day 1. Any of these replicates would be suitable for further analysis, but due to the varying quantity of RNA (at least 5 µg is needed for CAGE seq), experiments 1 and 3 were selected for CAGE. In Experiment 4, *BAP1* day 0 total RNA was less than 5 µg, and the same for Experiment 2 at day 1.

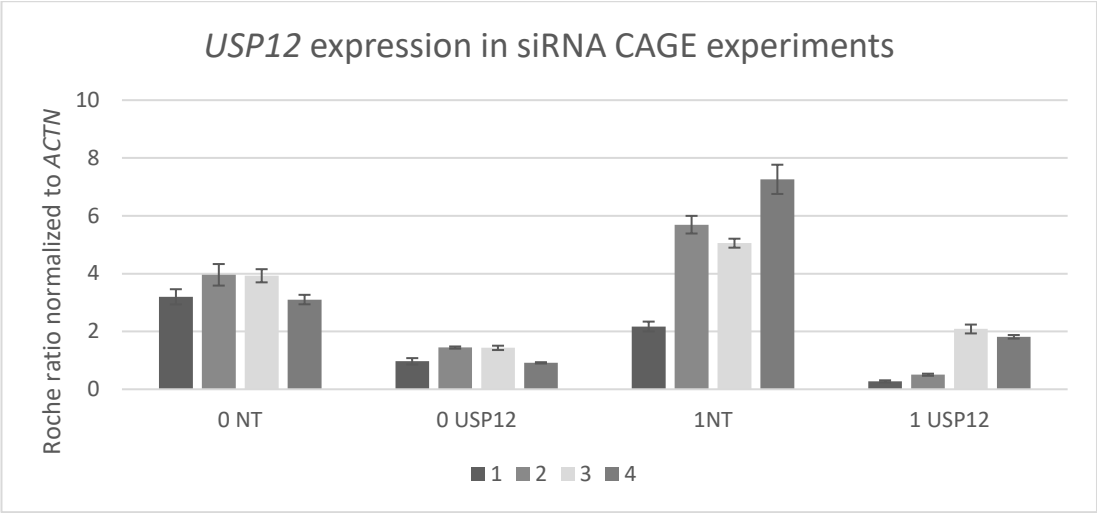

Figure 4-10 qRT-PCR results of USP12 expression in four siRNA experiments for CAGE sequencing. Pictured here is the USP12 expression normalised to ACTN expression and error bars showing standard error (calculated by the by Roche LightCycler 480 software). 0 NT is Non-targeting siRNA pool at day 0, 1 USP12 is USP12 siRNA pool at day 1.

	% of *USP12* expression after knock down	
Experiment	day 0	day 1
1	30%	13%
2	36%	9%
3	36%	41%
4	29%	25%

Table 4-3 USP12 expression after knock down by siRNA pools compared to NT control levels (% of control level). Calculated with qRT-PCR levels normalized to ACTN by Roche LightCycler 480.

USP12 downregulation by siRNA pools at day 1 was not as consistent in replicates as for *BAP1* (Figure 4-10). *USP12* gene expression levels for the Non-targeting control at day 1 (1 NT) varied across the replicates as well. Nevertheless, *USP12* siRNA pools managed to knock down the expression of *USP12* to around 30% of the control for day 0 and from 9% to 41% for day 1 (Table 4-3). Due to poor quantity and quality of RNA for replicates 1 and 4, the replicates chosen for RNA sequencing were from experiments 2 and 3.

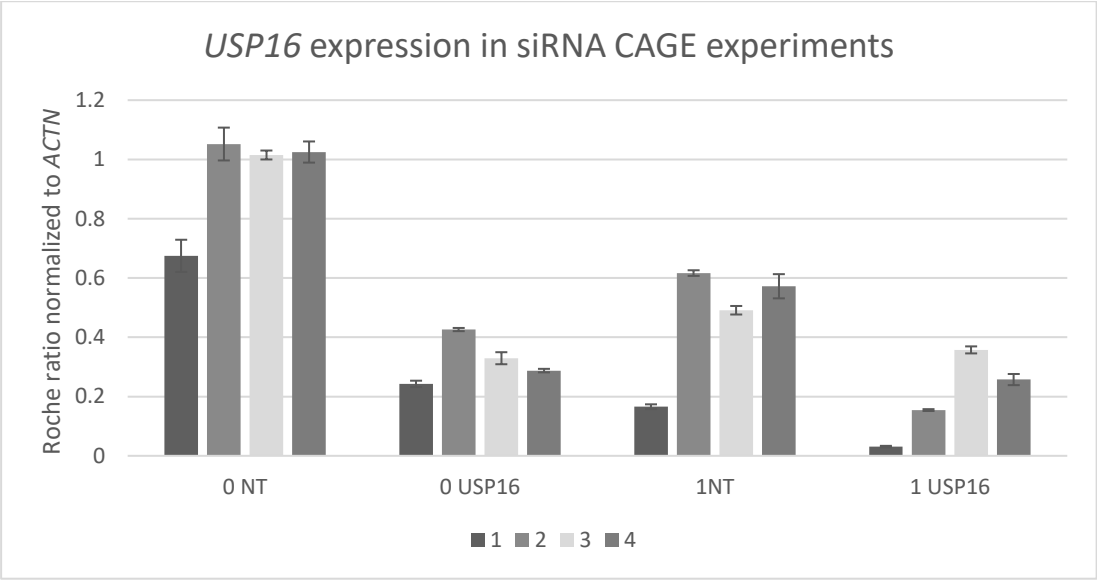

Figure 4-11 qRT-PCR results of USP16 expression in four siRNA experiments for CAGE sequencing. Pictured here is the USP16 expression normalised to ACTN expression and error bars showing standard error (calculated by the by Roche LightCycler 480 software). 0 NT is Non-targeting siRNA pool at day 0, 1 USP16 is USP16 siRNA pool at day 1.

	% of *USP16* expression after knock down	
Experiment	day 0	day 1
1	36%	19%
2	40%	25%
3	32%	73%
4	28%	45%

Table 4-4 USP16 expression after knock down by siRNA pools compared to NT control levels (% of control level). Calculated with qRT-PCR levels normalized to ACTN by Roche LightCycler 480.

USP16 siRNA experiment replicates also showed inconsistency in the USP16 levels (Figure 4-11). However, most samples showed reduction of mRNA levels of USP16 to about 30% of the control for day 0 (Table 4-4). Day 1 expression levels were highly varied, from 19% of Non-targeting control in experiment 1 to 73% in experiment 3. For further CAGE sequencing, replicates from experiments 1 and 2 were selected due to the best quality and quantity of RNA and also highest levels of downregulation in day 1 samples.

4.3.3 CAGE sequencing

As discussed in Chapter 3, all CAGE reads were mapped to hg38 and normalised between each other using power law distribution. The 26 CAGE samples included in the analysis for this chapter were all siRNA samples and controls, and wildtype controls from THP-1 differentiation from Chapter 3 for 0 hours and 24 hours.

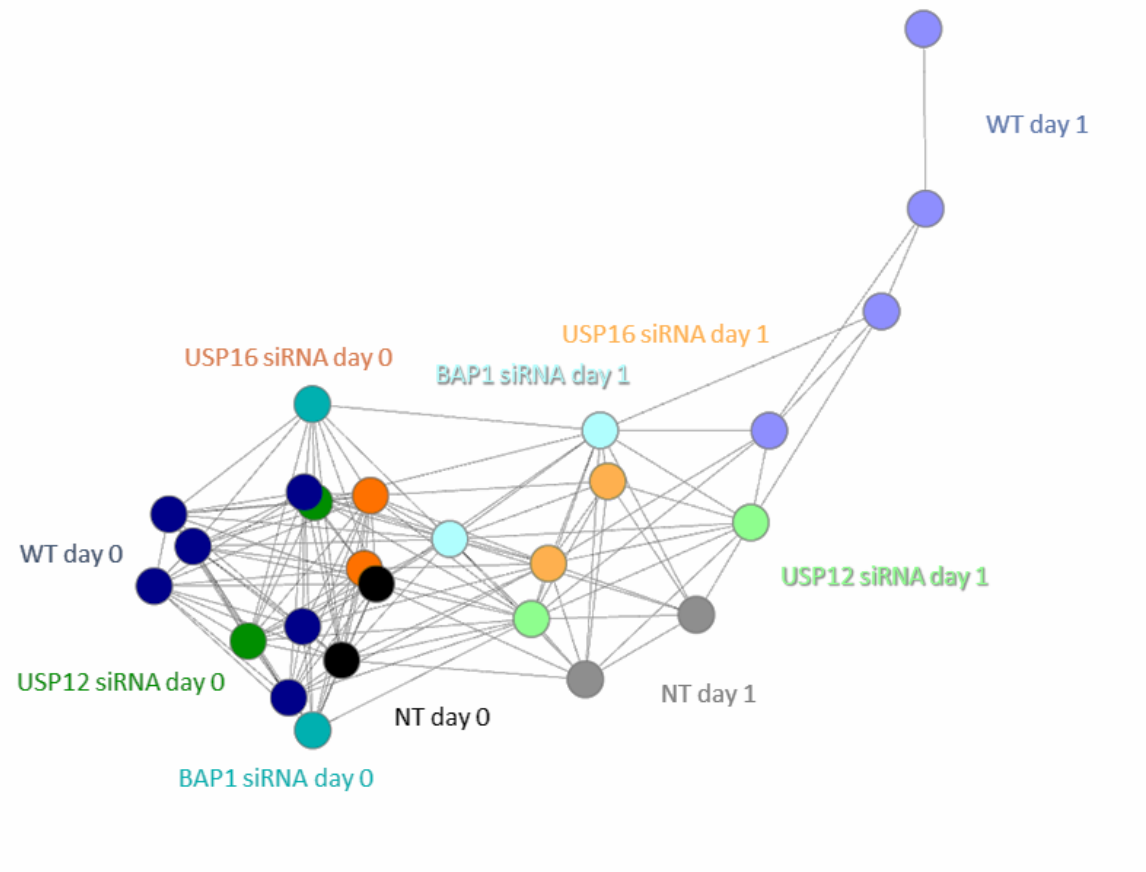

Figure 4-12 Sample-sample correlation of correlation coefficient of 0.88 derived from Miru programme. Samples are nodes and edges are connections between them with a correlation coefficient of 0.88 or greater, i.e. how closely correlated the expression patterns are to each other across all samples. Nodes which are close are highly correlated. Each sample type is different colour: wildtype – blue, Non-targeting siRNA control – black, BAP1 siRNA – aquamarine, USP12 siRNA – green and USP16 siRNA - orange. Day 0 is bolder colour, Day 1 paler.

All samples were then correlated to each other with Pearson's correlation coefficient of 0.88 (Figure 4-12) using Miru software. From the Figure 4-12, it is clearly seen that the most important divide between the samples is the differentiation status. All undifferentiated samples (Day 0; bolder colours) are on the left of the figure and their expression correlates between each other more than the differentiated samples (Day 1; paler colours), which are primarily on the right. The greatest difference between Day

0 and Day 1 was in the wildtype, indicating that the treated samples all had some response to the treatment (including the Non-targeting control). All the treated samples showed a similar shift in expression to the NT control, suggesting that much of the observed difference may be due to non-specific effects following transfection of RNA using lipofectamine.

The question that was considered was whether the siRNA knock down of *USP12, USP16* and *BAP1* also had an effect on THP-1 gene expression. The sample-to-sample correlation in Figure 4-12 shows that the biological replicates of many time points were not strongly correlated. For example, for *BAP1* at day 0 or *USP12* at day 1 (though these had a different levels of knock down), replicates were quite distant in the network layout. It was noted that the two *USP12* Day 1 replicates had very different reductions in mRNA levels measured by qRT-PCR (9% compared with 41% of NT control; Table 4-3), while the two *BAP1* Day 0 replicates chosen had very similar levels (Table 4-2), so the distance between replicates is not entirely due to differential knock down.

The gene expression levels of the three deubiquitinases detected by CAGE in each of the knock downs were assessed using Zenbu genome viewer. BAP1 levels at day 0 were knocked down on average to 45% of the Non-targeting control, and 46% for day 1 in the *BAP1* knock down cells (Figure 4-13). In the *USP12* knock down, BAP1 levels were higher than in the NT control and in the *USP16* knock down, BAP1 levels were similar to the NT controls. This indicates that BAP1 may compensate for *USP12* knock down.

USP12 expression was knocked down on average to 45% of the Non-targeting control (NT) at day 0 (Figure 4-14). At day 1, Non-targeting control and *USP12* siRNA values for *USP12* expression were more variable in the biological replicates. Nevertheless, on average, the expression was lowered to 47% of the Non-targeting control. In the *BAP1* and *USP16* knock downs, USP12 levels were similar to Non-targeting control at day 0. USP12 levels were marginally lower than Non-targeting control at day 1 of

the *USP16* knock down, but the levels in the *BAP1* knock down were variable with one replicate much higher and one lower than the Non-targeting control.

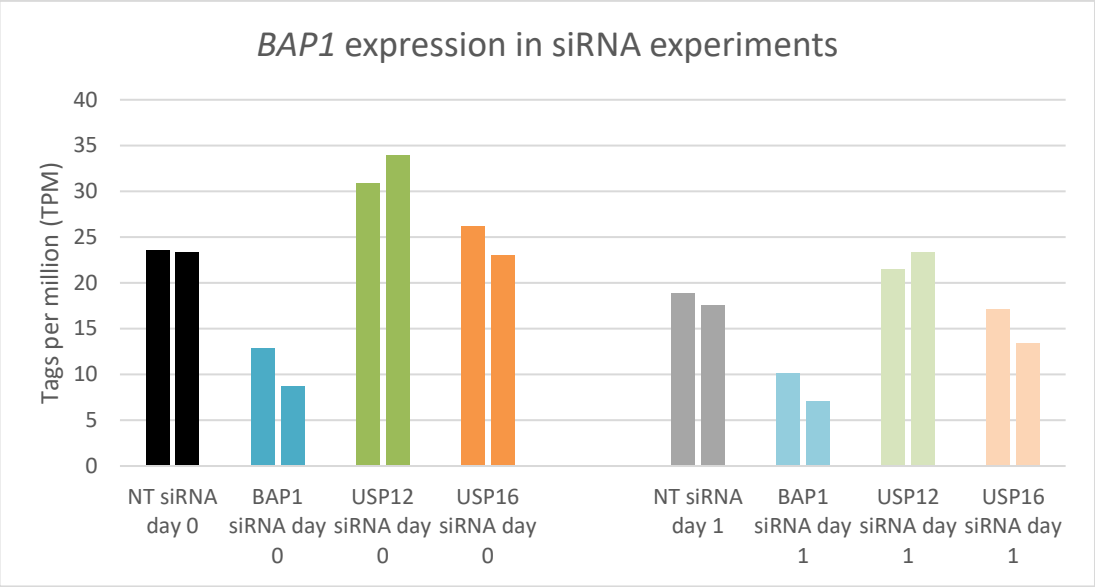

Figure 4-13 BAP1 expression in siRNA experiments assessed by CAGE sequencing. The gene expression levels are normalised and in Tags per million (TPM). Samples are Non-targeting control (NT siRNA), and treatment with BAP1 siRNA. USP12 siRNA and USP16 siRNA. Two biological replicates are shown.

Unlike the initial qRT-PCR experiment (Figure 4-11). *USP16* expression detected by CAGE did not seem to decrease after treatment with *USP16* siRNA either for day 0 or day 1 in the samples sent for CAGE sequencing (Figure 4-15).

CSF1R and *MYB* expression levels were similar to those of the Non-targeting control at both day 0 and day 1, so there is no indication that THP-1 cells subjected to siRNA treatment to reduce USP12 and BAP1 deubiquitinases were different from untreated cells, in their differentiation state.

USP12 downregulation was found to increase expression of *USP46* (Joo *et al*., 2011), but the level of USP46 was not found to be affected by any siRNA samples in both replicates (Figure 4-16).

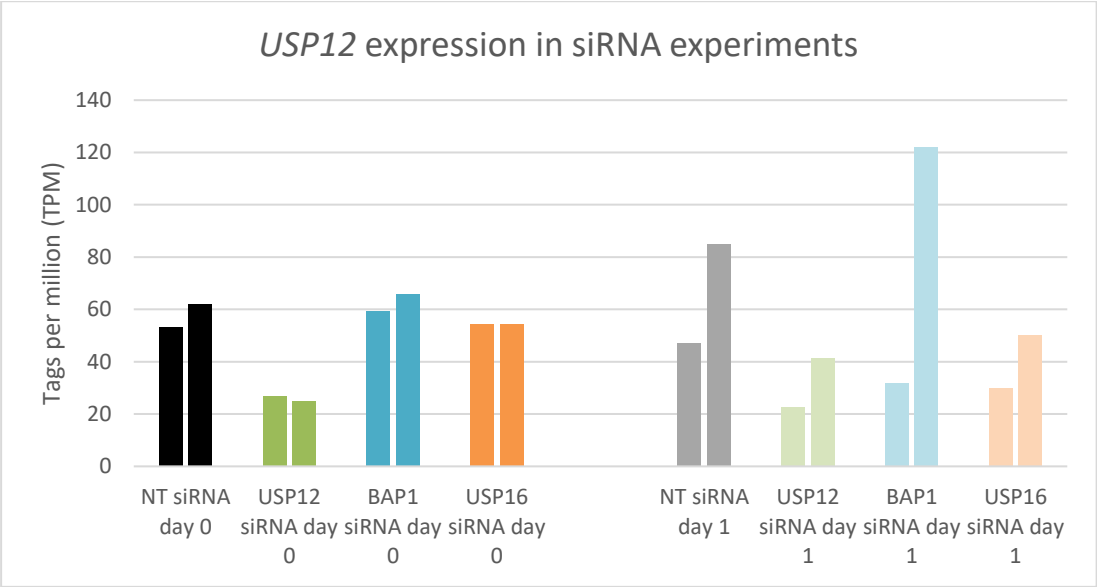

Figure 4-14 USP12 expression in siRNA experiments assessed by CAGE sequencing. The gene expression levels are normalised and in Tags per million (TPM). Samples are Non-targeting control (NT siRNA), and treatment with BAP1 siRNA. USP12 siRNA and USP16 siRNA. Two biological repeats are shown.

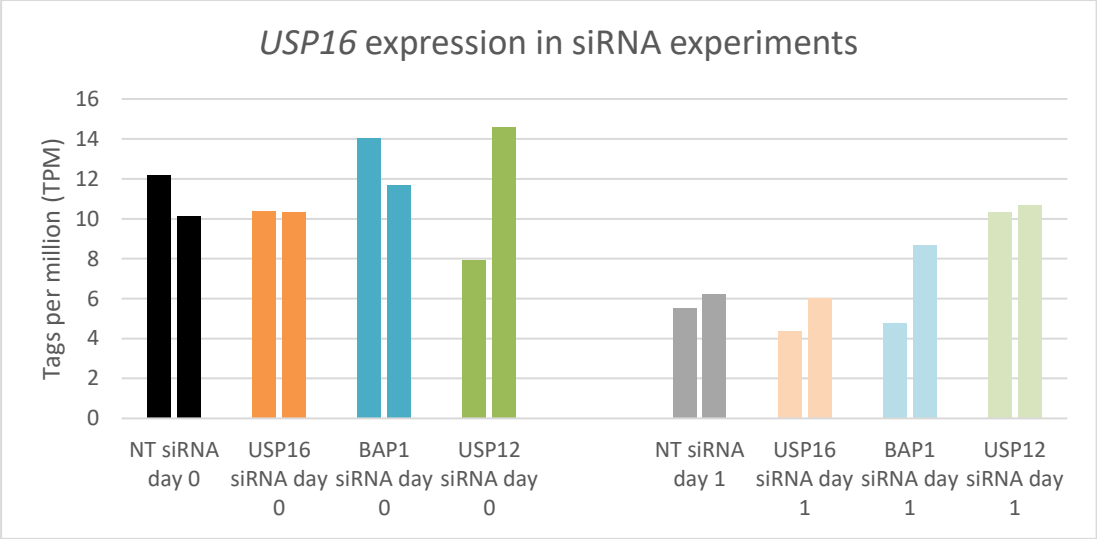

Figure 4-15 USP16 expression in siRNA experiments assessed by CAGE sequencing. The gene expression levels are normalised and in Tags per million (TPM). Samples are Non-targeting control (NT siRNA), and treatment with BAP1 siRNA. USP12 siRNA and USP16 siRNA. Two biological repeats are shown.

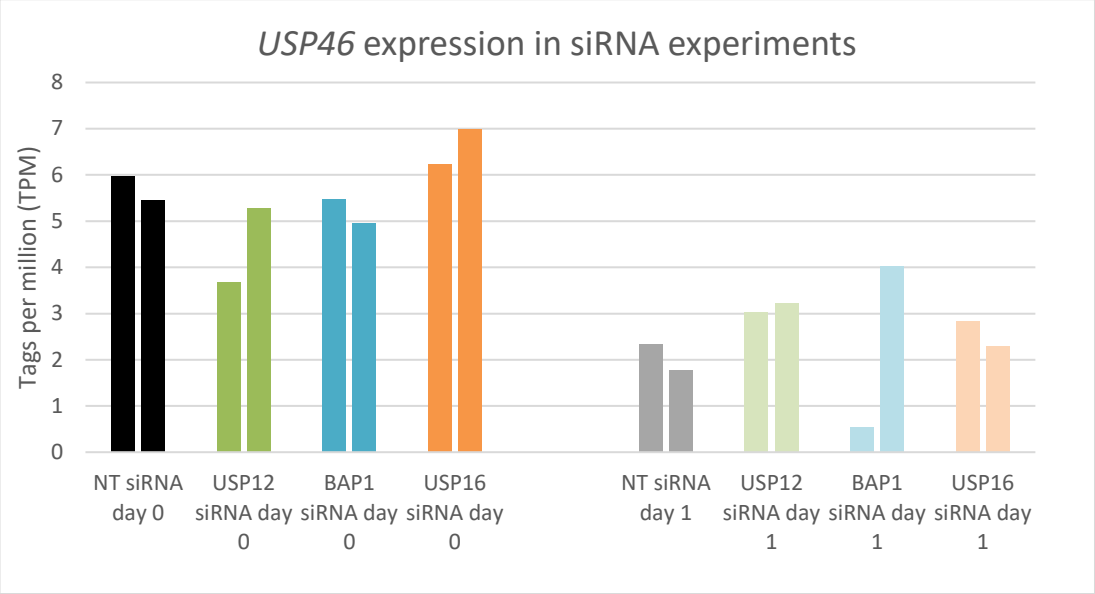

Figure 4-16 USP46 expression in siRNA experiments assessed by CAGE sequencing. The gene expression levels are normalised and in Tags per million (TPM). Samples are Non-targeting control (NT siRNA), and treatment with BAP1 siRNA. USP12 siRNA and USP16 siRNA. Two biological repeats are shown.

All siRNA samples were also analysed using Miru with Pearson's correlation coefficient of 0.95 (Figure 4-17). The two biggest clusters were similar to those seen for the wildtype samples (Chapter 3) and *USP16* knockout samples (Chapter 6) – a cluster of monocyte markers with high expression at the beginning of the time course and one of macrophage markers with high expression after treatment with PMA, which showed the same expression pattern across all samples.

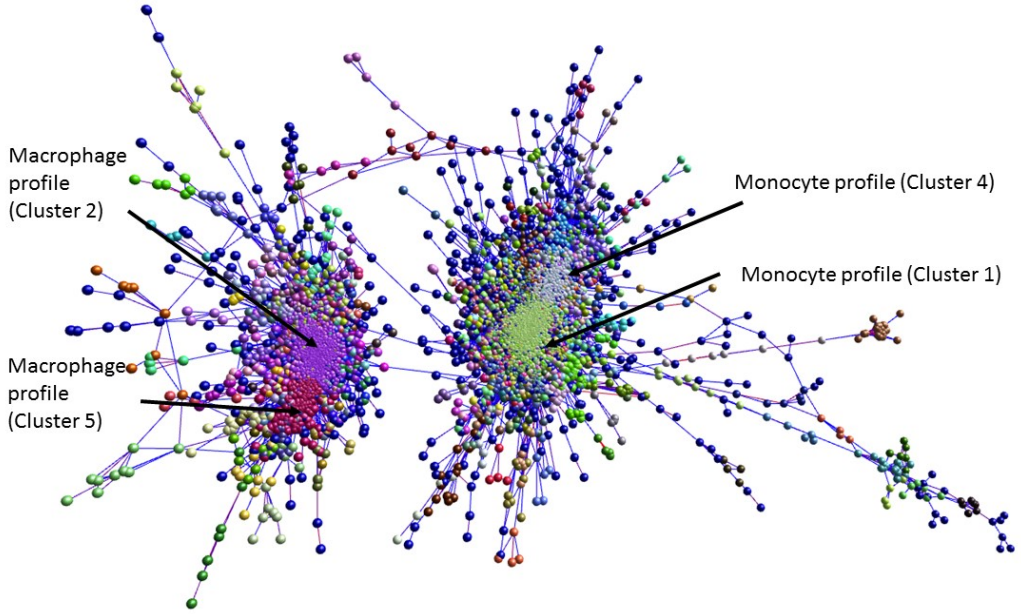

Figure 4-17 Miru layout with CTSS-CTSS Pearson's correlation ≥ 0.95. The largest clusters are annotated.

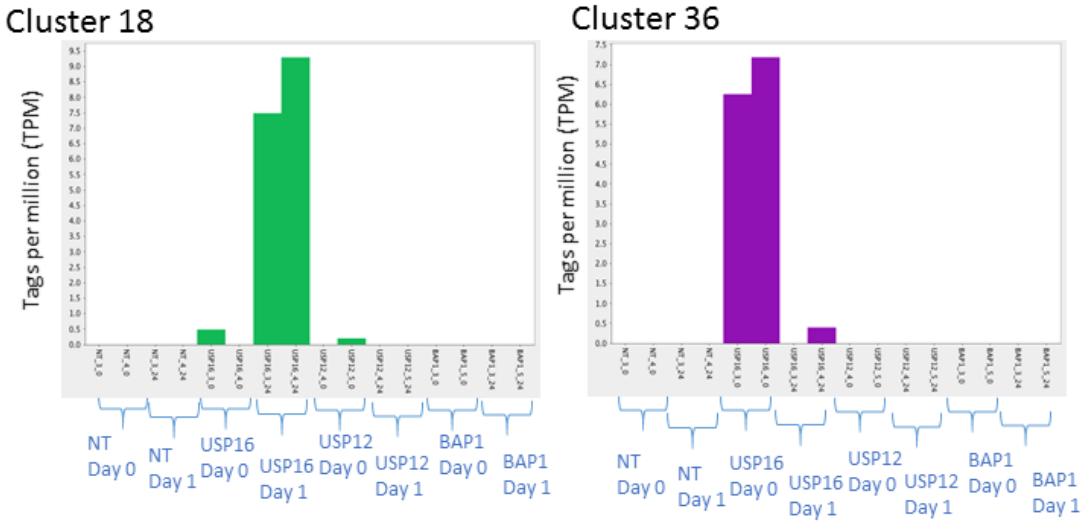

Figure 4-18 Interesting clusters from Miru software comparing all siRNA samples with Pearson's correlation ≥ 0.95. Cluster 18 contains CTSS which were only upregulated in macrophages treated with USP16 siRNA and cluster 36 which are upregulated in USP16 siRNA treated monocytes before differentiation.

The two interesting clusters found were clusters 18 and 36, which, were only expressed in *USP16* knock down samples - one in monocytes and one in macrophages (Figure 4-18). Cluster 18 (high after PMA treatment) was significantly enriched in GO terms relating to alternative splicing, whereas cluster 36 (high before treatment) was

enriched in GO terms relating to ATP-binding. Since *USP16* knock down did not reduce USP16 levels (measured by CAGE), these results possibly indicate an off-target effect of the siRNA pool.

No *USP12* or *BAP1* specific clusters were found, possibly because the biological replicates did not correlate well, or due to the different levels of knock down. Most clusters contained genes that were highly expressed in one of the replicates and not the other. While these clusters might have revealed the reasons for the lack of replication, the data could not be used to study the effects of the knock down and were not analysed further.

Differential expression in siRNA-treated samples was also assessed using the EdgeR package (Table 4-5). There were only 30 CTSS that were significantly (p-value < 0.01) dysregulated between Non-targeting control and wildtype in non-differentiated state (day 0). There were more CTSS on day 1 (507 CTSS) which were up or down regulated than before differentiation, and the sample-sample correlation graph (Figure 4-12) shows that the Non-targeting control was more similar to other siRNA samples than to wildtype.

Possibly due to the lack of replication in the two samples (as seen by Miru clustering), there was a very small number of CTSS changed in *BAP1*, *USP12* or *USP16* siRNA-treated samples. Most CTSS which were differentially expressed in *BAP1* or *USP12* siRNA samples were not located in gene promoters or any other known regulatory elements.

	NT 0 - WT 0	NT 1 - WT 1	BAP1 0 - NT 0	BAP1 1 - NT 1
-1	2	181	1	0
1	28	326	2	19
	USP12 0 - USP12 1	**USP12 1 - NT 1**	**USP16 0 - NT 0**	**USP16 1 - NT 1**
-1	2	48	6	81
1	0	32	35	9

Table 4-5 Number of CTSS which were upregulated (1) or downregulated (-1) in comparison between Non-targeting control (NT)or wildtype (WT) or BAP1 siRNA (BAP1) or USP12 siRNA (USP12) or USP16 siRNA (USP16) on day 0 (0) or day 1 (1) with smaller p-value than 0.01 as calculated by EdgeR package.

To make sure no important clusters were missing from the analysis, all 26 samples from this chapter were clustered using Miru with Pearson's correlation threshold of greater than or equal to 0.93. Aside from the obvious clusters of monocytic or macrophages markers present in all relevant samples, the most interesting cluster was Cluster 32, which contained 24 CTSS that were lower in all Lipofectamine-treated (with siRNA) differentiated macrophages at day 1 (Figure 4-19) than in wildtype or genome edited cells at day 1. Significant GO annotation enrichment for this cluster included protein transport and splice/sequence variant.

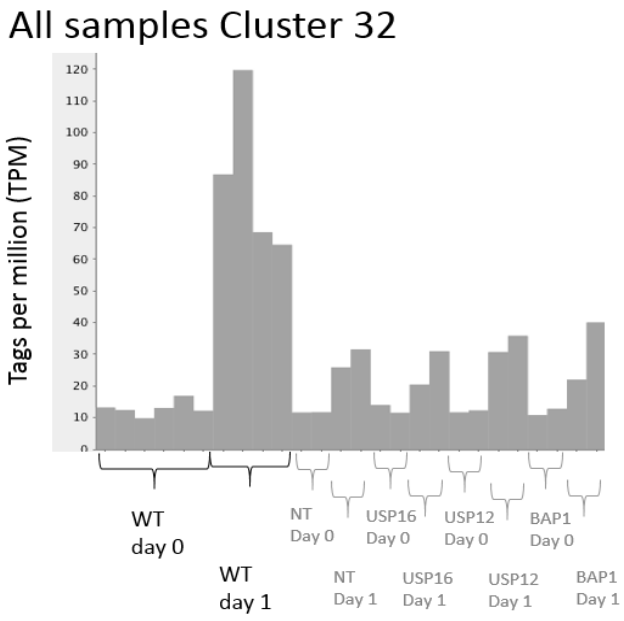

Figure 4-19 Cluster 32 from comparing all Chapter 4 samples in Miru software with Pearson's correlation ≥ 0.93.

4.4 Discussion

4.4.1 Chapter overview

Histone H2A deubiquitinases remove the repressive ubiquitin mark on Lysine 119 and thus pave the way for the activation of gene expression during differentiation or in response to stimuli. In this chapter, depletion of three H2A deubiquitinases, BAP1, USP12 and USP16 in acute monocytic leukaemia THP-1 cells was examined using siRNA technology. Knock down of these three genes in THP-1 cells was analysed using CAGE technology. CAGE libraries were successfully created. Unfortunately, biological replicates for *USP12* and *BAP1* siRNA samples did not correlate with each other and thus not many genes were significantly affected by BAP1 or USP12 depletion. The discrepancy between the qRT-PCR results and CAGE quantification of mRNA levels for *USP16* made it difficult to interpret the impact of *USP16* knock down.

4.4.2 Using siRNA to knock down deubiquitinases in THP-1 cells did not yield any significant results

siRNA was used to knock down the level of these chromatin modifying enzymes. Lipofectamine2000 was chosen as a mode of delivery into the cells due to its previous use in FANTOM4 (Suzuki *et al*., 2009), although other lipofection reagents have been tried in THP-1 with very similar delivery efficiency (not shown). ON-TARGETplus SMARTpool siRNA was chosen, as it contains 4 different siRNAs to maximise the knock down and had been tried before in monocytic cells (Troegeler *et al*., 2014).

BAP1 siRNA and *USP12* siRNA pools successfully downregulated the mRNA levels of *BAP1* and *USP12* respectively, according to both qRT-PCR and also CAGE sequencing. However, even though *USP16* siRNA downregulated *USP16* according to qRT-PCR (with primers in exons 15 and 16), this was not shown by CAGE sequencing. This discrepancy might be due to CAGE only picking up the 5' end of mRNA, not anywhere else. The 4 siRNAs in the pool targeting *USP16* bind to exons 8+9, 15 and 17, but the siRNAs for *BAP1* and *USP12* target exons much earlier – exons 4 and 3, respectively (section 2.3.1). However, due to the mechanism of siRNA

silencing, the 5'end of mRNA would be immediately targeted for degradation and the remnants of functional mRNA would have a very transient lifespan. The way to solve this question is to either perform qRT-PCR with primer sets mapping to other exons (and see if the *USP16* expression truly goes down) or perform western blotting on the samples.

Even though both *BAP1* and *USP12* siRNA biological replicates had similar levels of downregulated transcripts, they did not correlate closely with each other and all replicates had variable global gene expression (as seen in Figure 4-12). This is in contrast with the Non-targeting control, where replicates did group together well. Miru did not find any relevant clusters which were either upregulated or downregulated in both biological replicates for *BAP1* or *USP12*; most of genes were differentially expressed in one replicate, but not the other. This can also be seen in the very small number of differentially expressed genes found by EdgeR. It could be due to the nature of siRNA knock down, when different amounts of protein might be made. Knock downs should therefore be validated by western blotting as well as measurement of mRNA. Without protein validation, it can be only speculated why the two replicates were so variable. Both targets are chromatin modifying enzymes and thus their deficiency might result in different responses in cells – especially in cells which were not synchronised in cell cycle phases. These replicates were done on different days, and even though every precaution was taken to preserve the same experimental variables, cells might have behaved differently. Another explanation is that because *BAP1* expression is upregulated in *USP12* siRNA samples, BAP1 or other deubiquitinase might have compensated differently for this loss in different experiments.

4.4.2.1 Limitations of siRNA use

Using siRNA technology to reduce mRNA translation levels and therefore study the protein function in cells had been widely reported, as it is simple to do. siRNAs are smaller than plasmid DNA and therefore are easier to deliver to cells. However, knocking down gene expression using siRNA also has its disadvantages. First, the reduction in mRNA level is transient and therefore does not permanently affect protein

levels, especially in proliferating cells. Lipofectamine was found to stimulate proliferation in immune cells (Przybylski *et al*., 2017), which is then stopped by addition of PMA in THP-1 cells. However, there have also been reports of Lipofectamine slowing down cell proliferation in HeLa cells (Zhang *et al*., 2011). There are genes which are affected by the Lipofectamine treatment itself (like genes in Cluster 32, Figure 4-19) and thus proper controls (like the Non-targeting control sample) are needed.

The use of Lipofectamine might interfere with various macrophage responses, for example response to LPS (Leon-Ponte *et al*., 2005). Monocytic cells also do not respond well to foreign nucleic acid, since one function is to detect viral infections which can be indicated by the presence of viral DNA or RNA. Nucleic acids can induce cell death (Paijo *et al*., 2016). Optimisation of siRNA delivery methods into THP-1 which do not interfere with macrophage function might be required (as seen for lipofection of primary monocytic and macrophage cells (Troegeler *et al*., 2014)).

4.4.3 Conclusions and future directions

In the *USP12* siRNA samples, when the levels of one deubiquitinase was lower, gene expression of other deubiquitinases was upregulated. This has been documented previously by literature – for example *USP12* downregulation results in *USP46* upregulation (Joo *et al*., 2011). However, *USP46* gene expression was not significantly affected by downregulation deubiquitinases in this study. As documented in the introduction, there are various deubiquitinases that might share redundant function. Future knockdowns or knockouts of multiple deubiquitinases might answer this question. The use of CRISPR-Cas9 technology to knockout the deubiquitinase genes is explored in Chapter 6.

To properly distinguish between different roles of H2A deubiquitinases and concentrate on the regulation of gene expression, ChIP-seq against these H2A deubiquitinases could be performed during wildtype THP-1 differentiation to analyse which gene loci are the ones most likely to be affected by their dysregulation.

Chapter 5: Demethylases KDM6A, KDM6B and UTY and their functional role in macrophage differentiation and monocytic leukaemia

5.1 Introduction

The previous chapter described a study of several deubiquitinase enzymes during the transition of THP-1 monocytes to macrophages. As mentioned in Chapter 1, demethylase enzymes are also important in chromatin modification during cell state changes and the present chapter examines the impact of several of these enzymes on THP-1 differentiation. There are at least two demethylases containing a Jumonji C (JmjC) catalytic domain (described further in section 1.2.3.1 and 1.2.3.2), which are known to be responsible for removing the di-and trimethyl groups on H3K27 (Hong *et al*., 2007). These are KDM6A also known as Ubiquitously Transcribed Tetratricopeptide Repeat on chromosome X (UTX) and KDM6B which was first described as the Jumonji domain containing protein 3 (JMJD3) (Van der Meulen *et al*., 2014, Hong *et al*., 2007). Ubiquitously Transcribed Tetratricopeptide Repeat Containing, Y-Linked (UTY) is a third protein which contains JmjC domain (Figure 5-1) but is frequently overlooked as a non-functional Y-chromosome-encoded counterpart of KDM6A. UTY is also known as KDM6C (Allis *et al*., 2007).

KDM6B, unlike KDM6A or UTY, lacks the tetratricopeptide repeats (TPRs) (Figure 5-1). TPRs are important for protein-protein interactions (Van der Meulen *et al*., 2014), and the fact that KDM6B does not have them might suggest a lack of redundant functions between KDM6A and KDM6B. The *KDM6A* gene is located on chromosome X in eutherian mammals. Although not located in the pseudoautosomal region, *KDM6A* escapes X-inactivation (Greenfield *et al*., 1998). Eutherian females therefore have higher levels of this dosage-sensitive protein than eutherian males. However, eutherian males have a gene called *UTY,* the homologue of *KDM6A* on the Y chromosome, which shows similar patterns of expression to *KDM6A* (Xu *et al*., 2008). The UTY protein was thought to be enzymatically inactive (Shpargel *et al*., 2012) but recent studies suggest it may have some residual activity.

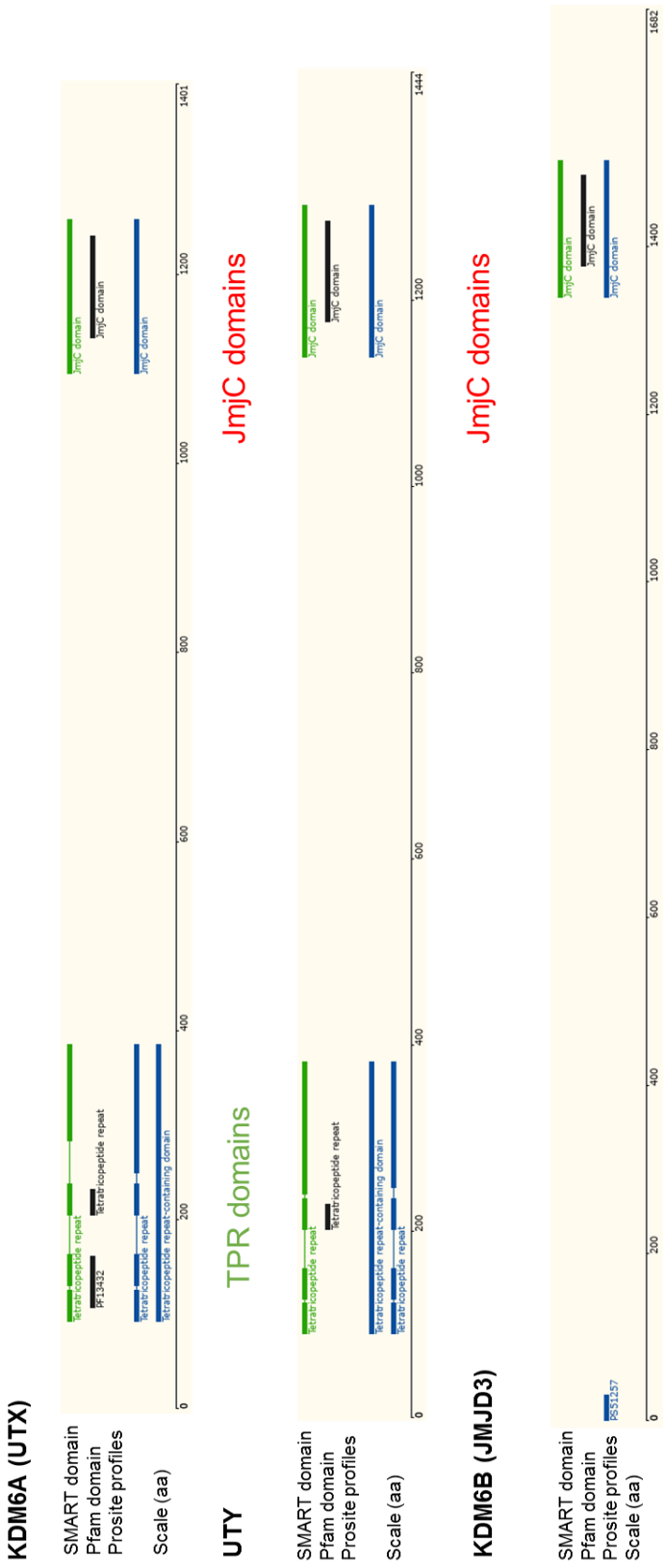

Figure 5-1 Protein domains of H3K27 demethylases. The main transcripts and their translated protein domains as identified by SMART, Pfam and Prosite tools are shown. Information downloaded from Ensembl database. KDM6A and UTY both have TPR domains and all three proteins have a catalytic JmjC domain. Scale shows the length of the proteins in amino acid residues.

As mentioned briefly in section 1.2.3.1, complete deletion of *Kdm6a* in mice results in fully penetrant embryonic lethality (Welstead *et al*., 2012). Shpargel *et al*. (2012) described *Kdm6a*$^{-/-}$ embryos at E10.5 stage to display severe deformities of neural tube, yolk sac and heart, and their entire development was delayed. These embryos were dead and reabsorbed by stage E12.5. Surprisingly, there was a subset of *Kdm6a*-null male mice (~25%) who survive to adulthood. These mice were smaller with reduced lifespan, but they were fertile and viable. This disparity was thought to be due to compensation for the absence of *Kdm6a* by *Uty*. Further experiments generated *Kdm6a*$^{-}$ *Uty*$^{-}$ male mice, which shared the fate of female homozygous *Kdm6a*$^{-/-}$ mice and died *in utero* with heart deformities.

Thus, although UTY was previously thought to be catalycally inactive (Shpargel *et al*., 2012), these results indicate that UTY performs some function. Indeed, Walport *et al*. (2014) found that human UTY does have a low level of H3K27me3 demethylase activity *in vitro*. Together with the knockout mouse studies, this suggests that UTY might be functionally redundant with KDM6A at least in some cells and tissues. However, it is not certain whether this depends on its demethylase activity or its protein-protein interactions.

A recent study identified *KDM6B* as an oncogene and *KDM6A* as a tumour suppressor gene in T-cell acute lymphoblastic leukaemia (T-ALL), but did not completely address the role of UTY in T-ALL (Ntziachristos *et al*. 2014). Further understanding of the activity of these enzymes, particularly UTY, in acute monocytic leukaemia cell lines would potentially be helpful for cancer treatments and improvement of outcomes.

5.1.1 Aim of this chapter

The aim of this chapter was to assess the importance of these three demethylase enzymes in the monocytic leukaemia cell line THP-1, and their impact on cell viability and ability to differentiate these cells into macrophages using a range of bioinformatics analysis tools and experimental approaches.

The focus of the bioinformatic investigation was on the relationship between *KDM6A* and *UTY* due to their localisation on sex chromosomes and the possibility that they might have complementary roles. The expression of *KDM6A* and *UTY* was assessed using FANTOM 5 datasets and available genome data from the Ensembl browser (see Methods below). The THP-1 cell line was used to experimentally test how perturbing gene expression or the catalytic function of these demethylases influenced the viability of the leukaemic cells and/or their competence in cell differentiation. THP-1 cells contain a deletion of the first exons (1 to 16) in *KDM6A*, and do not produce KDM6A protein at all (van Haaften *et al*., 2009). Consequently, studying THP-1 cell line helps to find more clues about UTY functionality. As mentioned in previous chapters, *KDM6B* expression is upregulated during early stages of THP-1 differentiation to macrophages and this enzyme thus might be one of the key factors that control changes in cell state and/or cell lineage definition. Perturbing the expression of *KDM6B* and *UTY* was attempted by silencing their expression by RNAi and using specific inhibitors against their enzymatic activity.

5.2 Materials and methods

5.2.1 Bioinformatic analysis of *KDM6A*, *KDM6B* and *UTY* genes

Gene structure, protein sequences and other gene information if not mentioned otherwise were taken from Ensembl database (https://www.ensembl.org; (Yates *et al*., 2016). Data is up to date as of September 2017. Protein sequences were aligned using MegAlign and MegAlign Pro; sequences were visualised using SeqBuilder (Lasergene suite 14, DNASTAR, Inc, Madison, Wisconsin, USA). Domains in protein sequences were determined using ScanProsite (http://prosite.expasy.org/scanprosite/) (de Castro *et al*., 2006). Human FANTOM 5-time course and baseline data was downloaded from http://fantom.gsc.riken.jp\zenbu (Severin *et al*., 2014). TPM for *KDM6A* and *UTY* were determined by highlighting the entire promoter region of the genes. Samples with *UTY* gene expression of 0 TPM were filtered to remove female samples; where available the sex of the sample was confirmed from sample information (http://fantom.gsc.riken.jp/5/sstar/Main_Page). For the remaining samples TPM for *UTY* and *KDM6A* were plotted against each other. Cancer samples were separated manually from tissue samples based on information at http://fantom.gsc.riken.jp/5/sstar/Main_Page. A full list of sample names and their respective TPM used for this study is found in Appendix 8.6. The *KDM6B, KDM6A* and *UTY* promoter structures were also assessed using FANTOM 5-time course and static datasets (Arner *et al*., 2015, Forrest *et al*., 2014). Transcription factor binding sites associated with *KDM6A*, *UTY* and *KDM6B* were discovered using Harmonizome available at http://amp.pharm.mssm.edu/Harmonizome/ (Rouillard *et al*., 2016).

5.2.2 RNA interference

siRNA pools (GE Dharmacon) against *UTY* and *KDM6B* were used as described in section 2.3. Both *UTY* and *KDM6B* siRNAs were employed in initial experiment to test the efficacy of siRNA pools and for validating their mRNA expression silencing capacity for the subsequent 6-well CAGE experiments.

5.2.3 Inhibition KDM6A, KDM6B and UTY enzymatic activity

GSK-J4 and GSK-J5 (Cayman chemicals, Ann Arbor, MI, USA) were diluted in DMSO (Sigma-Aldrich) at a stock concentration of 1 mg/ml (2.2 mM) and frozen at -20 ºC in aliquots. The stock was then diluted to the desired concentration in THP-1 media and added to cells. Cytotoxicity of different concentrations of both, GSK-J4 and GSK-J5, was tested using the MTT assay (as described in section 2.4.1). For cell adherence assays, THP-1 cells were seeded in 6-well plates at a density of $5x10^5$ cells per plate. Cells were differentiated with 30 ng/ml PMA as described in section 2.1, with the inhibitor added at the same time. Non-adherent cells, collected by removing the medium without lifting the adherent cells, were counted with a haemocytometer after 24 and 48 hours post differentiation and treatment with inhibitors. DMSO without PMA control was included to see if DMSO concentration alone could trigger THP-1 differentiation. The cell permeable isomer of GSK-J4, GSK-J5 was used as an inactive control of GSK-J4 (Kruidenier *et al*., 2012).

5.3 Results

5.3.1 H3K27 demethylases – a bioinformatics study

5.3.1.1 Genes, transcripts and protein comparison

Firstly, human *KDM6A, KDM6B* and *UTY* gene structures were compared based on the Ensembl database and FANTOM 5 data to gain more knowledge about these genes which would be needed for genetic editing at the later stage of the project.

KDM6B gene has 6 different splice variants, with only different 3 protein-coding transcripts (Table 5-1). The shortest protein-coding transcript (203, ENST00000570632.1) does not code for the KDM6B catalytic domain, whereas the other two do. These two long transcripts (6713 bp and 5422 bp) code for similarly long proteins (1682 amino acids and 1643 amino acids, respectively). Out of these two, the longer transcript (201, ENST00000254846.9) was selected from Ensembl using APPRIS bioinformatic tools (Rodriguez *et al*., 2013) as a candidate principal isoform, which was used for further analysis.

	Transcript	Ensembl ID	Type	Length (bp)	Protein length (aa)	Domains
1	201	ENST00000254846.9	Protein-coding	6713	1688	JmjC
2	202	ENST00000448097.6	Protein-coding	5422	1643	JmjC
3	203	ENST00000570632.1	Protein-coding	1339	361	none
4	205	ENST00000572030.1	Processed transcript	432	NA	NA
5	204	ENST00000571047.5	Processed transcript	362	NA	NA
6	206	ENST00000575521.1	Processed transcript	219	NA	NA

Table 5-1 KDM6B and its transcripts and encoded proteins as identified by Ensembl datasets.

Examination of the *KDM6B* promoter region in humans revealed a more complex picture (Figure 5-2). Looking at FANTOM 5 data, which show transcription start sites (TSS), the highest number of TSSs for *KDM6B* seem to be in an area within the first intron of the 201 transcript, not at the beginning of the identified transcripts. This region is annotated as a CpG island, and three different enhancers have also been identified in this region through bioinformatics analysis (Andersson *et al*., 2014). FANTOM 5 has identified the main promoter p1 for *KDM6B* in the centre of the gene (labelled as p1@KDM6B in intron 1/2), which is not consistent with the Ensembl and Gencode gene models (Figure 5-2). For further *KDM6B* gene expression analysis, the

region from the beginning of the gene to the end of the CpG island (boxed in Figure 5-2) was taken into consideration.

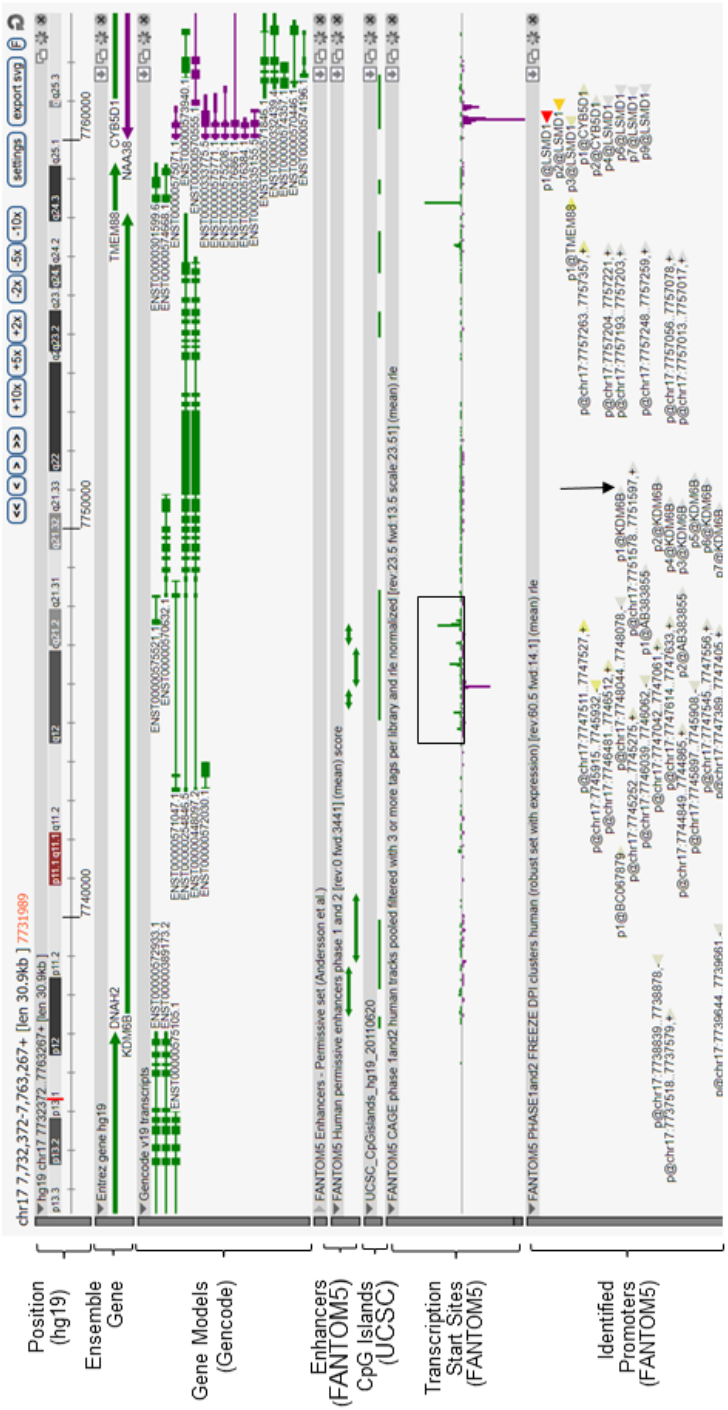

Figure 5-2 KDM6B gene structure and its TSS. Image taken from FANTOM5 genome viewer Zenbu (section 5.2.1). Upper tracks show the position, extent of the locus as determined by Ensembl, gene models are from Gencode data, enhancers are from FANTOM5 data and identified CpG islands from UCSC data. Lower tracks show the number of tags at the TSS detected in the FANTOM5 study and the promoters identified after clustering of TSS (Forrest et al 2014). Green indicates transcription from the forward strand; purple indicates transcription from the reverse strand. TSS associated with KDM6B are shown by the box and p1@KDM6B is indicated by the arrow.

According to Ensembl, the *KDM6A* gene has 14 different transcripts, 10 of which are predicted to be protein encoding (Table 5-2). The transcripts range from lengths of 5924 bp to 671 bp. Some shorter protein-coding splice variants do not code for KDM6A JmjC or TPR functional domains (for example transcripts 205, 203, 206 and 214). Interestingly, Ensembl has flagged the second longest protein-coding transcript rather than the longest one as its principal candidate isoform by the bioinformatics method APPRIS (Rodriguez *et al*., 2013). The principal isoform (currently named KDM6A-201), which codes for a 1401 amino acid long protein, has been chosen for further analysis in this study. The FANTOM 5 study identified only the main promoter (p1@KDM6A) at the start of the gene, no other alternative promoters can be detected or have been described (Figure 5-3).

	TRANSCRIPT	ENSEMBL ID	TYPE	LENGTH (BP)	PROTEIN LENGTH (AA)	DOMAINS
1	201	ENST00000377967.8	Protein-coding	5438	1401	TPR, JmjC
2	213	ENST00000611820.4	Protein-coding	5924	1429	TPR, JmjC
3	202	ENST00000382899.8	Protein-coding	5789	1384	TPR, JmjC
4	212	ENST00000543216.5	Protein-coding	5655	1269	TPR, JmjC
5	211	ENST00000536777.5	Protein-coding	5633	1332	TPR, JmjC
6	205	ENST00000433797.5	Protein-coding	4324	1044	JmjC
7	203	ENST00000414389.5	Protein-coding	4189	999	JmjC
8	214	ENST00000621147.4	Protein-coding	2876	224	TPR
9	204	ENST00000431196.2	Protein-coding	735	161	none
10	206	ENST00000451692.5	Protein-coding	671	224	TPR
11	208	ENST00000479423.1	Processed transcript	768	**NA**	**NA**
12	209	ENST00000484732.1	Processed transcript	640	**NA**	**NA**
13	207	ENST00000475233.1	Processed transcript	612	**NA**	**NA**
14	210	ENST00000485072.5	Processed transcript	382	**NA**	**NA**

Table 5-2 KDM6A and its 14 transcripts according to Ensembl database.

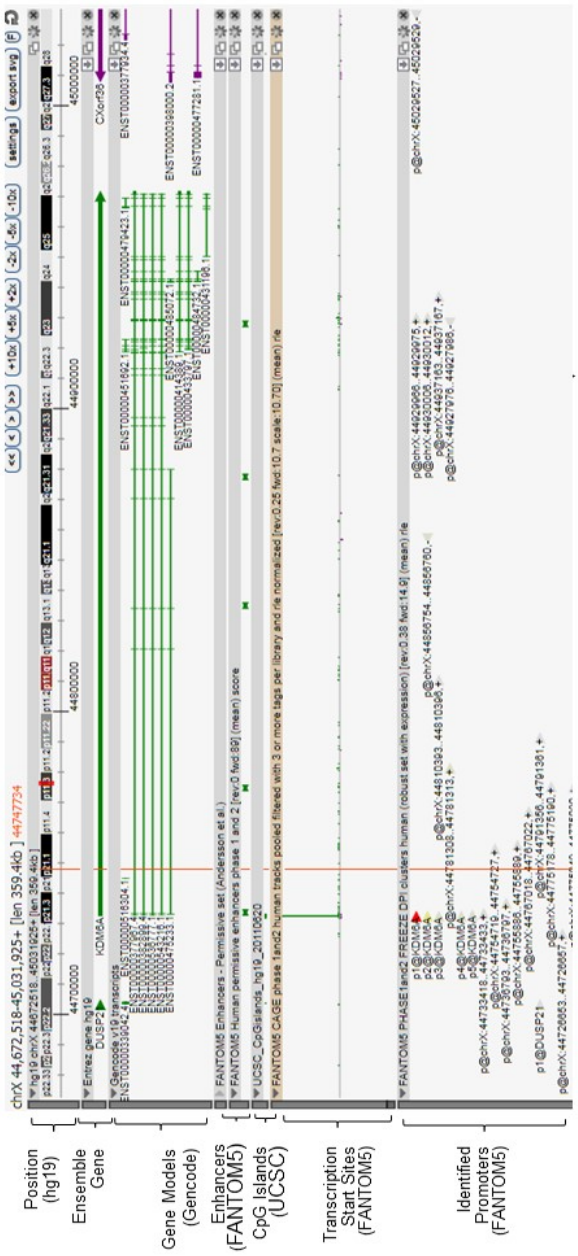

Figure 5-3 KDM6A gene structure and its TSS. Image taken from FANTOM5 genome viewer Zenbu (section 5.2.1). Upper tracks show the position, extent of gene determined by Ensembl, gene models from Gencode data, enhancers from FANTOM5 data and identified CpG islands from UCSC data. Lower tracks show the number of tags at the TSS detected in the FANTOM5 study and the promoters identified after clustering of TSS (Forrest et al., 2014). Green indicates transcription from the forward strand; purple indicates transcription from the reverse strand.

Similarly to *KDM6A, UTY* has 16 transcripts the in Ensembl database, 13 of them are protein-coding (Table 5-3), with various transcript and peptide lengths. Some splice variants do not contain the JmjC domain or full length TPR domains. None of the transcripts have been verified as a principal isoform by APPRIS for this gene based on

protein structural information, cross-validation from other species and important protein residues (Rodriguez *et al*., 2013). For further analysis, the longest transcript (currently UTY-212, as of August 2017) was chosen (6817 bp, which encodes a 1444 amino acid protein). FANTOM 5 transcription start site database did not show any alternative promoters to the main promoter named p1@UTY (Figure 5-4).

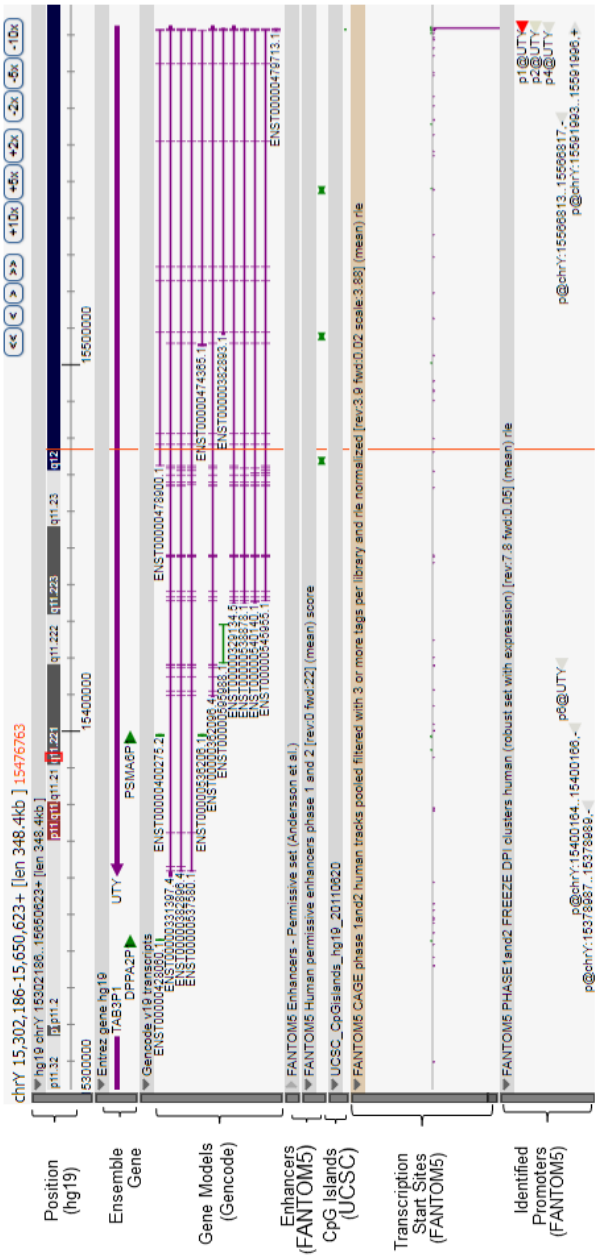

Figure 5-4 UTY gene structure and its TSS. Image taken from FANTOM5 genome viewer Zenbu (section 5.2.1). Upper tracks show the position, extent of the locus as determined by Ensembl, gene models are from Gencode data, enhancers are from FANTOM5 data and identified CpG islands from UCSC data. Lower tracks show the number of tags at the TSS detected in the FANTOM5 study and the promoters identified after clustering of TSS (Forrest et al 2014). Green indicates transcription from the forward strand; purple indicates transcription from the reverse strand. P1@UTY is the main promoter identified by FANTOM5 (red arrowhead), and no other TSS were found to be expressed at high rate.

	TRANSCRIPT	ENSEMBL ID	TYPE	LENGTH (BP)	PROTEIN LENGTH (AA)	DOMAINS
1	212	ENST00000545955.5	Protein-coding	6817	1444	TPR, JmjC
2	214	ENST00000617789.4	Protein-coding	6682	1399	TPR, JmjC
3	205	ENST00000382896.8	Protein-coding	6661	1392	TPR, JmjC
4	211	ENST00000540140.5	Protein-coding	6652	1389	TPR, JmjC
5	213	ENST00000612274.4	Protein-coding	6586	1667	TPR, JmjC
6	210	ENST00000538878.5	Protein-coding	6574	1363	TPR, JmjC
7	202	ENST00000331397.8	Protein-coding	6529	1347	TPR, JmjC
8	209	ENST00000537580.5	Protein-coding	6490	1335	TPR, JmjC
9	215	ENST00000618474.4	Protein-coding	6277	1264	TPR, JmjC
10	203	ENST00000362096.8	Protein-coding	4990	1240	TPR, JmjC
11	201	ENST00000329134.9	Protein-coding	4325	1079	TPR
12	216	ENST00000624098.3	Protein-coding	3636	1211	TPR, JmjC
13	204	ENST00000382893.2	Protein-coding	1539	207	TPR
14	206	ENST00000474365.1	Processed transcript	1387	NA	NA
15	207	ENST00000478900.5	Processed transcript	747	NA	NA
16	208	ENST00000479713.1	Processed transcript	686	NA	NA

Table 5-3 UTY and its 16 transcripts according to the Ensembl database.

Human *KDM6A* and *UTY* show up to 88% cDNA homology and up to 86% amino acid homology (Appendix 8.7, Ensembl database, accessed November 2016). *KDM6B* showed only 29% cDNA homology with both *KDM6A* and *UTY*, and even less for the protein. KDM6A and UTY proteins look very similar, which is consistent with their amino acid sequence homology (Figure 5-1). Unlike KDM6B, both contain 7 TPR domains. However, all three proteins contain a catalytic JmjC domain, which is conserved across all members of the KDM family (Klose *et al*., 2006).

Because of the high similarity between the KDM6A and UTY proteins, their amino acid sequences were compared further using the MegAlign programme. The results show that the sequences of the TPR and JmjC functional domains are highly homologous between both proteins (Figure 5-5, Figure 5-6). The catalytic residues (red asterisks in Figure 5-6) in their JmjC domains are also sequence conserved. Walport *et al*. (2014) reported that a G -> S substitution in the JmjC domain (green rectangle in Figure 5-6) reduces the UTY demethylase activity compared to KDM6A. The linker sequence between the domains is not as highly conserved compared to both functional domains (not shown).

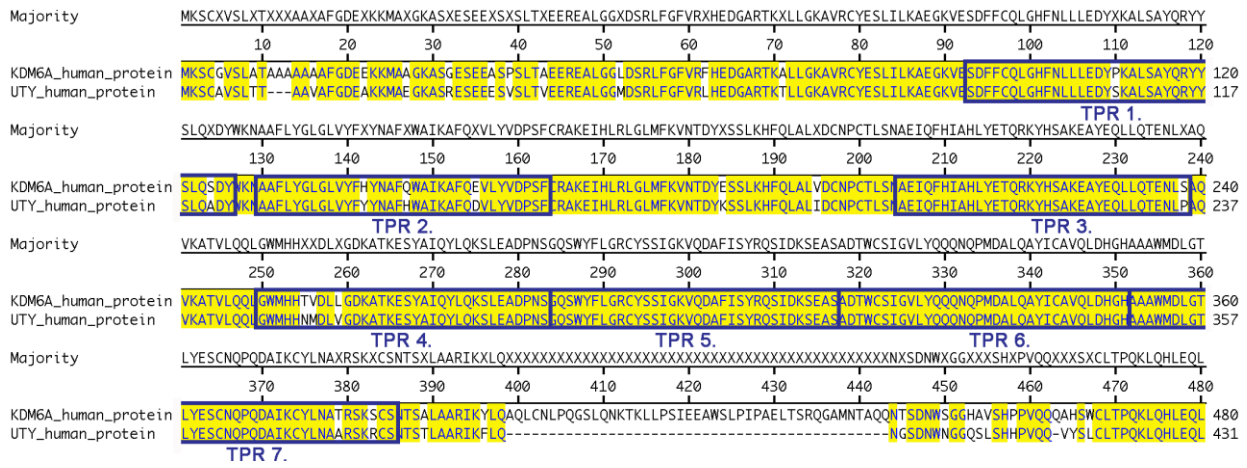

Figure 5-5 MegAlign CLUSTAL W comparison between Ensembl predicted amino acid sequences of the KDM6A and UTY main transcripts. 7 TPR domains are boxed in blue. Highlighted yellow residues indicate sequence conservation.

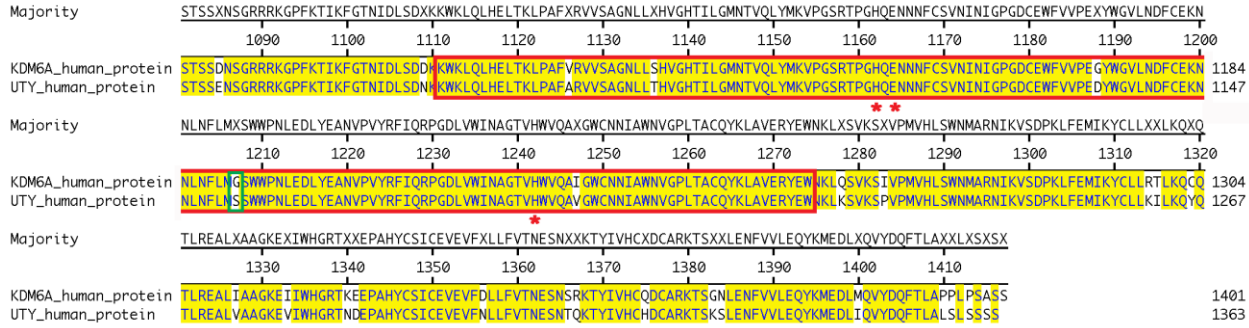

Figure 5-6 MegAlign CLUSTAL W comparison between Ensembl amino acid sequences of KDM6A and UTY main transcripts. JmjC domain is boxed in red. Highlighted yellow residues indicate sequence conservation between both proteins. Red asterisks denote catalytic sites in the JmjC domains. The green rectangle points to the amino acid change that has been previously reported to be associated with a reduction of the catalytic activity of UTY (Walport et al, 2014).

5.3.1.2 Investigating co-regulation of *UTY* and *KDM6A*

Due to the similarity of the UTY and KDM6A proteins, the vast FANTOM5 dataset was analysed to determine whether or not the genes were expressed at the same time in the same tissues (suggesting co-regulation) or with opposite expression patterns (suggesting functional redundancy in different tissues). An initial survey showed that most samples did not express *UTY* (Table 5-4). However, all testes and prostate samples did express *UTY* (Appendix 8.6), suggesting that those not expressing UTY were from female donors. Where possible this was validated using the metadata for samples available as part of the FANTOM5 data set (section 5.2.1). All samples with *UTY* TPM = 0 were excluded from further analysis (Table 5-4) to properly assess the

gene expression correlation of *KDM6A* and *UTY*. *KDM6A* and *UTY* TPM values were positively correlated (Pearson's correlation = 0.720, Figure 5-7). These results show that *UTY* probably does not substitute for *KDM6A* in *KDM6A* low expression cells, since when there is a high *UTY* expression, there is also high *KDM6A* expression. However, there seems to be less correlation between *UTY* and *KDM6A* expression levels in cancerous tissue as further analysed below.

SAMPLE TYPE	SUBSET	NUMBER OF SAMPLES	CORRELATION COEFFICIENT
ALL	Together	1829	
	UTY TPM = 0	1312	
	UTY TPM > 0	517	0.720
TISSUES	Together	1358	
	UTY TPM = 0	919	
	UTY TPM > 0	439	0.727
CANCER	Together	471	
	UTY TPM = 0	393	
	UTY TPM > 0	78	0.322

Table 5-4 FANTOM5 static and time course samples. Pearson's correlation coefficient between expression in non-zero UTY samples and KDM6A was calculated in Microsoft Excel.

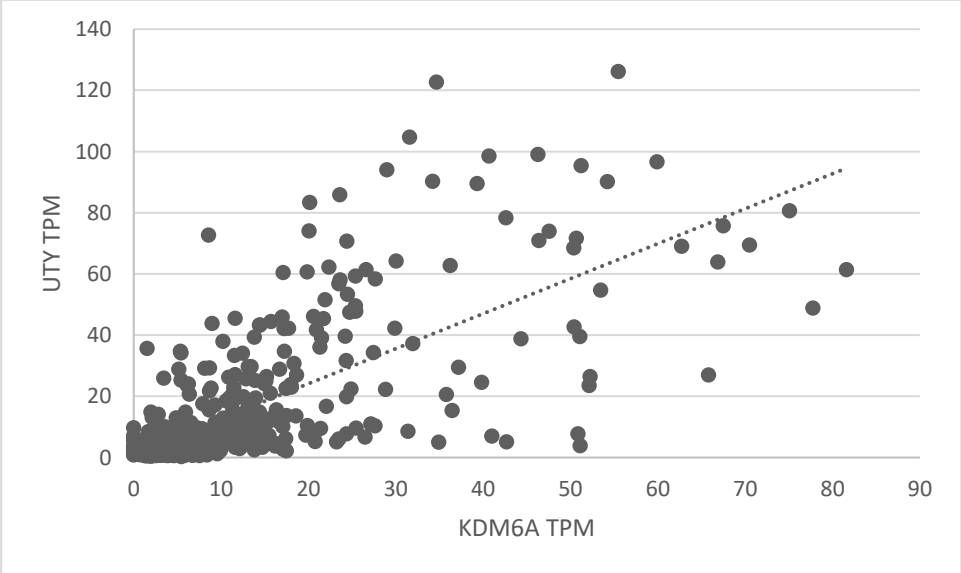

Figure 5-7 Correlation of mRNA expression of UTY versus KDM6A for all male FANTOM5 human samples. Samples which had no expression of UTY were taken out to accurately assess the correlation coefficient. Pearson's correlation coefficient = 0.720, N= 517.

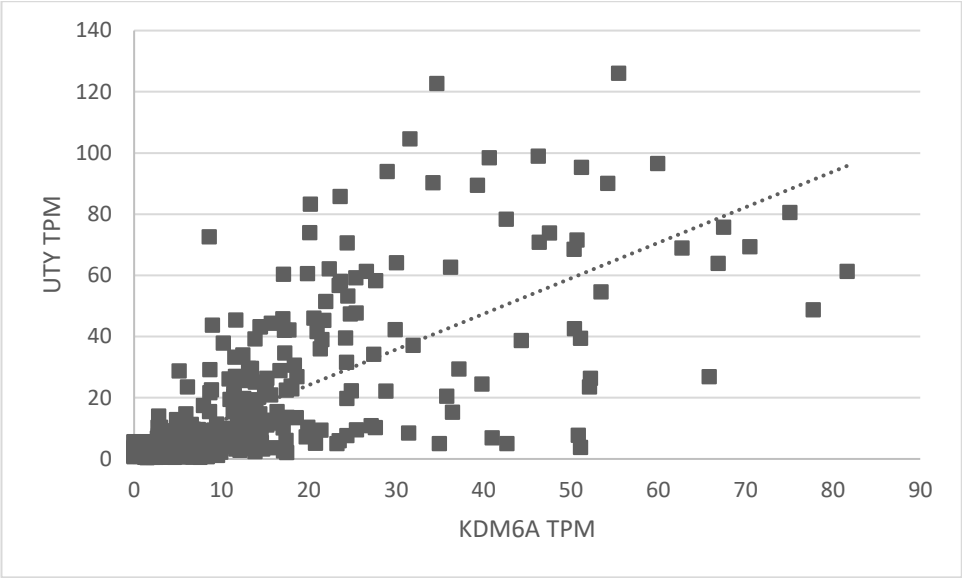

Figure 5-8 Correlation of mRNA expression of UTY versus KDM6A for all male tissues and non-cancerous cell samples. Pearson's correlation coefficient =0.727, N= 439.

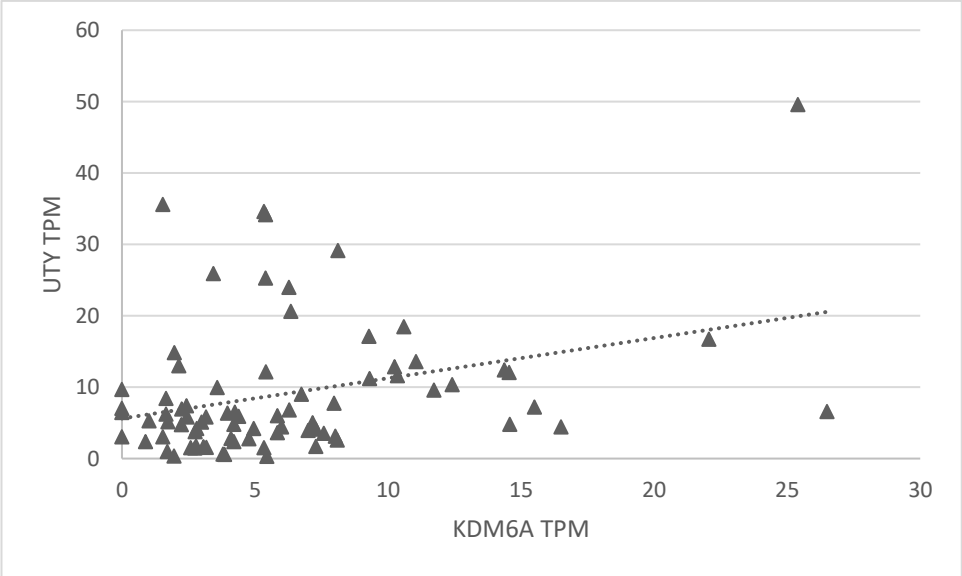

Figure 5-9 Correlation of mRNA expression of UTY versus KDM6A for all cancer samples. Samples with no UTY expression were removed. Pearson's correlation coefficient = 0.322, N= 78.

Interestingly, comparing tissue only data (Figure 5-8) to cancer samples (Figure 5-9), the high correlation in tissues did not apply to cancerous samples (Pearson's correlation 0.727 vs 0.322). There were fewer cancer samples than tissue samples, but modelling using a simulated larger dataset with the same distribution of expression levels did not change the correlation coefficient. *KDM6B* expression did not have a high positive correlation with *UTY* or *KDM6A* expression (Table 5-5).

COMPARISON	SAMPLE SIZE	CORRELATION COEFFICIENT
KDM6B* VS *KDM6A	1829	0.469
***KDM6B* VS *UTY* (*UTY* TPM > 0)**	517	0.261

Table 5-5 FANTOM 5 samples comparison for KDM6B with UTY and KDM6A.

Next, potential co-regulation of *KDM6A* and *UTY* by the same transcription factors was assessed. Alignment analysis of promoter regions of *KDM6A* and *UTY* in human sequences (where the transcription factors would bind) downloaded from FANTOM 5 dataset, using MegAlign Clustal W, showed the sequences were not strikingly similar (data not shown).

Using protein and gene database Harmonizome, different predicted transcription factor binding sites (TF) for *KDM6A*, *UTY* and *KDM6B* were identified. Various databases for transcription factors binding sites were used such as TRANSFAC, JASPAR, CHEA and ENCODE (Matys *et al*., 2003, Matys *et al*., 2006, Mathelier *et al*., 2014, Sandelin *et al*., 2004, Lachmann *et al*., 2010, ENCODE, 2004). TRANSFAC and JASPAR predict TF binding using known binding site motifs, whereas CHEA and ENCODE use ChIPseq data. The TRANSFAC datas*et al*so provided curated data, which were manually selected from low-throughput or high-throughput TF functional studies. The data from these five different sources gave very variable results, with very few TF appearing in more than 2 studies (all TF tables are available in Appendix 8.8).

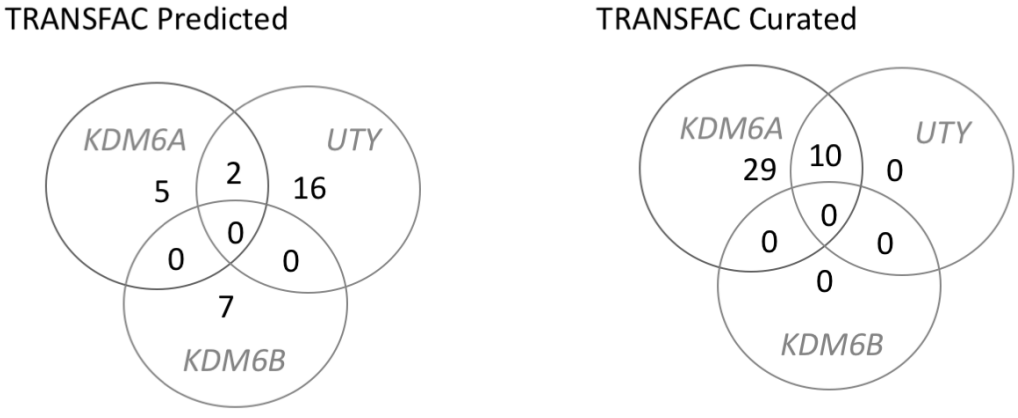

Figure 5-10 TRANSFAC predicted and curated transcription factor binding sites in KDM6A, KDM6B and UTY promoters. Data taken from Harmonizome.

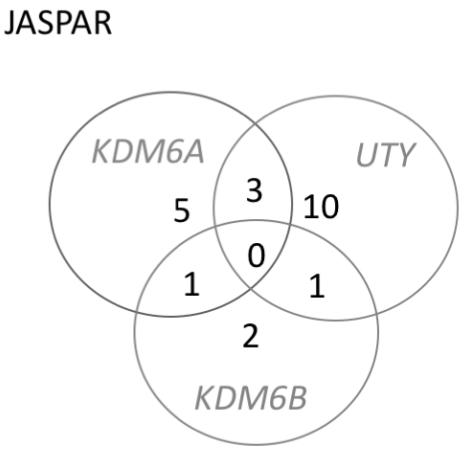

Figure 5-11 JASPAR predicted transcription factor binding in KDM6A, KDM6B and UTY promoters. Data taken from Harmonizome.

A number of TF binding sites appeared common to *KDM6A* and *UTY*, especially in TRANSFAC curated dataset (Figure 5-10), where all 10 *UTY* TF were shared by *KDM6A*. Fewer TF sites were shared between *KDM6A* and *KDM6B* or *UTY* and *KDM6B* in all three datasets, with no TF shared by all three gene promoters (Figure 5-10, Figure 5-11)

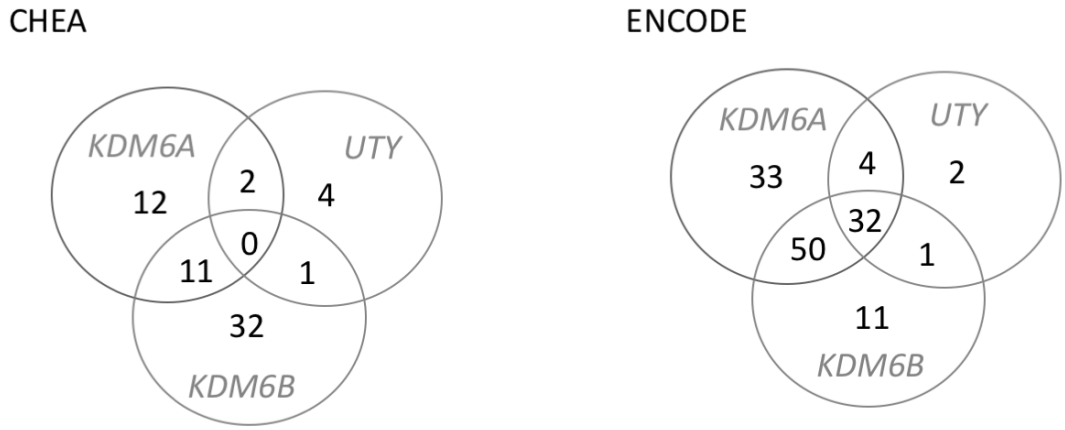

Figure 5-12 CHEA and ENCODE transcription factor binding based on ChIPseq studies in KDM6A, KDM6B and UTY promoters. Data taken from Harmonizome.

This trend is different in ChIP-seq based data (Figure 5-12) where the *UTY* promoter does not have as many TF binding sites as the other two. In both CHEA and ENCODE datasets *KDM6A* and *KDM6B* have a number of TF binding sites in common, unlike *UTY*. The ENCODE dataset also shows 32 different TF sites which are common to all three gene promoters.

Using Harmonizome and its pathway commons protein-protein interactions database, KDM6A and UTY share a higher number of common protein interaction partners, even though KDM6B has a higher number of targets overall (Figure 5-13). This database also shows that UTY and KDM6A interact together (Appendix 8.8).

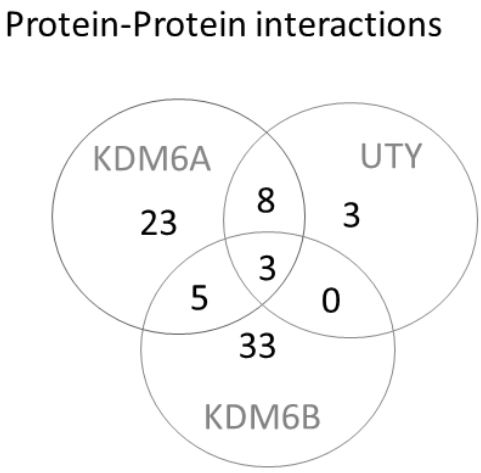

Figure 5-13 Pathway Commons protein-protein interactions of KDM6A, KDM6B and UTY. Data taken from Harmonizome.

5.3.1.3 Evolutionary analysis of *UTY* and *KDM6A*

Surveying orthologues of genes in other species provides clues about evolutionary constraints and thus the importance and possible function of a gene itself. In this section, the evolutionary relationships of the demethylases are examined, with the emphasis on the *KDM6A* and *UTY* genes.

According to Ensembl database, the *KDM6B* gene has 82 orthologues, *KDM6A* has 90 and *UTY* 51 orthologues in different species. Placental mammal orthologues are most similar to the human ones, but depending on the evolutionary distance of other species from humans, the less clear become the available annotations. However, even in animals with distant shared ancestors, for example in *C. elegans*, there is a gene encoding a JmjC domain, however, based on the available information of the protein

structure and its encoded domains this gene is probably not a homologue or paralogue of mammalian *KDM6A*, *KDM6B* or *UTY* as its protein structure is much less complex. *KDM6A* is located on human chromosome band Xp11.3 and at a syntenic region in other eutherian mammals, whereas it is located on an autosome (chromosome 4) in opossum (a marsupial) according to Ensembl database and UltraContig222 and is predicted to be located on chromosome 18p in platypus (a monotreme) (Veyrunes *et al*., 2008). *KDM6A* is also located on autosomes in birds and fish (Ensembl database). This would suggest that the *KDM6A* gene became associated with the sex chromosome with a block of other genes, somewhere at the time when the common ancestor for all eutherian mammals evolved. Therefore, only eutherian males do not have two expressing alleles of *KDM6A,* as *KDM6A* escapes X inactivation in human females (Greenfield *et al*., 1998).

UTY lies on the mammalian Y chromosome, right next to two other genes with X chromosome equivalents, *USP9Y* and *DDX3Y* and a testis specific non-coding element *TTTY15*. *KDM6A* also neighbours the X-equivalent genes, *USP9X* and *DDX3X*, although not as close by. There was only moderate correlation when comparing *TTTY15* gene expression with *UTY* or *DDX3Y* in the FANTOM5 human dataset (Figure 5-14, Figure 5-15). There are almost no gene expression data for *USP9Y* in the FANTOM5 results, and the reason for this lack of data is not clear. According to data from BioGPS database, *USP9Y* should be expressed at least in T-cells (Wu *et al*., 2016).

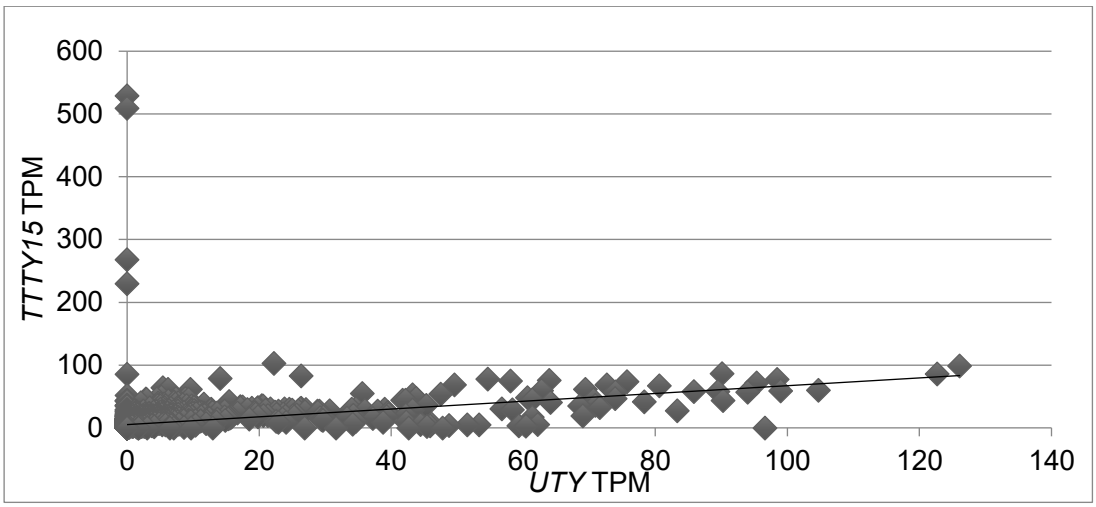

Figure 5-14 Correlation of RNA expression of TTTY15 versus UTY for all FANTOM5 human samples, including time courses. Pearson's correlation coefficient = 0.358, N= 1829.

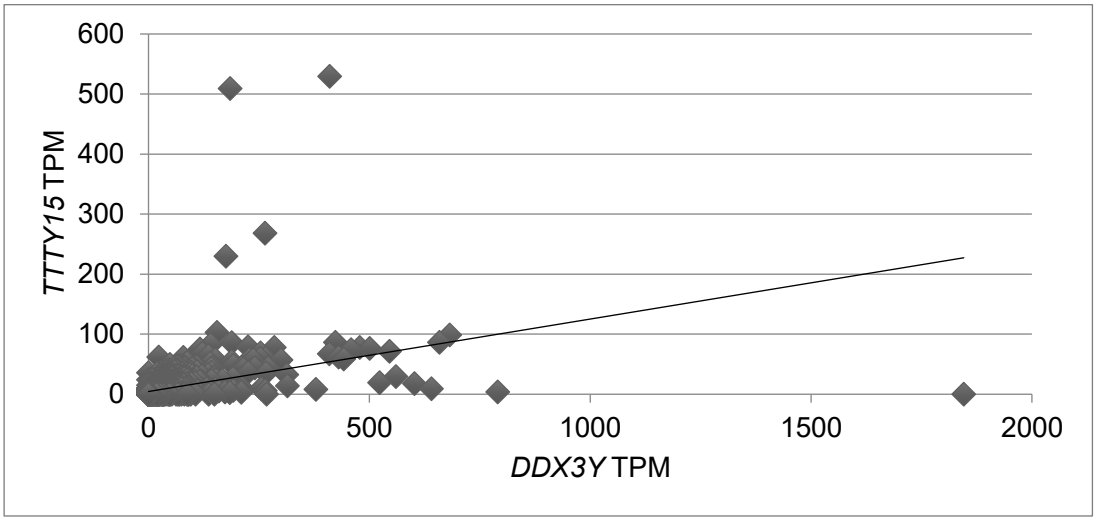

Figure 5-15 Correlation of RNA expression of TTTY15 versus DDXY3 for all FANTOM5 human samples. Pearson's correlation coefficient = 0.409, N= 1829.

5.3.1.4 Conclusions of bioinformatics analyses

In conclusion, analysis of available data for KDM6A and UTY shows that they share many similarities, in both amino acid sequence homology (Figure 5-5, Figure 5-6) and regulation of gene expression or common protein interactions (Figure 5-10, Figure 5-13). In contrast, KDM6B shows only a low level of similarity to either KDM6A or UTY.

Perturbing the gene expression or inhibiting the demethylase activity of these enzymes in THP-1 cells should show whether these demethylases can affect the normal function

of leukaemic cell lines and their differentiation from monocyte-like cells into macrophages. The next section reports the results of experiments changing expression of these genes through the use of RNAi as well as using an inhibitor of the catalytic activity of their JmjC domains.

5.3.2 The inhibitor GSK-J4 does not prevent THP-1 cells from differentiating into adherent macrophage-like cells

The inhibitor GSK-J4 has been found to specifically inhibit the demethylase activity of both KDM6B and KDM6A, with no data available of testing its effect on UTY (Kruidenier *et al*., 2012, Heinemann *et al*., 2014). To assess whether the ability of THP-1 cells to differentiate into macrophages is dependent on the demethylase activity of KDM6B or UTY, adherence assays were performed.

Firstly, an MTT assay of inhibitor toxicity was carried out on THP-1 cells. Identical numbers of THP-1 cells were cultured in varying concentrations of both GSK-J4 (inhibitor) or GSK-J5 (control) for 48 hours, and then their metabolic activity was measured by a colorimetric MTT assay. Results showed (Figure 5-16) that GSK-J4 had higher toxicity effect on THP-1 cells at lower concentrations than the control GSK-J5. Both inhibitors did not affect THP-1 viability and metabolic activity up to a concentration of 5 μM. However, from 10 μM upwards GSK-J4 decreased THP-1 viability to ~50% of THP-1 without inhibitor. GSK-J5 did not reach 50% viability until a concentration of 20-30 μM.

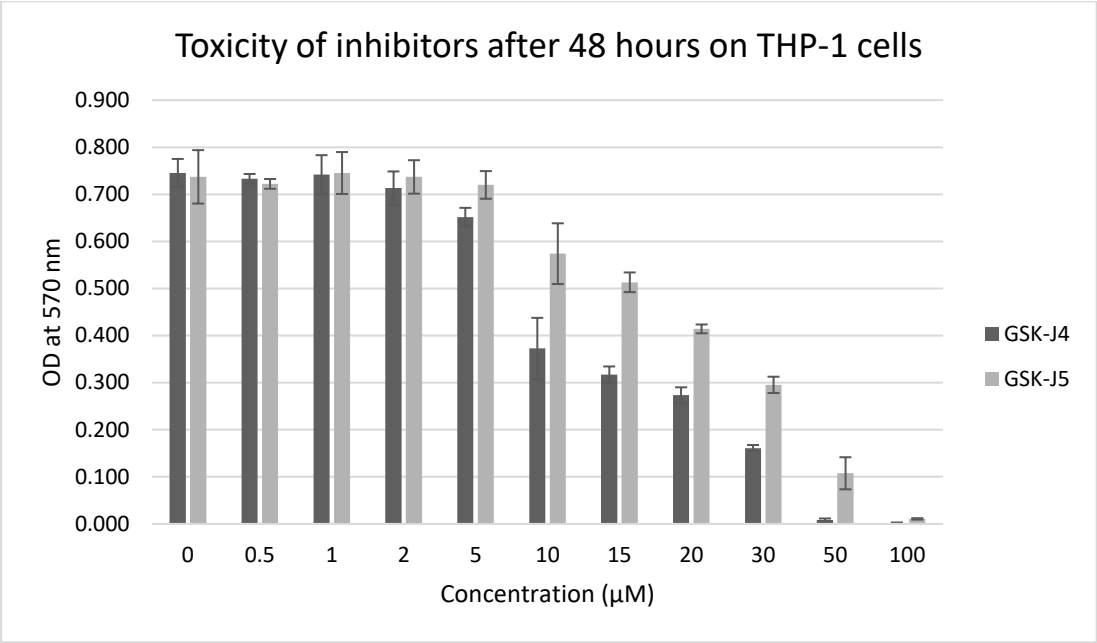

Figure 5-16 MTT assay results for THP-1 cells treated with varying concentrations of GSK-J4 and GSK-J5. Results after 48 hours in culture are shown. Results were measured at 570 nm, samples were quadruplicate. Blank measurement (only medium) was subtracted from the mean of samples. Error bars show standard deviation (SD).

A known number of THP-1 cells were then subjected to an adherence assay, where different concentrations of the inhibitors were added together with the differentiating agent PMA and adherence and morphology of the cells were assessed after 24 and 48 hours. Adherence of the cells indicates that they have differentiated towards macrophages, and failure to adhere suggests that they have not differentiated.

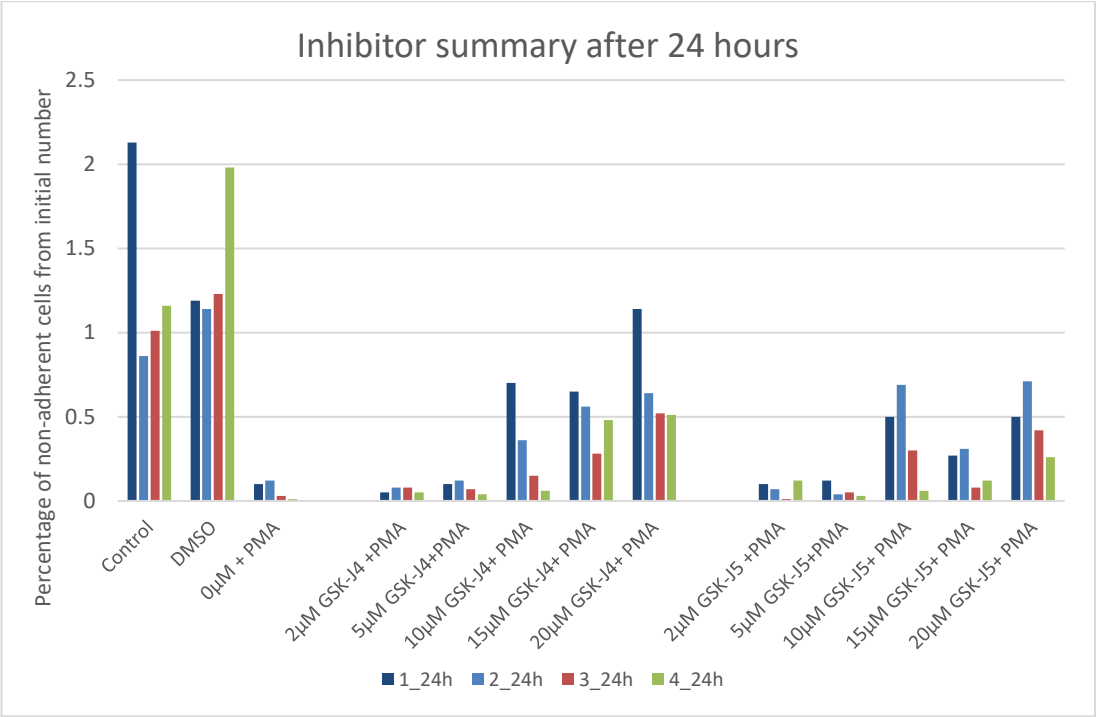

Figure 5-17 Summary of four different experiments testing the adherence of THP-1 cells 24 hours after adding PMA and varying concentrations of inhibitors. Control is THP-1 cells without any additions, DMSO is THP-1 cells with DMSO (the carrier of the inhibitors) to test whether DMSO itself make the cells differentiate.

The results of four replicates of the adherence assays (Figure 5-17, Figure 5-18, Table 5-6, Figure 5-19) are quite variable. THP-1 cells without any treatment (Control) and with control amount of the vehicle (DMSO) seemed to not differentiate and adhere to the plate in 24 or 48 hours. Control differentiation of THP-1 (PMA, without inhibitors) proceeded successfully and more than 90% of cells were adherent after 48 hours. Therefore, if the inhibitors were suppressing differentiation, the proportion of non-adherent cells would increase.

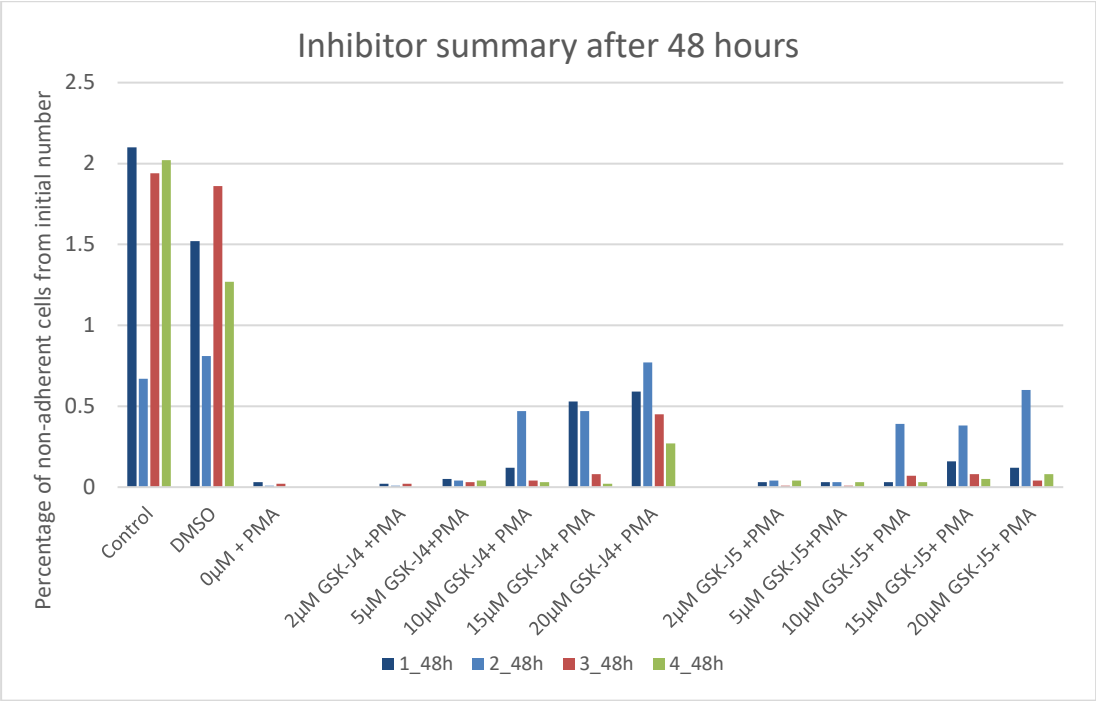

Figure 5-18 Summary of four different experiments testing the adherence of THP-1 cells 48 hours after adding PMA and varying concentrations of inhibitors. Control is THP-1 cells without anything, DMSO is THP-1 cells with DMSO (the carrier of the inhibitors) to test whether DMSO itself make the cells differentiate.

Each replicate of the assay for inhibitor GSK-J4 and the control inhibitor GSK-J5 showed different results and the average of these four replicates has a very high standard deviation (

Table 5-6, Figure 5-19). Nevertheless, a consistent finding was that a larger population of cells seemed to be non-adherent after 24 or 48 hours in high concentrations of GSK-J4 than for GSK-J5. The non-adherent cells counted were always viable, as dead cells were excluded by Trypan Blue staining, so it is not clear whether the inhibitor actually killed the cells or just stopped their metabolic activity (as detected by MTT assay) at concentrations of equal to or higher than 10 μM.

Average values	24 h	24 h SD	48 h	48 h SD
Control	1.290	0.50	1.683	0.59
DMSO	1.385	0.35	1.365	0.38
0μM + PMA	0.065	0.05	0.015	0.01
2μM GSK-J4 +PMA	0.065	0.02	0.013	0.01
5μM GSK-J4+PMA	0.083	0.03	0.040	0.01
10μM GSK-J4+ PMA	0.318	0.25	0.165	0.18
15μM GSK-J4+ PMA	0.493	0.14	0.275	0.23
20μM GSK-J4+ PMA	0.703	0.26	0.520	0.18
2μM GSK-J5 +PMA	0.075	0.04	0.030	0.01
5μM GSK-J5+PMA	0.060	0.04	0.025	0.01
10μM GSK-J5+ PMA	0.388	0.23	0.130	0.15
15μM GSK-J5+ PMA	0.195	0.10	0.168	0.13
20μM GSK-J5+ PMA	0.473	0.16	0.210	0.23

Table 5-6 Average values of percentage of non-adherent THP-1 cells (compared with initial number seeded into plates) after being treated with inhibitors for 24 hours (24 h) and 48 hours (48 h). Average values of 4 different technical replicates. Standard deviation (SD) was calculated from these 4 technical replicates.

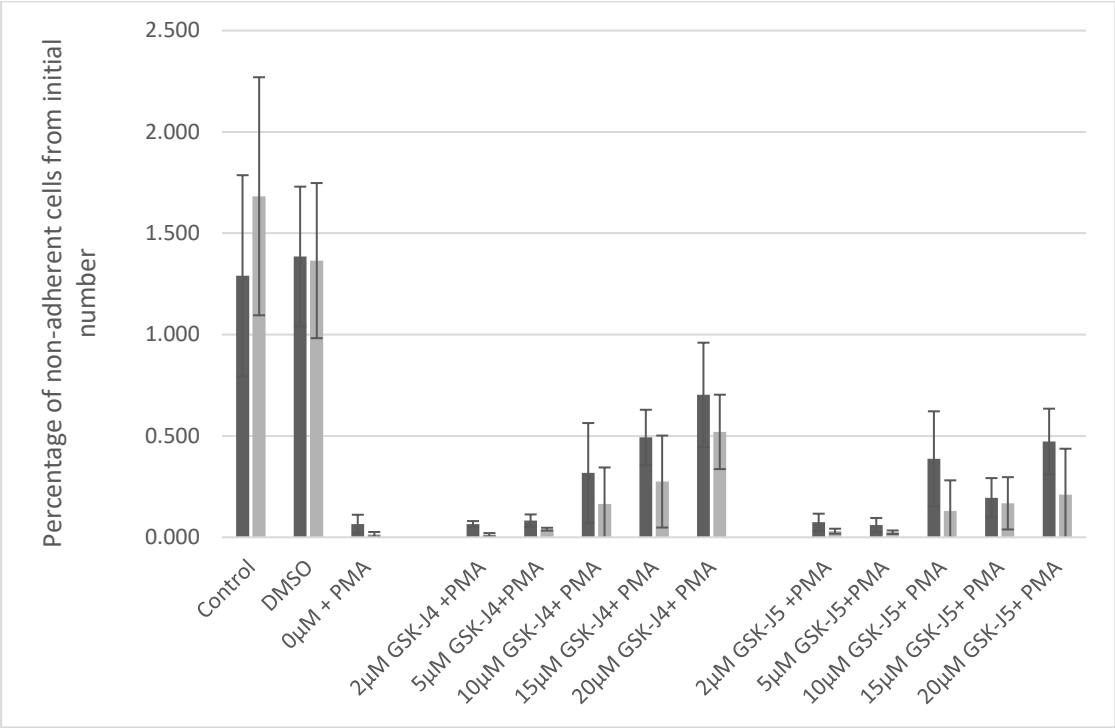

Figure 5-19 Average values of non-adherent THP-1 cells after being treated with inhibitors for 24 hours (24 h) and 48 hours (48 h). Average values of 4 different technical replicates. Error bars represent standard deviation (SD), which was calculated from these 4 technical replicates.

5.3.3 Using siRNA technology to knock down the expression of *UTY* and *KDM6B* in THP-1 cells

5.3.3.1 Preliminary experiments assessing siRNA knock down efficiency showed no effect on *UTY*, and moderate effect on *KDM6B*

RNAi was used to knock down expression of these *UTY* and *KDM6B*. A pool of four different siRNAs was ordered (Chapter 2 section 2.3) for each gene and their efficiency was tested using an 8-day time course. Lipofection with siRNA of the THP-1 cell line was carried out on day -1, and cells were stimulated by PMA to differentiate into macrophages on day 0 (section 2.1). RNA was taken on day -1, 0, 1, 2, 3 and 7.

For *KDM6B*, a Non-targeting control (NT) and normal THP-1 differentiation assay without lipofection or siRNA (control) were included. For *UTY* there was also a lipofection-only control, without any siRNA (Lipofectamine). The Non-targeting (NT) siRNA pool is specially designed to not bind anywhere specifically in the human genome, and thus is considered a perfect negative control for siRNA experiments (LifeSciences, 2017).

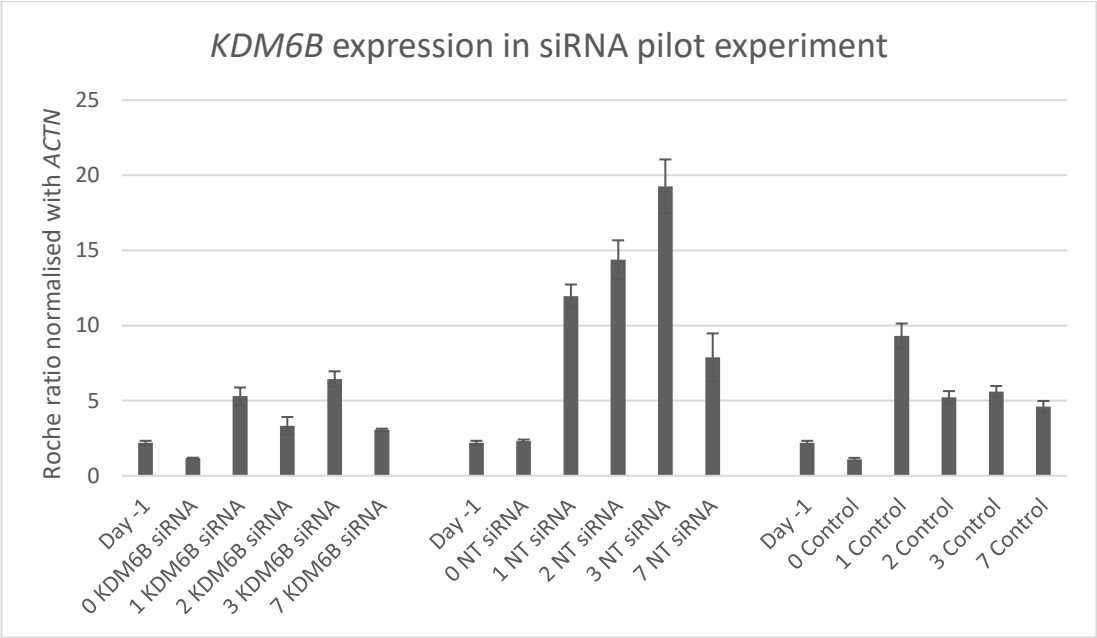

Figure 5-20 qRT-PCR results of an initial siRNA experiment for KDM6B. Pictured here is the KDM6B expression normalised to ACTN expression and error bars (calculated by the by Roche LightCycler 480 software). Error bars show standard error. Day -1 is the same for all treatments.0 means day 0, 1 is day 1 and etc. 0 KDM6B siRNA stands for day 0 sample with KDM6B siRNA, 1 NT siRNA stands for day 1 Non-targeting control siRNA, and 2 Control stands for day 2 wildtype THP-1 differentiation.

KDM6B was expressed at low levels at day -1 and 0 in control differentiation experiment (Control), but it increased 9-fold from day 0 to day 1, and then stabilised its expression for the rest of time course by day 2 (Figure 5-20). The Non-targeting control (NT) followed the pattern of the Control time course for days 0 and 1, but *KDM6B* expression continued to increase up to day 3.

siRNA pool (*KDM6B* siRNA) knocked down the expression of *KDM6B* in day 0 to about half the value of NT. The values were around the same level as the Control experiment for day 0. *KDM6B* expression at day 1 was knocked down to up to ~50% of NT or Control expression using the siRNA pool. Some level of knock down was maintained until around day 3, if compared to Control, and throughout the whole time course if compared to NT control. The cells exhibited the same macrophage morphology as controls at day 1 and later (not shown).

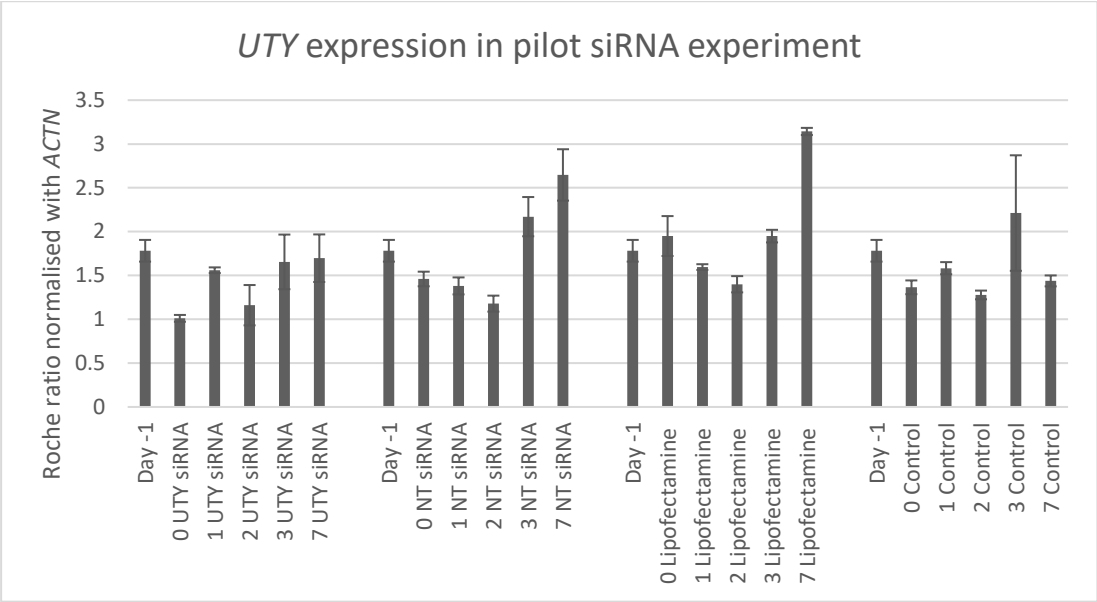

Figure 5-21 qRT-PCR results of an initial siRNA experiment for UTY. Pictured here is the UTY expression normalised to ACTN expression and error bars (calculated by the by Roche LightCycler 480 software). Error bars show standard error. Day -1 is the same for all treatments.0 means day 0, 1 is day 1 and etc. 0 UTY siRNA stands for day 0 sample with UTY siRNA, 1 NT siRNA stands for day 1 Non-targeting control siRNA, 2 Lipofectamine stands for day 2 cells with Lipofectamine only, and 3 Control stands for day 3 wildtype THP-1 differentiation.

The *UTY* siRNA pilot experiment also included a third control – Lipofectamine 2000 used on day -1, without siRNA (Lipofectamine) (Figure 5-21). Generally, *UTY* expression did not change much throughout the normal time course of THP-1 differentiation (Control), with similar values for also Lipofectamine and NT controls. The only exception was day 7 for both Lipofectamine and NT control, where expression was almost twice as high as for the Control. Other than that, Lipofectamine or Non-targeting siRNA pool did not affect the *UTY* expression during THP-1 differentiation.

UTY siRNA pool only reduced *UTY* expression to ~70% of the NT control on day 0. On day 1, *UTY* siRNA pool expression was higher than NT. *UTY* siRNA did not prevent THP-1 differentiating into macrophages, as they became adherent to the surface and exhibited a macrophage-like morphology. It can be concluded that *UTY* siRNA pool did not knock down the expression of *UTY* to a sufficient level at day 0, and not at all after start of the differentiation, so the impact of reducing *UTY* mRNA level could not be assessed in this experiment.

5.3.3.2 siRNA experiment harvesting RNA for proposed CAGE sequencing did not show a satisfactory knock down for *KDM6B* and almost none for *UTY*

The siRNA experiments were carried out again, aiming to harvest sufficient RNA for CAGE sequencing.

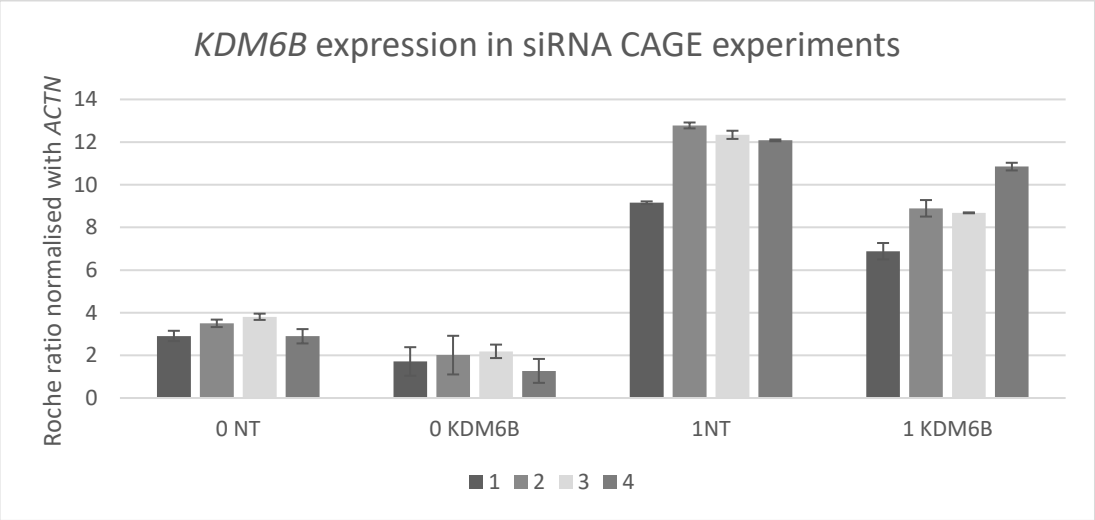

Figure 5-22 qRT-PCR results of KDM6B expression in four siRNA experiments for CAGE sequencing. Pictured here is the KDM6B expression normalised to ACTN expression and error bars (calculated by the by Roche LightCycler 480 software). Error bars show standard error. 0 NT is the Non-targeting siRNA pool at day 0, 1 KDM6B is KDM6B siRNA pool at day 1.

KDM6B siRNA pool experiment was repeated 4 times, with lipofection of siRNAs on day -1, harvesting RNA and simultaneously adding PMA at day 0, and harvesting RNA on day 1 again. Only Non-targeting control (NT) was used in this instance (Figure 5-22). Replicates seem to have similar levels of *KDM6B* expression on day 0, with more variation at day 1. This might be due to the greater biological variability of the cells after differentiation, as also seen in FANTOM4 (Suzuki *et al*., 2009).

KDM6B expression level was only reduced to around half in all four cases for day 0 (Table 5-7) and remained at more than 70% on day 1. This was deemed an unacceptable level of knock down to properly assess the effect of change in gene expression and these samples were not used for CAGE sequencing.

	% of *KDM6B* expression after knock down	
Experiment number	day 0	day 1
1	59%	75%
2	57%	70%
3	57%	70%
4	44%	90%

Table 5-7 KDM6B expression after knock down by siRNA pools compared to NT control levels (% of control level). Calculated with qRT-PCR levels normalized to ACTN by Roche LightCycler 480.

Because *UTY* knock down was not very efficient in the pilot experiment, *UTY* siRNA pools were only repeated twice in the siRNA experiments for possible CAGE sequencing. *UTY* siRNA pools did not knock down the *UTY* expression to less than 65% in any sample (Figure 5-23, Table 5-8). In the second experiment day 1, *UTY* expression with the *UTY* siRNA was even higher than the control (Figure 5-23, Table 5-8).

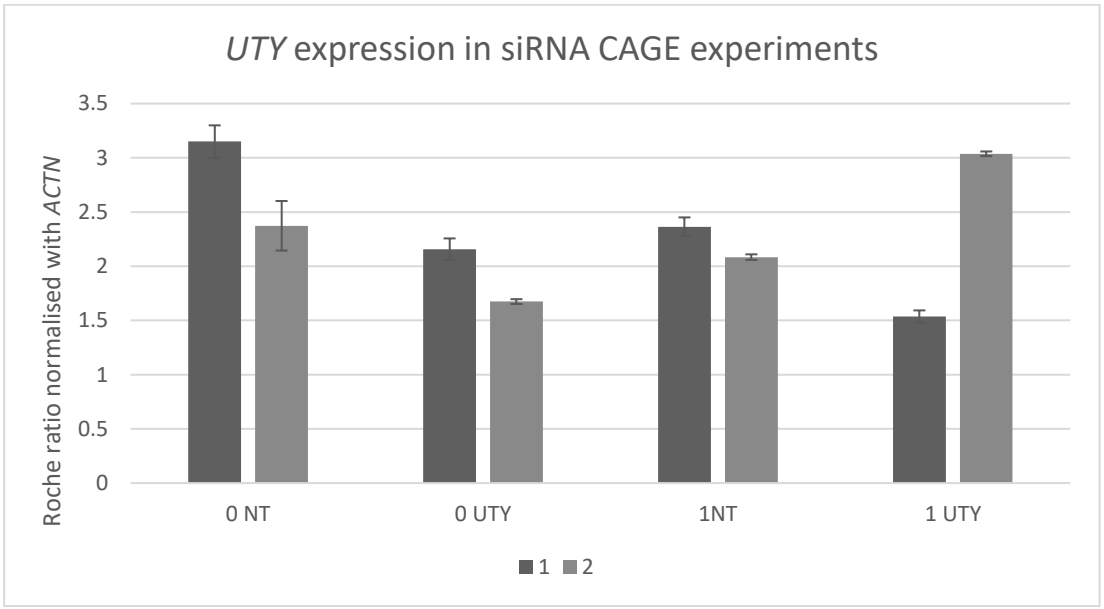

Figure 5-23 qRT-PCR results of UTY expression in four siRNA experiments for CAGE sequencing. Pictured here is the UTY expression normalised to ACTN expression and error bars (calculated by the by Roche LightCycler 480 software). Error bars show standard error. 0 NT is Non-targeting siRNA pool at day 0, 1 UTY is UTY siRNA pool at day 1.

	% of *UTY* expression after knock down	
Experiment number	day 0	day 1
1	68%	65%
2	71%	146%

Table 5-8 UTY expression after knock down by siRNA pools compared to NT control levels (% of control level). Calculated with qRT-PCR levels normalized to ACTN by Roche LightCycler 480.

Thus, it was not appropriate to continue with CAGE sequencing for any demethylase siRNA treated samples.

5.4 Discussion

5.4.1 Overview of this chapter

H3K27 demethylases perform an important catalytic function in mediating change in gene expression, whether it is cell differentiation or activation, because they remove repressive marks from histones which opens the chromatin and facilitates transcription. In this chapter, functional roles of H3K27 demethylases were more closely studied using widely available open source databases like Ensembl and FANTOM5, and also using the THP-1 monocyte to macrophages experimental system.

Looking at bioinformatics analysis of KDM6A, UTY and KDM6B showed that *KDM6A* and *UTY* genes are very similar in structure and their regulation, and thus potentially their function as well. *KDM6A* and *UTY* expression was found to be co-regulated, but this seemed to diminish in cancerous samples. Using the THP-1 experimental system, which contains a mutation deleting *KDM6A* expression, allows exploration of the function UTY and KDM6B in leukaemic monocytic cells and in the monocyte to macrophage transition. Demethylase-specific inhibitor GSK-J4 prevented THP-1 differentiation, consistent with these enzymes being important in removing repressive marks from key genes. It was hoped to study this effect by targeting each demethylase specifically. Unfortunately, treatment with RNAi was not able to knock down the expression of *KDM6B* and *UTY* sufficiently to have an impact on enzyme function.

5.4.2 Bioinformatics study of H3K27 demethylases

5.4.2.1 Structural differences between H3K27 demethylases

Analysis of the different transcripts, proteins, promoter structure and gene expression of KDM6A, KDM6B and UTY demethylases showed that KDM6B is quite different from the other two. As mentioned in section 5.1, UTY has been frequently overlooked as a catalytically inactive member of this protein family, with much more focus directed at its X chromosome paralogue, *KDM6A*. Even though the protein structure of KDM6B lacks the TPR domains found in KDM6A, there is functional catalytic

redundancy between these two proteins in some instances (Manna *et al*., 2015). However, *KDM6B* expression did not have a high positive correlation with *UTY* or *KDM6A* expression in the comprehensive FANTOM5 data (Table 5-5), suggesting a weak link between gene regulation and function of *KDM6B* and *KDM6A/UTY*.

KDM6B was shown to have quite a complex promoter structure based on expression data from FANTOM5 (Figure 5-2). There was no main TSS peak as seen in other simpler promoter structures (Figure 5-3 or Figure 5-4). Instead, there are multiple TSS for transcription from both strands, which is a sign for enhancers (Arner *et al*., 2015). The enhancer area actually comprised of one of the most expressed *KDM6B* TSS, whereas the areas where the two main protein-coding transcripts should start, had very low transcript levels. Enhancers having higher levels than the promoters they regulate is unusual, and it might be possible that this enhancer region interacts with another promoter. However, the highest TPMs of this region (*KDM6B*), reflecting gene expression, were found in stimulated immune cells (for example monocyte-derived macrophages stimulated with LPS) which is highly documented by literature (De Santa *et al*., 2009). The *KDM6B* promoter 1 annotated by FANTOM5 (p1@KDM6B) is found in the middle of the gene, where none of the known transcripts have start sites, and this position was not used as the promoter region for further analysis. Instead, the whole CpG region was used to determine expression for *KDM6B*, which could mean that some enhancer activity was included. However, Ensembl lists this whole region as promoter flanking region, with two promoters, either at the beginning of the two long protein-coding transcripts, or further away, at the beginning of the short protein-coding transcript. All in all, it seems that even though the CpG region taken for consideration does include some expression associated with enhancers, it is also identified as a promoter for the short transcript (Ensembl ID ENST00000570632.1). It is therefore possible that most of the expression for *KDM6B* detected in the FANTOM5 samples codes for the short transcript without the JmjC catalytic domain. However, for example, *KDM6B* activity in monocyte-derived macrophages stimulated with LPS depends on its demethylase activity (Kruidenier *et al*., 2012) and thus it is highly improbable that these transcripts would be without the catalytic domain. Further

investigation by looking at RNAseq data would be able to confirm or to deny this hypothesis.

UTY and *KDM6A* did not have the complex promoter structure of *KDM6B* (Figure 5-3, Figure 5-4). However, *UTY* has been reported as having a huge frequency of alternative splice forms (Laaser *et al*., 2011). They found as many as 284 new transcripts, but these findings have not been validated by other reports. Thus, only the 16 validated transcripts from Ensembl were included in this project to elucidate the role of *UTY* in THP-1 cells. As described in the results section, a few of the protein-coding transcripts do not code for UTY or KDM6A functional domains. It is not clear whether these incomplete peptides are able to carry out the functional role of these two enzymes.

5.4.2.2 Co-regulation of demethylase gene expression

KDM6A and *UTY* were analysed in detail to observe the level of similarity between these two genes, and assess the importance of *UTY* in cells due to its frequent neglect in publications about demethylases. Several lines of evidence suggested that UTY might have some function in cells. For example, while *Kdm6a*$^{-/-}$ female mice die *in utero*, a proportion of *Kdm6a*$^{-}$/Y male mice, which have a normal *Uty* gene on the Y chromosome, survive. In addition, THP-1 cells, which have a deletion of *KDM6A* which renders it non-functional, are also able to survive in culture, possibly due to the presence of a functional *UTY* gene.

Firstly, it was hypothesized that UTY might compensate for KDM6A in tissues where there is low *KDM6A* expression. In this case, there would be an inverse correlation between expression of *KDM6A* and *UTY.* Instead, these results showed that *UTY* and *KDM6A* gene expression are positively correlated (Figure 5-7), suggesting that the two genes are co-regulated and that UTY might be a mechanism to compensate in males for the single copy of *KDM6A* (which is not inactivated in females who therefore have expression from both chromosomal gene copies).

This positive correlation was shown to be disrupted in cancer-only cells (Figure 5-9). *KDM6A* is a known tumour suppressor gene, so the disruption could contribute to the malignancy of cancerous cells. For example, Ahn *et al*. (2016) reported that both *UTY* and *KDM6A* enhance proliferation of two male urothelial bladder cancer cell lines. *UTY* knockout increased cell proliferation to the same rate as *KDM6A* knockout, whereas double knockout of *KDM6A* and *UTY* increased it even more. The authors argue this is due to the loss of dosage-dependent suppression effect of KDM6A/UTY in urothelial cancer. *KDM6A*, like other developmentally important chromatin modifying genes and tumour suppressor genes, is tightly regulated in most tissues (Bellott *et al*., 2014). Unlike other genes on the chromosome X, X genes with a Y chromosome paralogue have even been mostly found to have a role in transcription, translation and nucleic acid binding (Bellott *et al*., 2014). There is a strong assumption that these genes (many of them coding for chromatin-modifying enzymes) are needed for proper gene regulation and thus are more sensitive to the altered dosage. Perturbation of this fine balance of regulation of demethylase activity might result in cancer or developmental issues.

KDM6A missense, nonsense and deletion mutations were found to cause Kabuki syndrome (KS, OMIM #147920) (Miyake *et al*., 2013), which is a rare multi-systemic disorder first reported in Japan by two research groups (Kuroki *et al*., 1981, Niikawa *et al*., 1981). Patients with KS have an unusual facial appearance, resembling the traditional make up by Japanese Kabuki artists. However, the majority of cases were found by whole-genome sequencing to have mutations in *KMT2D* (histone lysine N-methyltransferase 2D, also known as *MLL2*), not *KDM6A*. KMT2D is part of a complex involved in methylating H3K4 (and thus activating gene expression), which also includes KDM6A (Cho *et al*., 2007). So far, 33 germline mutations in KDM6A gene have been found in a comprehensive study of KS mutations (Bogershausen *et al*., 2016). No mutations have been reported in *UTY*.

Some patients with Turner syndrome (TS) were found to have KS like features (Rodriguez *et al*., 2008). TS is a 45,X monosomy and is the only human chromosomal haploinsufficiency where affected individuals survive after birth. Nevertheless, it

significantly affects foetal mortality as only 1% of 45X monosomy foetuses survive to term (Hook and Warburton, 2014). TS patients surviving to adulthood are most probably mosaic cases (Hook and Warburton, 2014) where some cells have two X chromosomes or an X and a Y chromosome.

Trolle *et al*. (2016) hypothesized that *KDM6A* is a potential candidate gene for premature ovarian failure in TS. *KDM6A* was found to be more highly expressed in 46XX karyotype than in 45X karyotype, which is consistent with the observation that *KDM6A* escapes X-inactivation (Greenfield *et al*., 1998). However, they found that *KDM6A* was expressed at the same level in 46XX females as in 46XY males. If *UTY* is also active, this should mean that males have much higher levels of demethylase activity (from *KDM6A* and *UTY*) than females with just *KDM6A*. These results are also not consistent with other findings in mouse (Xu *et al*., 2008), where higher levels of KDM6A in liver and brain in females than in males were reported. Xu *et al*. (2008) postulated that even though *Kdm6a* does escape X inactivation, its expression from the inactivated X is lower than from the active X chromosome.

Miyake *et al*. (2013) created a hypothesis for the *KDM6A* abnormality observed in KS and TS. They ascertain that a certain level of *KDM6A/UTY* gene expression is needed for proper developmental function, either two copies of *KDM6A* or one copy of *KDM6A* and one of *UTY*. This threshold level is higher than that coming from one copy of *KDM6A*, as TS patients have one copy of *KDM6A* and may manifest a KS-like phenotype (discussed above). KS syndrome female patients also can have skewed inactivation of the X chromosome for the *KDM6A* mutation (Lederer *et al*., 2012), which could bring the overall level of *KDM6A* (one X active copy) higher than in males with *KDM6A* mutation, but still less than the expression level reached with two *KDM6A* copies (one X active copy and one X inactive less expressing copy). Thus, this hypothesis needs to be investigated further by assessing absolute levels of KDM6A and UTY in the same male and female tissues.

Analysis of potential TF binding on *KDM6A, UTY* and *KDM6B* promoters either by motif predicting (Figure 5-10, Figure 5-11) or various ChIPseq data (Figure 5-12) gave

very variable results. More validation would be needed, especially by ChIPseq in the same cell type, but the present results indicate that *KDM6A* and *UTY* have a higher number of common TF than either with *KDM6B*.

According to the database Common Pathways protein-protein interactions, UTY protein interacts with KDM6A (Appendix 8.8), but with also other parts of the H3K4 methylation complex such as KMT2B, KMT2C and RBBP5, just like KDM6A. This suggests that UTY binds to the same complex as KDM6A and also performs demethylation functions. It is not clear whether UTY binds to this complex independently or only with KDM6A present, so it still might have a catalytically autonomous function and could therefore target different substrates. Further investigation into protein-protein interactions with KDM6A should prove interesting.

Although UTY may serve a separate male specific function, for example in testes, it appears to be expressed globally (Appendix 8.6). The findings presented in this section suggest that in general *KDM6A* and *UTY* have shared regulation, which may allow for survival of males by compensating for the haploinsufficiency of *KDM6A*. In contrast *KDM6B* appears to have very different regulation and showed little redundancy with *KDM6A* and *UTY*, indicating that this protein likely has an independent role.

5.4.2.3 Evolutionary consequences of KDM6A position on X chromosome

Despite recent advances in high throughput sequencing of genomes of various organisms, Y chromosome sequence has been frequently overlooked, which makes it more difficult to examine *UTY* sequences in other organisms and its evolutionary history. Most genome releases in Ensembl did not include Y chromosomes, probably due to its small size, its presumed low gene content and a huge number of repetitive sequences which hinder scaffold assemblies (Li *et al*., 2013a). However, from the limited reports of Y chromosome sequencing it appears that *UTY* is one of the few Y chromosome genes that is present as a functional gene in all species studied so far (Li *et al*., 2013a).

The original placental and marsupial (therian) Y chromosome, containing the sex-determining gene *SRY*, emerged approximately 180 million years ago (Cortez *et al*., 2014). *UTY* is located in a Y chromosome region that stopped recombining approximately 100 million years ago (Lahn and Page, 1999), together with other Y-chromosome genes such as *ZFY, USP9Y, DDX3Y* and *TMSB4Y* (Cortez *et al*., 2014), just as *KDM6A* emerged on the X chromosome. As mentioned in the results section, *UTY* has always clustered together in a male-specific Y chromosome region (MSY) with two other Y-linked genes, *USP9Y* and *DDX3Y*, throughout eutherian Y chromosome evolution despite rampant rearrangements of Y chromosome (Li *et al*., 2013a). The high microsyntenic conservation of this cluster of genes *UTY+USP9Y+DDX3Y* suggests that they might be co-regulated. Li *et al* (2013) speculate that they could be regulated by the 'testis-specific' non-coding element *TTTY15*. However, expression was not correlated with either of these genes in the extensive FANTOM5 results (Figure 5-14, Figure 5-15) so this hypothesis was not confirmed.

Gerrard and Filatov (2005) reported findings of positive directional selection on human genes *UTY* and *USP9Y*, unlike their X chromosome paralogues which are under a strong purifying (negative) selection. This means that any *KDM6A* mutation negatively affects the cell, but *UTY* mutations may have been encouraged in evolution and they might be beneficial with a male-specific function. The question arises whether the loss of the larger part of the sequence encoding the demethylase / hydroxylase function of UTY was driven by positive selection or was an unfortunate consequence of evolutionary processes acting on the Y chromosome. Nevertheless, it seems that, despite the tight regulation of *KDM6A* and *UTY* gene expression, *UTY* with lower demethylase activity seems sufficient to rescue the embryonically lethal *Kdm6a*$^{-/-}$ genotype.

5.4.3 Inhibition of KDM6A, KDM6B and UTY

Initially, the effect of perturbing demethylase function on THP-1 differentiation was assessed using an H3K27 demethylase specific inhibitor. GSK-J4 is a cell permeable H3K27 demethylases inhibitor, which binds to the catalytic centre of JmjC domain

(Kruidenier *et al*., 2012). It was originally thought to bind only to KDM6A and KDM6B, but later was proved to bind weakly to other demethylases such as KDM5B and KDM5C (Heinemann *et al*., 2014). Neither of these papers show whether GSK-J4 has an effect on UTY demethylase activity. Kruidenier *et al*. (2012) also showed that GSK-J4 successfully inhibits LPS stimulation of human monocyte-derived macrophages, and decreases typical LPS response-related gene expression. Control for this experiment was GSK-J5, a cell permeable compound similar to GSK-J4, but without any inhibitor activity (Kruidenier *et al*., 2012). The half-maximum inhibitory concentration (IC_{50}) of GSK-J4 effect on KDM6B in unnamed cell based assays was 3.1 μM (Heinemann *et al*., 2014), whereas IC_{50} for GSK-J4 in TNFα blockade was 9 μM (Kruidenier *et al*., 2012).

The MTT results (section 5.3.2) show that viability/metabolic activity or proliferation of THP-1 decreased to about 50% at 10 μM GSK-J4 inhibitor. Other research projects used GSK-J4 at much lower dose in THP-1 cell line (1 and 2 μM) (Ha *et al*., 2017, Ntziachristos *et al*., 2014). Together with the MTT results, it seems that the reason why cells do not differentiate and adhere to the bottom of the well is due to the lethality or effect on metabolic activity of the inhibitor. Because GSK-J5 decreases cells' viability/metabolic activity at a lower rate than GSK-J4, it can be concluded that demethylase activity of KDM6 proteins is important for the cells' survival or metabolic activity. GSK-J4 only binds to the catalytic centre of the JmjC fold, so specific protein-protein interactions of KDM6B and UTY might be still active.

The adherence assay also had high variability of the results, suggesting that THP-1 cell populations were not completely uniform in response to the inhibitor, something else is affecting the differentiation or the methods of quantifying the results need to be improved. Counting cells with haemocytometer is not time effective nor precise. Establishing clear boundary between THP-1 monocytes and macrophages using macrophage markers such as CD11b for flow cytometry would be helpful. Unfortunately, an initial attempt showed that this approach needed to be optimised extensively and it was not possible to achieve this due to time constraints. Another

way would be to take the two populations of adherent and non-adherent cells and investigate gene expression levels of monocyte and macrophage markers.

MTT results suggest that the non-adherent cells present at >10 μM concentration levels are dead/metabolically inactive cells. However, due to using Trypan blue during counting, only living cells were included. It might be suggested that high concentrations of GSK-J4 only slow down the differentiation process, since, when these non-adherent cells were replated onto a fresh plate, they did appear to differentiate after few days (only repeated once, not shown). GSK-J4 might therefore slow down or pause the differentiation process, but it is not known how long GSK-J4 stays active in the cell. The differentiation adherence assay was also only performed at 24 and 48 hours post-differentiation, making it possible to miss the window at which differentiation of THP-1 cells affected by GSK-J4 is slowed down. As only a fraction of cells at 10 and 20 μM were adherent after 24 and 48 hours, it seems that perhaps the cells are not uniform in their response to the inhibitor. It might be useful to use lower passage number of THP-1 cells to assess whether passaging and time in culture affects this response. These experiments with inhibitor against KDM6B in THP-1 cells raised interesting questions about THP-1 uniformity. Further research could include gene expression analysis and assessing global H3K27me3 levels in THP-1 following the use of the inhibitor.

Since the results with the demethylase inhibitor suggested that KDM6B and possibly UTY were involved in the transition from monocyte to macrophage, other approaches to perturb the level of these proteins in THP-1 cells were then explored.

5.4.4 siRNA silencing efficiency for KDM6B and UTY in THP-1 cells

Next, silencing of the genes was attempted using RNAi in THP-1 cells. The pools of four different siRNA provide complete silencing of the gene according to the manufacturer (GE Dharmacon website, 2016). The mode of delivery chosen was lipofection by Lipofectamine 2000, following the procedure in the FANTOM4 project (Suzuki *et al*., 2009). THP-1 monocyte cells were hard to transfect with plasmids (discussed further in Chapter 6), it was easier to introduce siRNAs, as described in

previous chapter (Chapter 4). It might have been possible to achieve higher level of transfection using siRNA-specialised Lipofectamine reagents (for example Lipofectamine RNAiMAX, Thermo Fisher Scientific), but as shown in Chapter 4, where the level of *BAP1* expression was knocked down to less than 20% of the control expression, the problem of low efficiency of *KDM6B* and *UTY* knock down was not in the lipofection itself.

KDM6B expression was knocked down by siRNA to around 50% at day 0, which was not as low as required to explore the effects of the gene/protein (Chapter 5.3.3.2). At day 1 the knock down is even less at only around 30% reduction in mRNA level. As *KDM6B* expression levels before differentiation are very low and dramatically rise during first hours of differentiation (as seen in Chapter 3), it might be possible that the amount of siRNA present in the cell did not manage to degrade all the *KDM6B* transcripts at the peak of expression. However, the final concentration of siRNA used was higher than the minimum amount suggested by manufacturer (50 nM vs 20 nM) for all siRNAs.

It seems that lipofection itself increases the *KDM6B* levels, as the cells are in stress (Figure 5-20). All the comparisons between siRNA and control were therefore made using Non-targeting siRNA control (NT). This might explain why *KDM6B* knock down did not have an effect on THP-1 differentiation, as the absolute levels of *KDM6B* transcript in cells with KDM6B siRNA and Lipofectamine might have been around the same as in control THP-1 cells without Lipofectamine.

The *UTY* siRNA pool did not downregulate the amount of *UTY* transcripts in THP-1 cells at any point. The best knock down out of all repeats was around 30% reduction in mRNA level, and at one point the *UTY* mRNA level increased above the control level with the addition of *UTY* siRNA. It is unclear why the siRNA pool did not decrease *UTY* levels. The primer set used to quantify the *UTY* levels by qRT-PCR was designed specifically to detect only *UTY* transcripts (exon 16). This primer set was used in preliminary absolute expression qRT-PCR experiment comparing U937 cell line (with no *UTY* and with *KDM6A* expression) with THP-1, where the expression

was only found in THP-1 cells (not shown). The most similar transcript to *UTY* is *KDM6A*, which is not expressed at all in THP-1 in FANTOM5 data and very small amount in FANTOM4. Looking at THP-1 clone 5 genome using microarray-based comparative genomic hybridization, there is a deletion in X chromosome which also includes *KDM6A* (Adati *et al*., 2009). van Haaften *et al*. (2009) states that the deletion occurs precisely at exons 1 to 16 in *KDM6A* in THP-1 cells. It can therefore be concluded that this *UTY* primer set used in the qRT-PCR cannot pick up *KDM6A* contamination in THP-1 cells.

To ensure higher knock down of *UTY* and *KDM6B* using siRNA different sets of siRNAs for *KDM6B* and *UTY* could be designed and tested. Transcript levels could also be assessed by various sets of primers, and the protein levels should be checked by western blotting.

5.4.5 Conclusions and future directions

This chapter has shown that UTY, despite its low demethylase activity, should share some of the activities of KDM6A. This functionality is worth further investigation, either by assessing the levels of *KDM6A* and *UTY* in the same male and *KDM6A* in female tissues to see whether *KDM6A* and *UTY* expression together in males equates to *KDM6A* expression in females. The present study failed to knock down *UTY* using siRNA in THP-1 cells, but other options include using primary cells from male and female donors to see whether *UTY* has an effect on cell survival. A further possibility is to use other leukaemia cell lines and primary cells with or without *UTY* and *KDM6A* genes to elucidate the UTY function by adding exogenous WT or mutated KDM6A or UTY proteins or overexpression the genes. *KDM6B* gene expression was shown to rapidly increase during early monocyte to macrophage differentiation (Chapter 3). Thus, it is reasonable to assume that it is important in changing gene expression. It would be interesting to find out whether cells would change gene expression if more KDM6B protein was added or the gene was overexpressed. The use of CRISPR-Cas9 technology to knockout the demethylases' genes is explored in Chapter 6.

In summary, studying demethylases KDM6A, KDM6B and tentative demethylase UTY brings us more clues about how chromatin modifying enzymes orchestrate the change in gene expression and how their dosage is important in development and cancer.

Chapter 6: Genome editing of chromatin modifying enzymes using CRISPR-Cas9

6.1 Introduction

As discussed in Chapters 4 and 5, perturbing the expression of the chromatin modifying genes using inhibitors and siRNA showed variable results. Knocking down a gene using siRNA never brought its expression down to zero per cent, with protein still being made. Even though the dosage of chromatin modifying enzymes is extremely important in healthy cells (as discussed in Chapter 5), transient lower levels of these enzymes in leukaemic cell lines such as THP-1 might still permit their immediate survival. Therefore, attempts were made to generate stable mutations in these chromatin modifying genes using genome editing technology.

CRISPR (clustered regularly interspaced short palindromic repeats) - Cas9 is a new genome editing technology which utilises a bacterial system of recognition of sequences against viruses, such as bacteriophages and other genetic mobile elements (as reviewed by Doudna and Charpentier (2014)). This method is extremely efficient, versatile and easy to design, as to create a new target, only a 20 nucleotide guide sequence, which leads the endonuclease Cas9 to the desired place, needs to be provided (Ran *et al*., 2013). Cas9 creates a double strand break, which can be repaired by non-homologous end joining (NHEJ), if no template is present, creating an insertion or a deletion (indel) (Ran *et al*., 2013). In the present project, the 20 nt guide sequence has been inserted into a plasmid containing Cas9 nuclease and GFP genes, so that successfully transfected cells express GFP and are easily recognisable (Ran *et al*., 2013).

High levels of Cas9 activity or guide sequences which are highly complementary to multiple places in the genome might lead to off-target effects. Most researchers have found that CRISPR-Cas9 is an extremely precise technology, with minimum off target effects (Iyer *et al*., 2015), and the ones reporting extortionate high numbers of off target effects are currently being investigated (Schaefer *et al*., 2017).

At the start of this project, no genome editing employing CRISPR-Cas9 had been used in THP-1 cells. Since then, only two reports have been published (Goetze *et al*., 2017, Schmidt *et al*., 2016), with THP-1 cells declared to be difficult to transfect with CRISPR-Cas9 plasmid. Nevertheless, generating a stable knockout THP-1 cell line with multiple passage numbers in culture and thus diminishing the probability of functional protein being present in the cell after multiple cell divisions might give more information about the functional role of these enzymes in leukaemia.

6.1.1 Aims of this chapter

In this chapter attempts to edit the five chromatin modifying genes in THP-1 cells are described. As THP-1 cells have a deletion of the first exons (1-16) of *KDM6A* (van Haaften *et al*., 2009), these represent a natural knockout of this gene with no *KDM6A* expression found in FANTOM5 data (Chapter 3 and 5), and further manipulation was not necessary.

Initially bioinformatics analysis using Ensembl and FANTOM5 databases was done in Chapters 4 and 5 to decipher which transcripts and transcription start sites are used in *BAP1, USP12 USP16, KDM6B* and *UTY*. This knowledge was used to design CRISPR-Cas9 guides to efficiently knockout these five enzymes. The *USP16* gene was successfully altered by insertion of an adenine nucleotide causing a frameshift mutation. *USP16* knockout THP-1 clone proliferation, cell cycle progression and macrophage phagocytosis were assessed. The effect of *USP16* deficiency on THP-1 differentiation and its transcriptome was further investigated using whole genome CAGE sequencing.

6.2 Materials and methods

6.2.1 CRISPR-Cas9 editing of target genes

CRISPR sequence guides were designed using an online website tool (http://crispr.mit.edu), where the desired sequence to be edited was queried, and then good quality guides were selected based upon their base chemistry and possible off-target effects (Table 6-1). Information about protein and gene structures was taken from the Ensembl database: https://www.ensembl.org (Yates *et al*., 2016). Information was correct as of September 2017. Translation of altered *USP16* DNA exonic sequence into protein was provided by the online ExPaSy tool available at http://web.expasy.org/translate/. Transcription Start Site (TSS) information was taken from the FANTOM 5 database (Forrest *et al*., 2014) visualised on the Zenbu hg19 genome viewer (http://fantom.gsc.riken.jp/zenbu/gLyphs/) (Severin *et al*., 2014). The oligonucleotides were ordered through Sigma-Aldrich (0.025 µmol, DST purification) based on the protocol from Ran *et al*. (2013).

The guide sequences were initially phosphorylated and annealed together by adding 1 µl of 100 µM TOP sgRNA, and 1 µl of 100 µM BOTTOM sqRNA together with 1 µl of T4 PNK ligase 10X Buffer (NEB), T4 PNK (10 U/ µl, NEB) and 6 µl of water (total 10 µl per guide). The mixture was incubated at 37°C for 30 minutes, 95°C for 5 minutes and then cooled down to 25°C at 5°C per minute. The phosphorylated and annealed oligonucleotides were diluted 1:200 by adding 1 µl of annealed and phosphorylated oligonucleotide mix to 199 µl water. The sgRNA oligonucleotides were then cloned into a plasmid vector containing a Cas9 open reading frame sequence (pX458 from Addgene, plasmid#48138, kindly provided by Dr Peter Hohenstein, The Roslin Institute) in one step. Cutting and ligation was done by adding together 100 ng of pX458, 2 µl of the diluted annealed oligonucleotide, 2 µl of 10X T4 ligase buffer with 10mM ATP (NEB), 1 µl of *BbsI* restriction endonuclease (10 U/ µl , NEB), 0.5 µl of ligase from Quick Ligation kit (NEB) made up to 20 µl water. The mixture was incubated for 6 cycles of 37°C for 5 min and 21°C for 5 min.

Next, the cloned plasmids with sgRNA sequences were treated by Plasmid Safe Exonuclease kit (Epicentre, Madison, WI, USA) which digests any residual linear

DNA. 11 µl of ligation reaction was mixed with 1.5 µl of Plasmid safe buffer 10X, 1.5 µl of 10 mM ATP and 1 µl of Plasmid Safe Exonuclease (10 U/ µl), and incubated at 37°C for 30 min, then 70°C for 30 min.

Afterwards, DH5α strain *E. coli* bacteria were transformed with these plasmids as follows. 20-30 µl of DH5α bacteria per reaction were thawed on ice, mixed with 2 µl of plasmid and left on ice for 20-40 min. The bacteria were then heat-shocked by incubating in a water bath for 30-60 s at 42°C and left on ice for 2 min. Next, 100 µl of SOB (Super Optimal Broth, provided by the Roslin Institute core services) was added and the bacteria were streaked on an ampicillin plate (100 µg/ml ampicillin in LB (Lysogeny Broth) agar, provided by the Roslin Institute core services). After leaving the plate overnight at 37°C, there were around 50-100 colonies on the plate, no colonies on a no plasmid control and no colonies for a no oligonucleotide control.

Two colonies per plate were picked into 5 ml LB (provided by the Roslin Institute core services) + 100 µg/ml ampicillin and left overnight in a shaker at 37°C and 200 rpm. The plasmid DNA was extracted using the Qiagen MiniPrep kit (Qiagen) according to manufacturer's instructions. The right DNA sequence was validated by sequencing using U6 FWD primer at Edinburgh Genomics (see 6.2.3.2).

Once the sequences were validated, 5ml of LB with ampicillin was inoculated with the cell suspension from the previous step, and left in the shaker for 6-8 hours at 37°C and 200 rpm. In the evening, 500 µl of the growing colony was put into 250 ml of LB with ampicillin and left incubating overnight at 37°C and 200 rpm. In the morning, the bacteria were spun down at 3200 x *g*, 30 min at 4°C. The plasmid DNA was extracted using the Endo-free Maxi Prep (Qiagen) according to manufacturer's instructions. It was important to use endotoxin-free (Endo-free) reagents to avoid activating the cells during the transfection step. The final plasmid was resuspended in 500 µl of Endo-free TE. All maxi prep-prepared plasmids were verified by sequencing at Edinburgh Genomics.

Gene	Location	Name	sgRNA sequence	Score	Off-target sites	- in genes
BAP1	Before exon 4	g1	TTTGCACTGCGTCATCACTC	84	79	6
		g2	TGAGTGATGACGCAGTGCAA	83	114	15
	After exon 4	g1	AGTTCAGTTCGTTCTGCCAG	76	124	12
		g2	TCAGATATCCAGTGGGTATT	72	199	8
USP12	Exon 2	g1	CCGGTCAATGAGCACTATTT	78	60	5
	Exon 3	g2	ATCTTGTGATGAACTTCTTA	64	235	24
	Exon 4	g3	CAAAATGGTCGTTTACCTAA	73	144	14
	Before exon 3	g1	AGAGAGAAATTCCCGGTATA	85	187	11
	After exon 3	g1	AAGAGGCACTCGTTTAGTGA	84	111	14
USP16	Exon 5	g1	TGGCGTCAGATAGTGCTTCA	79	89	7
		g2	AAGATCTGAACCTCACTGTC	70	188	15
KDM6B	Exon 11	g1	CCATTACCAAACTCCCGCGC	96	24	0
	Before exon 4	g1	CGAGCTGGCTGGATGTACAC	84	79	8
	After exon 4	g1	GTCTGTTATATACGGCAAAA	85	114	10
	Before exon 17	g1	CGGTGCGGAAGTGCGACGAG	95	42	14
		g2	GCTGCTAGGACCTCGTGGCG	90	108	24
	After exon 17	g1	CGGGGATCGCAGTTCCGACC	95	29	8
		g2	CTTCCAGCATCAGCGCGCAC	92	93	24
UTY	Exon 1	g1	GGTAGTGAGCGACACTGCGC	92	46	11
		g2	GTCTGTTAGCCTGACAGTCG	87	77	9
	Exon 2	g1	GCCTAGTAGGGTCTTCGTTC	92	40	0
		g2	GCCTCACGAACCCGAAGAGA	88	88	15
	Exon 3	g1	AGAGGTTGAAGTGACCTAAT	73	184	10
	Exon 24	g1	ACTTCCTGCTTTTGCGCGTG	90	103	13
	Before exon 6	g1	TATATCTCACTTGATAGTAT	64	278	10
	After exon 6	g1	TTAATGTCACTGAGAAGCGT	77	173	19
		g2	AAATTGAGGATATGACCTAA	67	287	14

Table 6-1 CRISPR-Cas9 sgRNAs designed and scored using http://crispr.mit.edu page from 0-100 (100 most likely to give the appropriate result) based on the chemical structure of the sgRNA sequence and the number of potential off-target sequences in the human genome.

6.2.2 THP-1 nucleofection of CRISPR-Cas9 plasmids

The best rate for plasmid transfection efficiency in THP-1 was obtained using the 4D Nucleofector kit (Lonza, Germany), with a following optimised Lonza protocol for THP-1 cells. $1x10^6$ cells per sample were centrifuged at 400 x *g* for 5 minutes, the pellet resuspended in 100 µl of SG 4D Nucleofector solution with added supplement (SG Cell line 4D Nucleofector solution X kit, Lonza). Then, 0.5 µg of Endotoxin-free plasmid (section 6.2.1, 0.5 µg of DNA total) was mixed in per sample. The cell suspension was transferred to the Nucleocuvette vessels and the programme FF-100 was executed on the 4D Nucleofector. Afterwards, 500 µl of pre-warmed THP-1 media was added using the supplied pipettes and the cell suspension was transferred to a 12-well plate with 1 ml of THP-1 media already in each well. Next day (~24 hours after nucleofection), the cells were spun down at 400 x *g* for 5 min at room temperature, and the pellet resuspended in 300 µl of 10% FBS in PBS and passed down through the filter tube for FACS (flow cytometry assisted cell sorting). Single GFP+ cells were sorted into 96-well plates of 200 µl THP-1 media using BD FACS Aria IIIu (service provided by Flow Cytometry unit in the Roslin Institute).

6.2.3 Validation of knockout cell lines

The single cell clones were left to grow for multiple weeks in 96 well plate until there were enough of them to passage into a bigger vessel. These potential knockout cell lines were tested for an insertion or deletion by DNA extraction (section 6.2.3.1), sequencing (section 6.2.3.2), qRT-PCR (section 2.2.4) and western blotting (section 6.2.3.3).

6.2.3.1 DNA extraction

The THP-1 cells were spun down, and the pellet was resuspended in 500 µl of DNA buffer (100 mM Tris-HCl pH 8.5, 0.2% SDS, 200 mM NaCl and 5 mM EDTA) and 10 µl of Proteinase K (20mg/ml, Qiagen). The samples were incubated at 56-60°C for 1-2 hours. Then, phenol-chloroform extraction was performed to isolate the DNA from protein and RNA. 500 µl of Phenol-Chloroform-Isoamyl alcohol (25-24-1, Sigma-Aldrich) was added to the sample (1 volume), then the layers were thoroughly mixed

by vortexing. The mixture was centrifuged at 16,000 g for 5 min at room temperature. The upper layer containing DNA was transferred to a new 1.5 ml Eppendorf tube. After that, 500 µl of chloroform (1 volume) was added and thoroughly mixed again. Samples were centrifuged at 16,000 x *g* for 5 min at room temperature. The upper layer containing DNA was again transferred to a new 1.5 ml Eppendorf tube. The DNA was precipitated by adding 50 µl of sodium acetate 3 M (1/10 volume) and 1 ml of 100% ethanol (2 volumes) and left at -20°C overnight or for at least an hour. Afterwards, the sample was spun at 16,000 x *g* for 5-10 min at room temperature to form a pellet. This pellet was washed using 70% ethanol and air-dried. DNA was resuspended using a suitable volume of TE or water.

6.2.3.2 PCR and sequencing

The polymerase chain reaction used High Fidelity Q5 Polymerase (NEB) according to manufacturer's instructions. For a 25 µl reaction, 5 µl of 5X Q5 reaction buffer, 0.5 µl of 10 mM dNTPs, 1.25 µl of 10 µM Forward primer, 1.25 µl of 10 µM Reverse Primer, 1 µl of 100 ng/µl DNA, 0.25 µl of Q5 Polymerase and 15.75 µl of water were mixed together. For primer sequences see Table 2-1 together with estimated melting temperature (T_m) for each primer pair, which was calculated using NEB Tm Calculator (http://tmcalculator.neb.com). The samples were initially denatured at 98°C for 30 s, then incubated for 35 cycles of 98°C for 10 s, T_m (normally 65-68°C) for 30 s and 72°C for 30 s. The final extension was at 72°C for 2 minutes.

The PCR product was run on an agarose gel (1.5% of Agarose Ultrapure (Invitrogen) in 1X TAE (provided by the Roslin Institute core services) with 1X of Sybr Safe DNA stain (Invitrogen) and results viewed using a transilluminator. 20 µl of the PCR product was purified using Charge Switch PCR clean-up kit (Invitrogen) according to manufacturer's instructions.

For sequencing, 3 µl of water, 2 µl of the purified PCR product and 1 µl of 3.2 µM primer was mixed and sent for chain termination (Sanger) sequencing at Edinburgh Genomics. The results were viewed using FinchTV programme (Geospiza, Perkin Elmer, Waltham, MA, USA).

Gene	Oligo name	Sequence	Q5 T_m
BAP1	BAP1_ex4_site_F	AGGCCACTTCAGACACAGTG	68
	BAP1_ex4_site_R	AGTCACCCATACACAGCACC	
USP12	USP12_ex2g1site_F	AAAGACGCACCTGTGTAGGG	68
	USP12_ex2g1site_R	AGGCAGACGTTGCAGTGAG	
	USP12_ex3g2site_F	GCAGTTTGGGAATACCTGCT	62
	USP12_ex3g2site_R	AAAAGCCAAGCAAGCAAAAA	
	USP12_ex4g3site_F	GGTTTTGCCCTTTTCTCAGG	65
	USP12_ex4g3site_R	TTCATCTTCCATTCTGTCTTTCTCA	
USP16	USP16 crispr site F	CCTAGCGAGTGCATGGTTTT	65
	USP16 crispr site R	ACCCAAGAGGCAGAGGAACT	
KDM6B	KDM6B_ex11_F2	TGGAGCTTGTCTTGAGGCAG	64
	KDM6B_ex11_R2	GGGTCCCTCAGTCCCAAAGA	
	KDM6B_ex4_site_F	GGGTAGCGGGCACTCTTATC	69
	KDM6B_ex4_site_R	CCTTACCTCCCACTCCCAGA	
	KDM6B_ex17_site_F	GTTCCTGCTTCCTTCCCCTC	68
	KDM6B_ex17_site_R	AGAAAGCGCTGATGGTCTCC	
UTY	UTY_ex1+2_F	TGGTGCTGGCAAAGTTTGTG	68
	UTY_ex1+2_R	GGGCCTTGGTAGAGCAAAGT	
	UTY_ex3_F	TCCTCTCTGAAATGCCGAATT	65
	UTY_ex3_R	ACAAGGGCAAGCAAAATAGC	
	UTYbefex6_F	GTGAGCTGTGATTGTCCTGC	68
	UTYbefex6_R	TCGACAAAAGCTGGGGTCAA	
	UTYaftex6_F	TTGACCCCAGCTTTTGTCGA	68
	UTYaftex6_R	GGCAGTCCAAAAGCCTTAGA	
	UTY_bothex6_F	CCTGCTGTTCTTCCTTGAATGT	65
	UTY_bothex6_R	TCTCTCTTTGGTGGCAGACA	
	UTY_ex24_F	AAGAAAACCAAAGTACCCTCTCT	64
	UTY_ex24_R2	TGCTCTCTATGCCTGCTCC	

Table 6-2 Primers used to validate the CRISPR-Cas9 targetting, together with the annealing temperature (Q5 Tm) used for genotyping with Q5 High Fidelity Polymerase (NEB).

6.2.3.3 Western blot

$2x10^6$ cells per sample were resuspended in 50 µl PBS (provided by the Roslin Institute core services), then 50 µl of 2X Laemmli loading buffer with 50 mM DTT (BioRAD, Hercules, CA, USA, prepared by mixing 950 µl Laemmli loading buffer + 50 µl 1M

DTT (NEB) in water), was added and mixed thoroughly. Samples were then incubated at 95°C for 5 minutes and stored at -20°C until needed.

Samples stored in buffer as described above were run on a precast Mini-Protean TGX 4-15% 12-well gel (BioRAD), in a tank with running buffer (25 mM Tris (Fisher Scientific), 192 mM Glycine (Sigma-Aldrich) and 0.1% w/v SDS (Fisher Scientific)) at 100V for 5 min and then 120V till the end of the gel. The gel was then rinsed in water.

To blot the protein onto PVDF membrane (Immobilon-P, Sigma-Aldrich), the membrane and the gel were activated by soaking in 100% methanol. The western blotting apparatus was then assembled as follows: black side down, sponge, 2 pieces of Whatman paper (GE Healthcare, Little Chalfont, UK), gel, PVDF membrane, 2 pieces of Whatman paper, sponge and red/white side up (BioRAD). The apparatus was closed, making sure there were no bubbles between the gel and membrane and that it was completely soaked in the transfer buffer (25 mM Tris, 192 mM Glycine, 20% (v/v) Methanol). The blotting was run at 50 V for 1 hour at initial current of 400 mA.

Afterwards, the membrane was blocked for 1 hour in 5% milk powder (Marvel Dried Milk, Premier Foods Group Ltd, London, UK) in PBS-T (0.05% Tween in PBS, PBS was provided by the Roslin Institute core services). Then, primary anti-human antibodies (dilution of 1:1000 for rabbit anti-USP16 (ab121650, Abcam, Cambridge, UK); and 1:2000 dilution for mouse anti-ß-actin (C4) monoclonal IgG1 (sc-47778, Santa Cruz Biotech, Dallas, TX, USA)) were diluted in 5% milk powder in PBS-T and the membrane immersed rotating overnight at room temperature. The next day, the membrane was washed six times for 5 min in PBS-T, and then secondary antibodies were diluted at 1:2000 in 5% milk powder in PBS-T (horse anti-mouse HRP-linked for β-actin and goat anti-rabbit HPR-linked for USP16; both from Cell Signalling Technology, Danvers, MA, USA), and the membrane immersed rotating for 1 hour at room temperature. The membrane was again washed six times for 5 min in PBS-T, and then Pierce ECL western blotting substrate (Thermo Scientific) was applied

according to manufacturer's instructions. The image was developed onto Amersham Hyperfilm (GE Healthcare).

6.2.4 Phagocytosis assay

A phagocytosis assay was performed by incubating THP-1 macrophages, that had been differentiated with PMA for two days, with Zymosan A particles (Life Technologies, Z2841) coated with fluorescein isothiocyanate (FITC), at a ratio of 100 particles per cell; for 1 hour at 37°C. Cells were washed five times with cold PBS and fixed with 4% paraformaldehyde (PFA, WVR, Radnor, PA, USA) for 10 min at room temperature and washed twice with PBS. Images were viewed using a fluorescent microscope (Zeiss Vert.A.1, Carl Zeiss Limited, Cambridge, UK).

6.2.5 MTT assay and cell cycle analysis

MTT assays were performed as in section 2.4.1 to assess cell viability, and cell cycle analysis was conducted as in section 2.4.2 to determine the stages of the cell cycle.

6.2.6 CAGE library preparation and bioinformatics analysis

Libraries were prepared and sequenced as specified in section 2.2.6, together with processing the sequencing reads, and mapping to hg38 genome (section 2.5.1). Results for 38 samples (samples for this chapter and Chapter 4) were uploaded into the CAGEr programme (Haberle *et al*., 2015) (Appendix 8.3). The CAGE samples were normalised together with CAGEr using the power law with alpha of 1.19, with results being in Tags per million (TPM) (section 2.5.2). Samples' data (Table 6-3) was analysed (sections 2.5.4, 2.5.5, 2.5.6, 2.5.7) using Miru software and EdgeR package.

Sample	Time point	Number of biological replicates
Wildtype (from Chapter 3)	0 hours	6
	24 hours	4
USP16 homozygote knockout clones	0 hours	4
	24 hours	4
USP16 heterozygote knockout clones	0 hours	2
	24 hours	2

Table 6-3 CAGE samples used in the analysis in this chapter. All samples had two time points, 0 and 24 hours.

6.3 Results

6.3.1 CRISPR guide design for *BAP1, USP12, USP16, KDM6B* and *UTY*

Two guides were designed to make an indel in exon 5 of *USP16*, creating an out of frame mutation. Exon 5 (coding exon 3) is at the start of the transcript and codes for part of the zinc finger domain. Insertion or deletion of one or two nucleotides would create an out of frame mutation, which creates a wrongly translated amino acid sequence, with possible multiple stop codons along the way. This shortened peptide would not have a catalytic deubiquitinase domain, and thus would be unable to perform the function of USP16. Transcripts with multiple stop codons are sometimes degraded by the cell's nonsense-mediated mRNA decay mechanism and thus the production of *USP16* mRNA and resulting peptide would be reduced.

BAP1 CRISPR guides were designed to work in pairs to create two double strand breaks in the DNA on the opposite ends of an exon and therefore delete the whole exon. Deletion of a large segment of the gene makes it easier to screen the THP-1 clones by a simple PCR rather than having to sequence the product. Generally, deletion of a section of the genome, rather than creating a single base indel, would be preferred, as it is easier to screen the THP-1 clones by a simple PCR rather than more expensive Sanger sequencing. Exon 4 was selected, which is in the beginning of the gene and is 133 nucleotides long, so its deletion would create a shift in the reading frame downstream and introduce stop codons.

USP12 CRISPR guides were designed to both make an indel in the *USP12* coding sequence and also delete whole exons. The first few exons did not have multiple guides available with good scores (not pictured), so only one guide was designed for exon 2, exon 3 and exon 4 each to generate an indel which would create a frameshift mutation. Later, when a deletion of whole exon strategy was deemed more suitable, guides to delete exon 3 were designed. Exon 3 is 214 nucleotides long and its deletion would also shift the reading frame in subsequent exons.

CRISPR-Cas9 target sites were designed to make an indel inside the open reading frame of *KDM6B* which according to the design should result in a frameshift mutation

(single guide) or deletion of a whole exon from the locus by using two sgRNA guides in combination together. It was decided to create an indel in exon 11 (the longest exon), and to delete exons 4 (containing the start of the coding sequence) and exon 17 (which encodes the start of the JmjC domain coding sequence). Exons 4 and 11 start and end with a different reading frame, thus the deletion would also create a frameshift mutation with stop codons along the way. Exon 4 is also present in all protein-coding transcripts (Figure 6-1), so knocking out this exon should abolish all functional translation of *KDM6B* by disrupting the reading frame of all transcripts. To explore possible non-catalytic activity of KDM6B, guides which should result in deletion of exon 17 were designed. This mutation would produce a truncated KDM6B without the JmjC catalytic domain, if the transcript was not degraded. In either case the strategy chosen would allow assessment of the importance of the underlying demethylase activity in macrophage differentiation.

UTY CRISPR-Cas9 guides were designed similarly to *KDM6B*, to produce frameshift mutations using a single guide (for exons 3 and 24) and deletion of exons using two guides (exons 1+2 and exon 6). For indel-inducing single guides relying on non-homologous end joining to produce an imperfect repair of a break in DNA, exon 3 and exon 24 were chosen. Exon 3 codes for the start of TPR domains, and exon 24 starts the coding sequence for the JmjC domain. Exons 1 and 2 sit very close together (Figure 6-2), and there were no high score guides available for the intron before or after, so it was decided to use the guides which match the coding sequence inside the exons. Based on 4 different combinations of 2 guides for each exon, it would be possible to determine which guide pair was most efficient in creating a frameshift mutation. According to the Ensembl website, exon 6 is one of the few exons present in all *UTY* transcripts. It is also 121bp long (that is, not a multiple of three), so successful deletion would shift the reading frame and create multiple stop codons thereafter (not shown).

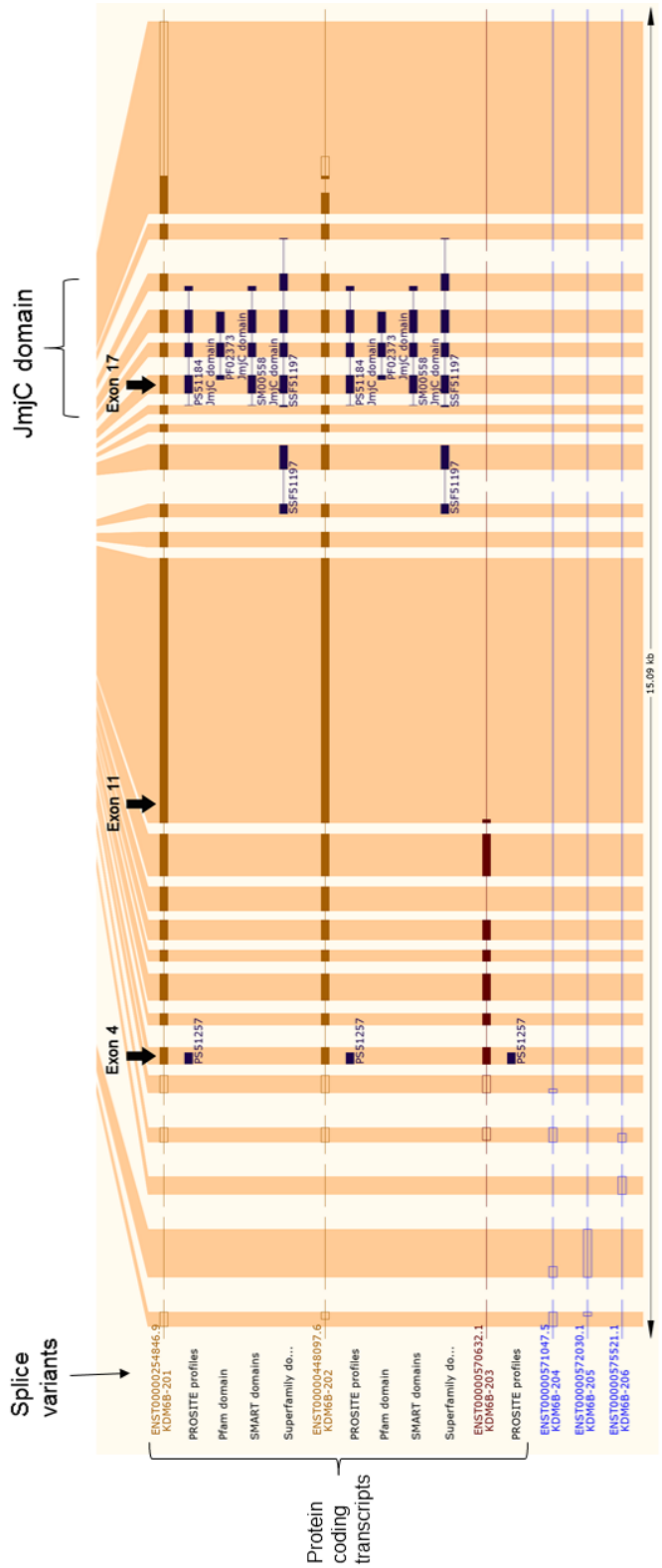

Figure 6-1 Six splice variants of KDM6B and chosen exons for CRISPR-Cas9 genome editing. Image adapted from Ensembl website. Three transcripts are protein coding (gold and red), the other three are not (blue). Depicted here are exons chosen for CRISPR-Cas9 editing. JmjC domains are shown as predicted by Prosite, Pfam and SMART domains.

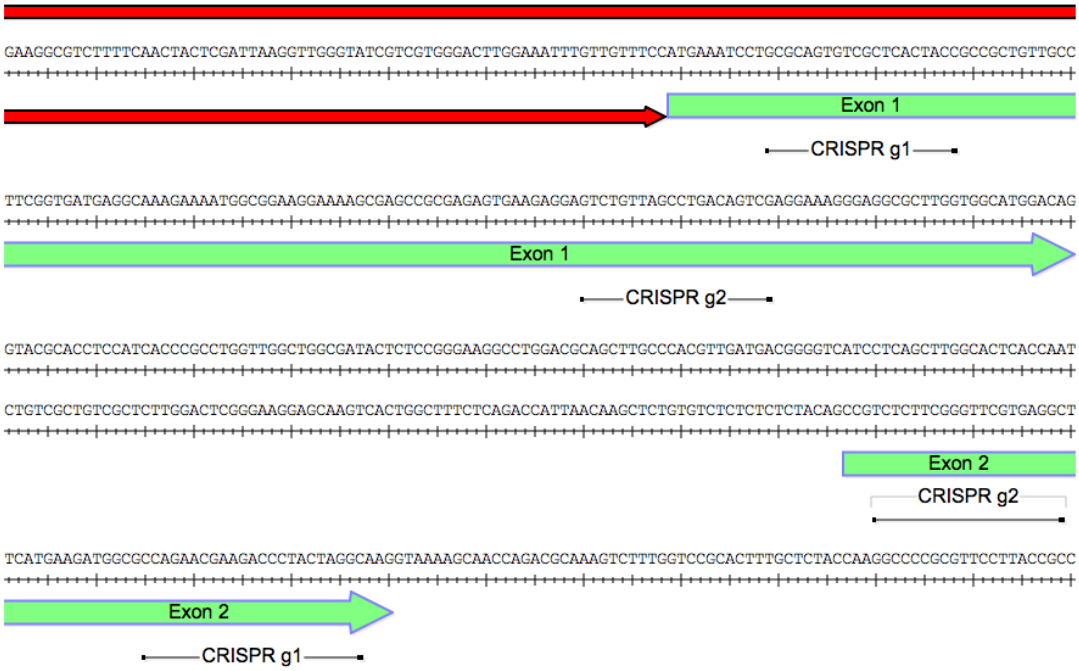

Figure 6-2 Location of CRISPR guides in exons 1 and 2 in UTY. Figure made from SeqBuilder, Lasergene suite. 5' UTR is pictured as red arrow, exons are shown as green arrows.

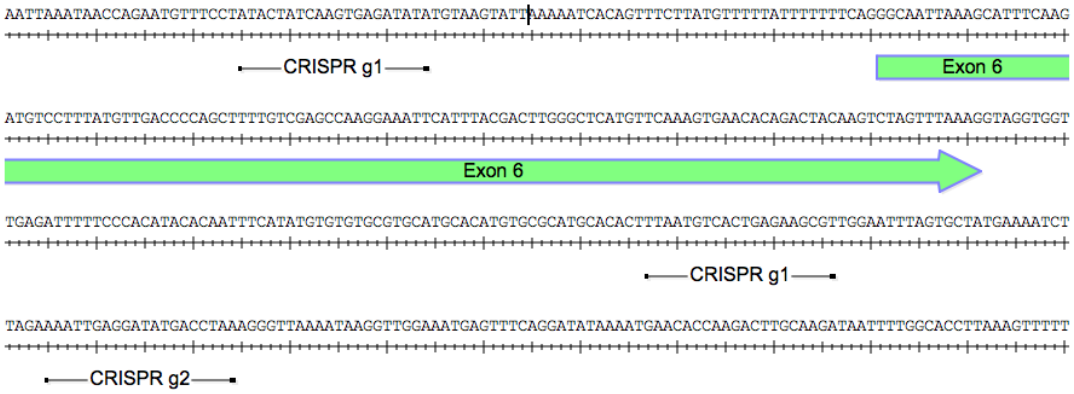

Figure 6-3 Location of CRISPR guides for deletion of exon 6 in UTY. Figure made from SeqBuilder, Lasergene suite. Exons are shown as green arrows.

6.3.2 Knockout of *USP16* in THP-1 cells was successful

Use of the *USP16* guides resulted in insertion of a single nucleotide. *USP16* g1 plasmid was nucleofected into THP-1 cells with low passage number, and GFP positive cells were single cell sorted into 96-wells. Out of two plates, 15 clones survived the sorting and 3 clones proved to be heterozygous for a one nucleotide insertion (A) (Figure 6-4). No homozygotes were present at the first targeting, so two of the heterozygotes were used for targeting the second time with the same guide (*USP16* exon 5 g1). Out of 3 plates for each heterozygous parental cell line, 67 clones survived the sorting (Table 6-4). 13 clones (19%) had reverted back to the wildtype sequence and 4 clones (6%) were homozygous for the same adenine insertion in exon 4 (Figure 6-4). One clone (hereafter named as Hom 1) was derived from a different parental heterozygote (Het C) than the other three (Hom 2, Hom 3, Hom 4) which came from another heterozygote not featured in the further experiments.

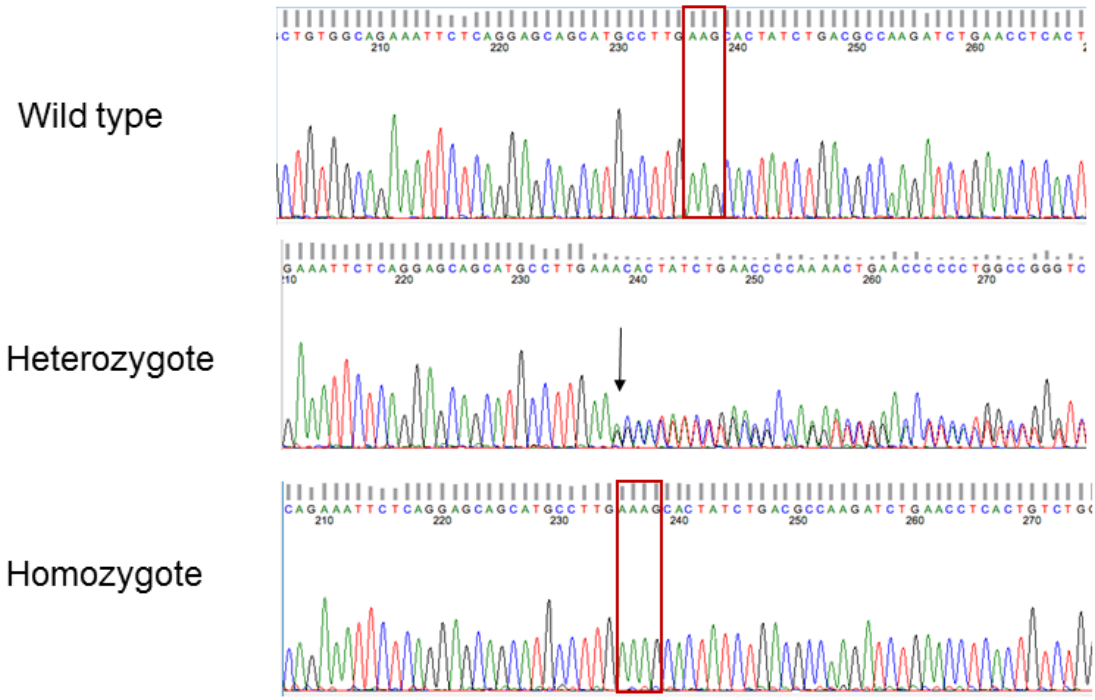

Figure 6-4 Sanger sequencing of exon 4 of USP16 clones. All three founder heterozygous clones had the same insertion of adenine (A) in one allele (indicated by arrow). Four final homozygous USP16 $^{-/-}$ clones had the same insertion of one additional adenine, but on both alleles (indicated by red box). The sequence after the insert was identical to wildtype.

	Number of clones	% of all clones
Wildtype	13	19
Heterozygotes	50	75
Homozygotes	4	6
All	67	100

Table 6-4 Results of the second targeting of USP heterozygotes. Clone genotypes verified by Sanger sequencing.

To determine the impact of the genetically edited transcript, the sequence was translated by bioinformatics tools (Figure 6-5). The insertion site of the adenine in exon 4 codes for the amino acid alanine in the zinc finger domain of USP16. When this single additional nucleotide is present, a further string of incorrect amino acids is abruptly terminated by a stop codon ten amino acids downstream. The transcript also has numerous other stop codons along the way, ensuring that the resulting peptide is small and contains only the start of the protein sequence and no functional domains (Figure 6-5).

To further validate the knockout of *USP16* in the homozygous clones, their RNA was extracted and cDNA synthesised (section 2.2). The cDNA was subjected to Sanger sequencing and the insertion was confirmed to be present in the transcribed RNA. All four putative homozygous clones had the extra adenine inserted in their cDNA.

The next step was to find out whether the level of *USP16* transcripts was lowered in the edited lines, potentially due to nonsense mediated decay. Quantitative reverse transcriptase PCR experiments were conducted with the four homozygotes, two heterozygotes and wildtype THP-1 cells for 3 sets of primers, each set in a different exon along the transcript. The results show that *USP16* homozygous clones had decreased level of expression compared to wildtype THP-1 cells (Figure 6-6, Figure 6-7, Figure 6-8).

Met G K K R T K G K T V P I D D S S E T L E P V C R H I R K G L E
Q G N L K K A L V N V E W N I C Q D C K T D N K V K D K A E E E
T E E K P S V W L C L K C G H Q G C G R N S Q E Q H A L K A L S
D A K I **Stop** T S L S G S **Stop** F G Q L E C **Met** V L
R **Met** **Stop** **Stop** **Stop** G P V L **Stop** F K P V G S S G **Stop** L C Q K
T S Q H Y N S K A S R E R **Stop** W K Y **Stop** T **Stop** K **Stop** K I R K
R E **Stop** E **Stop** T R E R K E G K H G **Stop** R E S S H E F S L P N
N R E R T Q **Stop** F G K H **Met** F L Q C S Y A E L V T N T S
A **Stop** R T T K R S E N V W N N C K N **Stop** T T **Stop** F G I N R T I
R N K P **Stop** A S R P S Y F S H E P V S **Stop** **Stop** D A R D Q K G
G C D T E R T L F S G L **Stop** K S S A V **Stop** R L S A A R Q P G A
A S L L I G W D E S R R T P K S E **Stop** R N T **Stop** S I W **Stop** F
Y **Stop** K V G **Stop** R T K K **Stop** S **Stop** R L**Stop** E E K I N A K F
C **Stop** P H L W W **Stop** T N **Stop** Y D H V **Stop** S **Met** Q N C L L G
S **Stop** I F P **Stop** F V P P S F R **Stop** S E W **Stop** E K C
K **Stop** **Stop** K S E K D S G G **Stop** R S R **Stop** **Stop** G R K

Figure 6-5 Translation of edited USP16 in homozygous clones by the ExPaSy tool (single letter code). Insertion of the single nucleotide induces a frame shift mutation, which introduces multiple stop codons in the translated sequence. The insert is present in the middle of the zinc finger domain, in codon 91 for the amino acid alanine (pictured here framed by yellow box). Ten amino acids later, translation would be terminated by an in frame stop codon. Met indicates potential start codons (amino acid methionine), and highlighted sequences show open reading frames. The first grey sequence up to the yellow box is the correct amino acid sequence.

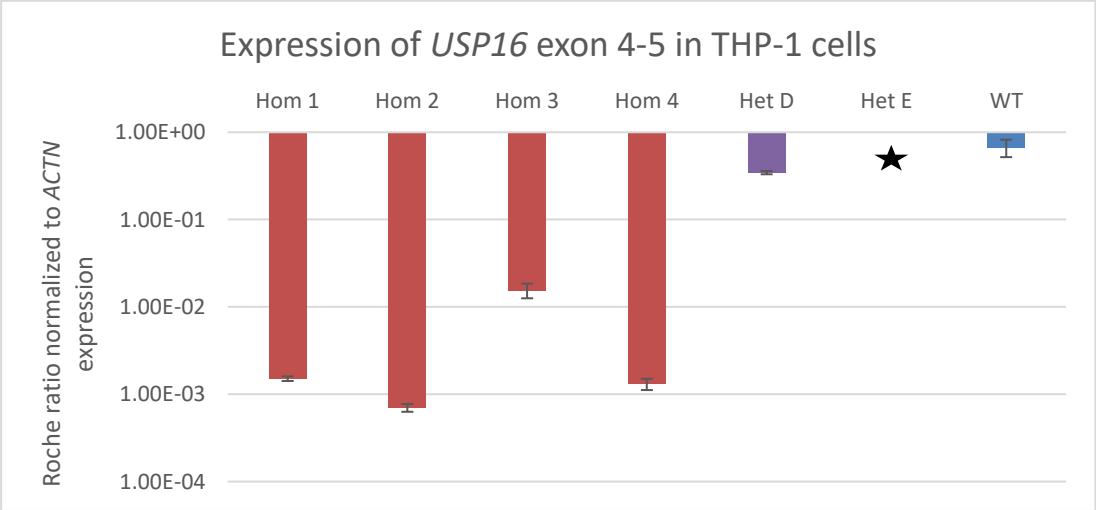

Figure 6-6 qRT-PCR results of USP16 knockout clones and wildtype THP-1. Pictured here is Roche ratio of USP16 exon 4-5 primers normalized by ACTN expression with error rates calculated by Roche LightCycler480 software (see section 2.2.4). Het E result (asterisk) had a technical failure. The Y-axis is in logarithmic scale. Hom = Homozygote for USP16 knockout, pictured in red, Het = Heterozygote for USP16 knockout, pictured in purple, and WT = wildtype THP-1 cells, pictured in blue. Error bars show standard error.

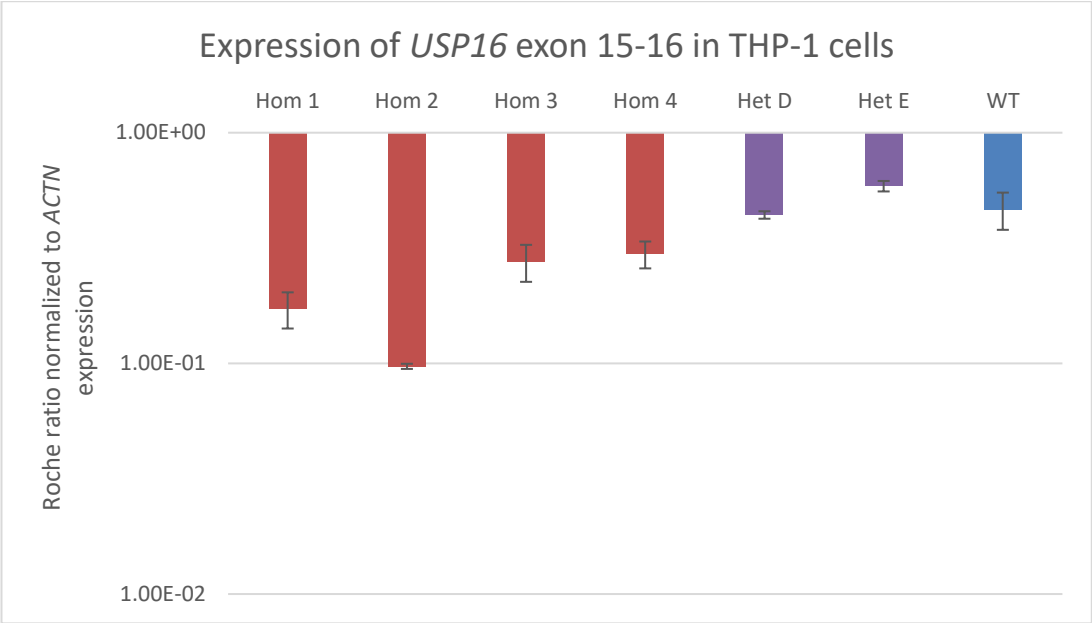

Figure 6-7 qRT-PCR results of USP16 knockout clones and wildtype THP-1. Pictured here is Roche ratio of USP16 exon 15-16 primers normalized by ACTN expression with error rates calculated by Roche LightCycler480 software (see section 2.2.4). The Y-axis is in logarithmic scale. Hom = Homozygote for USP16 knockout, pictured in red, Het = Heterozygote for USP16 knockout, pictured in purple, and WT = wildtype THP-1 cells, pictured in blue. Error bars show standard error.

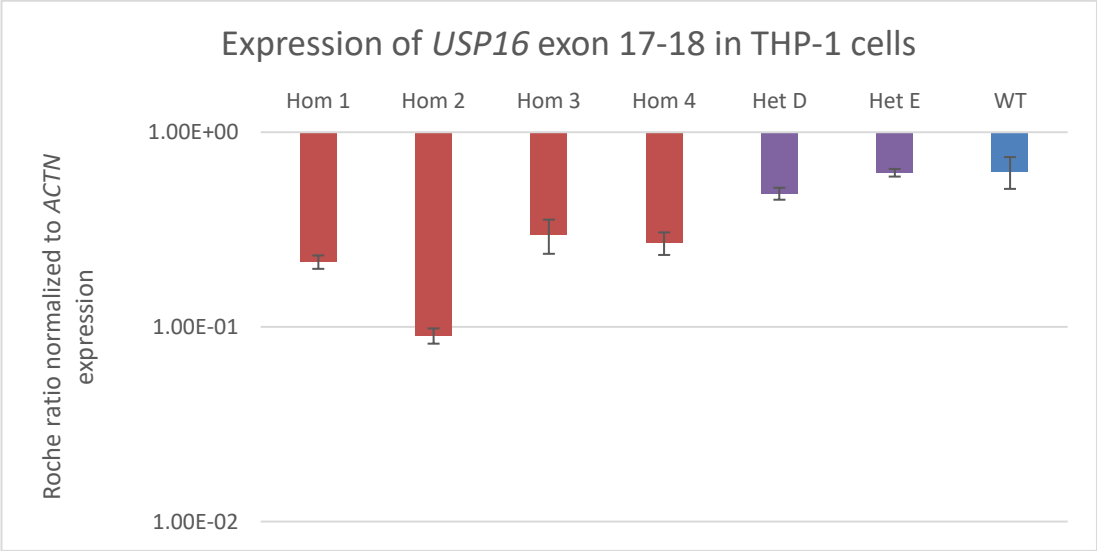

Figure 6-8 qRT-PCR results of USP16 knockout clones and wildtype THP-1. Pictured here is Roche ratio of USP16 exon 17-18 primers normalized by ACTN expression with error rates calculated by Roche LightCycler480 software (see section 2.2.4). The Y-axis is in logarithmic scale. Hom = Homozygote for USP16 knockout, pictured in red, Het = Heterozygote for USP16 knockout, pictured in purple, and WT = wildtype THP-1 cells, pictured in blue. Error bars show standard error.

The final validation of the *USP16* knockout was to see whether the altered (and reduced) transcript of the *USP16* knockout homozygotes made USP16 protein by

western blotting (Figure 6-9). In the figure, the lower actin bands, identified by an arrow, shows that the loading of protein lysates was constant for all samples. This actin antibody did not produce any other bands in these protein samples (not pictured), so all other bands were detected by anti-USP16 antibody. Various bands were present, but the one around 120 kDa (red box in the figure) was absent in homozygous clones, whereas it was the strongest one for wildtype and heterozygotes. Predicted size of USP16 according to ab121650 datasheet is 94 kDa, but Xu *et al*. (2013) confirmed the size of their USP16 to be 110 kDa and Joo *et al*. (2007) found the peptide to be 120 kDa in SDS-PAGE. It can be therefore concluded that *USP16* KO homozygotes do not produce a correct USP16 protein.

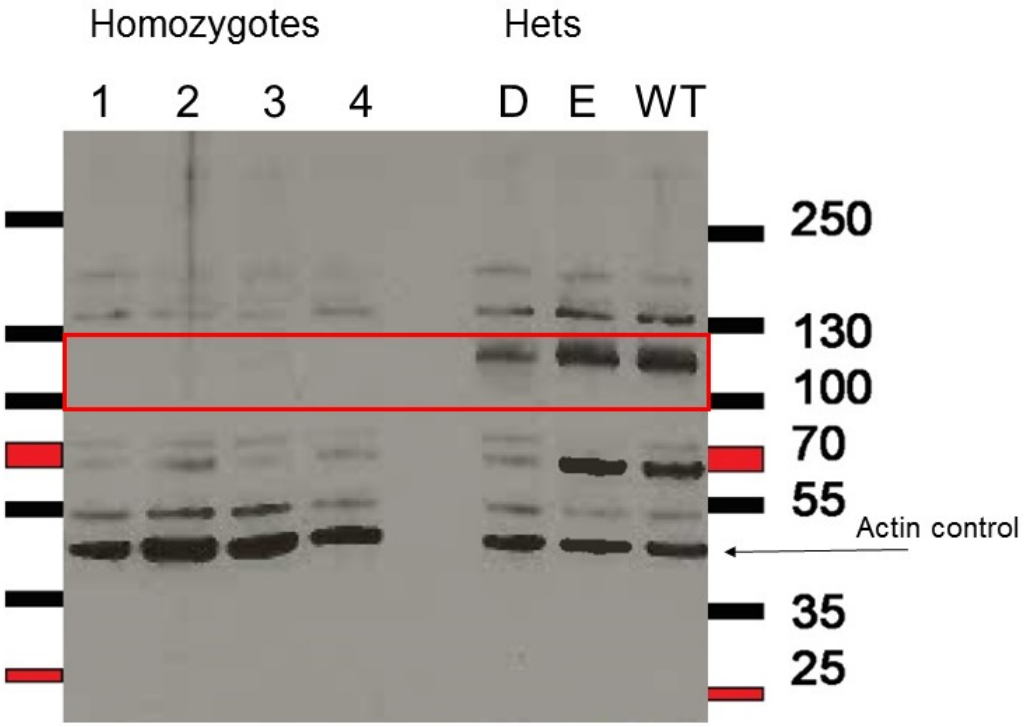

Figure 6-9 Western blotting for USP16 and actin in extracts of THP-1 clones. Ladder sizes are in kDa. Lanes 1, 2, 3, 4 are USP16 KO homozygotes; D, E are USP16 KO heterozygotes, and WT denotes wildtype THP-1 cell line. Actin loading control bands of size of 42 kDa are indicated by an arrow. All other bands are detected by USP16 antibody (ab121650), bands around 120 kDa (present in heterozygotes and wildtype) are not present in four homozygous knockout USP16 THP-1 clones.

6.3.3 Deletion of *USP16* in THP-1 cells did not prevent macrophage differentiation

After establishing that the homozygous knockout clones did not produce a functional USP16 protein, the first experiment was to determine whether USP16 protein deficiency interrupts the monocyte to macrophage differentiation of THP-1 cell line.

Knockout clones were treated with the differentiation agent PMA, and exhibited the same differentiation pattern as wildtype THP-1 after 48 hours (Figure 6-10). The next step was to see whether *USP16* homozygous clones' ability to phagocytose (one of the key functions of macrophages) would be impaired. THP-1 macrophages (derived by treatment with PMA) were exposed to Zymosan A particles for an hour, the excess of particles was washed away and then macrophages were viewed on a fluorescent microscope. The results showed that USP16 clones were able to phagocytose the particles at approximately the same level as wildtype cells (Figure 6-10).

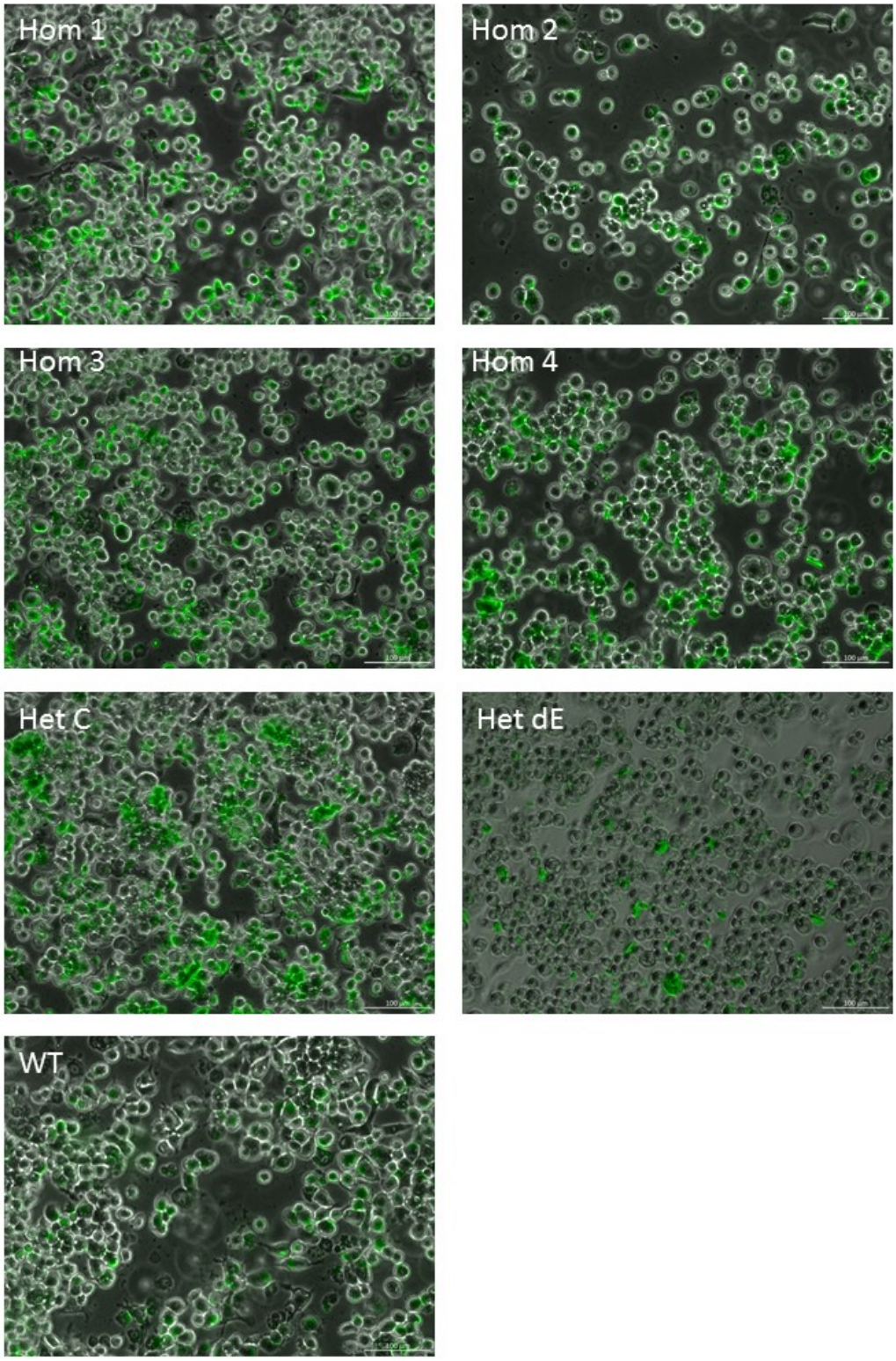

Figure 6-10 Brightfield and FITC signal overlay of phagocytosis assay (phagocytosis beads are FITC +) of USP16 clones and wildtype macrophages (2 days post- PMA treatment). Scale bar is 100 µm.

6.3.4 Deletion of *USP16* in THP-1 cells affected proliferation and cell cycle progression during differentiation

To see whether knockout of *USP16* affects THP-1 monocyte proliferation rate, MTT assays were conducted at multiple times throughout the culturing of these cells. Cells' metabolic activity was measured by this colorimetric assay at day 0 (the time of seeding) and after two days of culture (day 2). Two results from two assays are presented, one from the beginning of the project (I) and the other from the end (II).

The first MTT assay reading (Figure 6-11) was used to calculate the difference in metabolic activity over the two days, which should show the proliferation rate of the different clones (Figure 6-12). Wildtype (WT) had the highest rate of proliferation (three-fold increase in metabolic activity measured by MTT in two days), followed by heterozygotes and homozygote clonal line 1. *USP16* knockout clonal lines 2, 3, and 4 had smaller increase in metabolic activity compared to WT and the other clones. One sample t-test showed that homozygous cell lines proliferated more slowly than wildtype (p-value = 0.032).

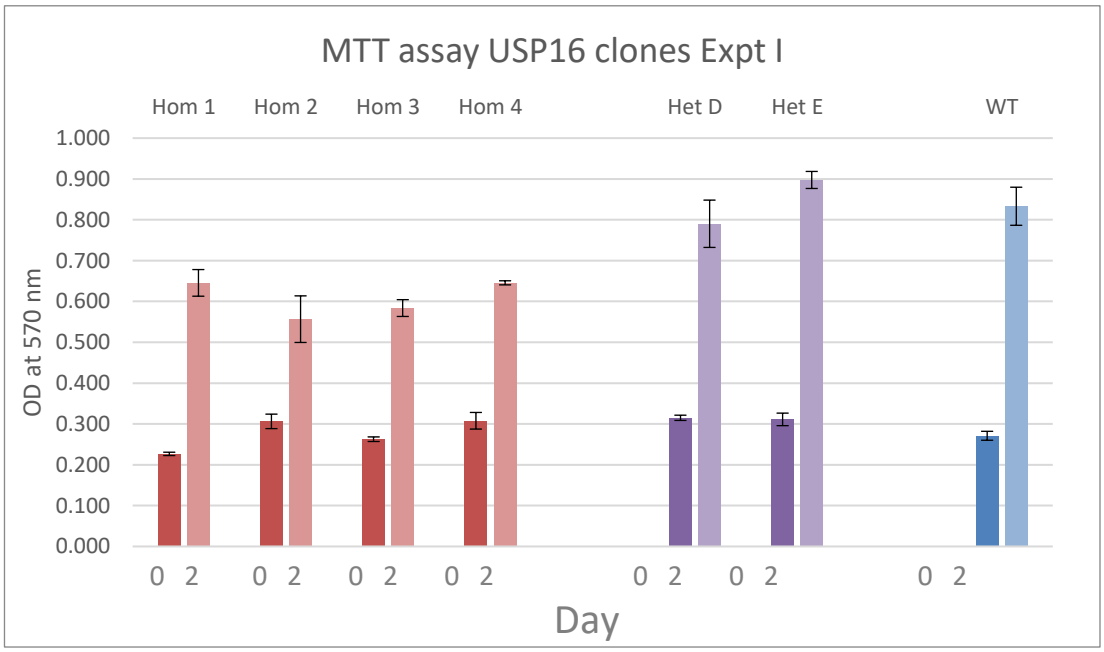

Figure 6-11 MTT assay USP16 knockout clones vs wildtype THP-1 cells, Experiment I, from early in the project. Day 0 is symbolized by bolder colour, day 2 by paler colour. Hom = Homozygote for USP16 knockout are pictured in red, Het = Heterozygote for USP16 knockout are pictured in purple, and WT = wildtype THP-1 cells are pictured in blue. Error bars represent standard deviation of the readings.

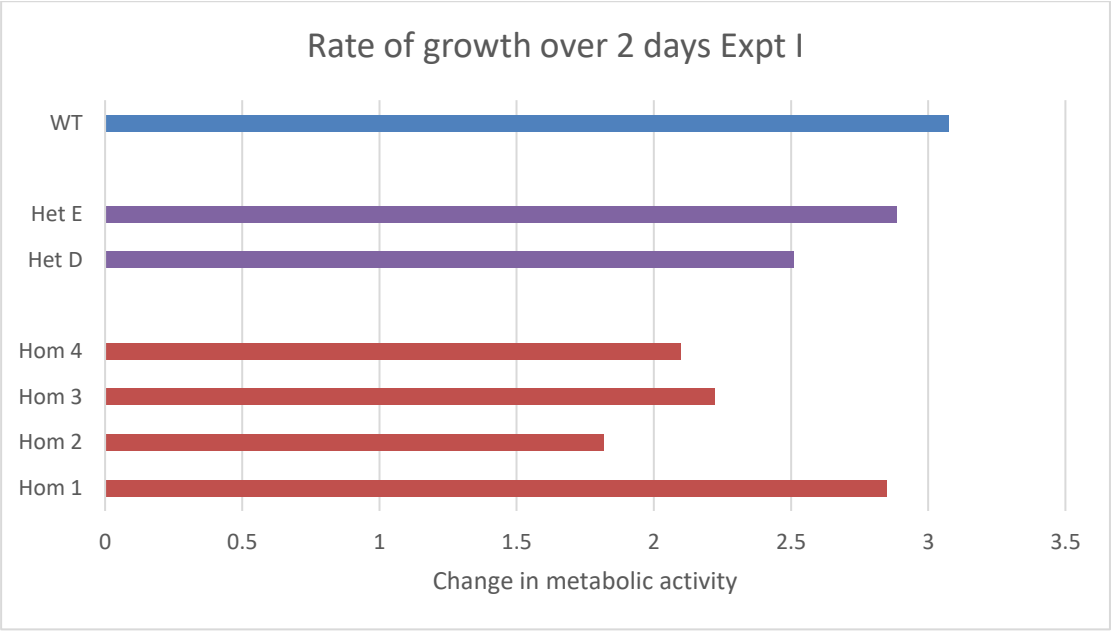

Figure 6-12 Rate of growth of USP16 clones calculated from MTT assays at the beginning of the project. Hom = Homozygote for USP16 knockout, Het = Heterozygote for USP16 knockout, WT = wildtype THP-1 cells.

To minimise technical errors and random biological variation and to assess any drift in clone phenotype over time, this experiment was repeated multiple times throughout the clones' time in culture. Figure 6-13 shows a different reading from the end of the project, where the clones had been in culture for multiple passages. At this time, the parental heterozygote cell line (Het C) for homozygous clone 1 (Hom 1) was included. Comparing the change in metabolic activity over two days for these clones (Figure 6-14), it seems that cells maintained a similar rate of proliferation as in the beginning of their time in culture (Figure 6-11, Figure 6-12). It was noted that the parental heterozygote for homozygous clone 1 (Het C) had similar rate of change of metabolic activity to the homozygous clone 1, which was at similar level to wildtype.

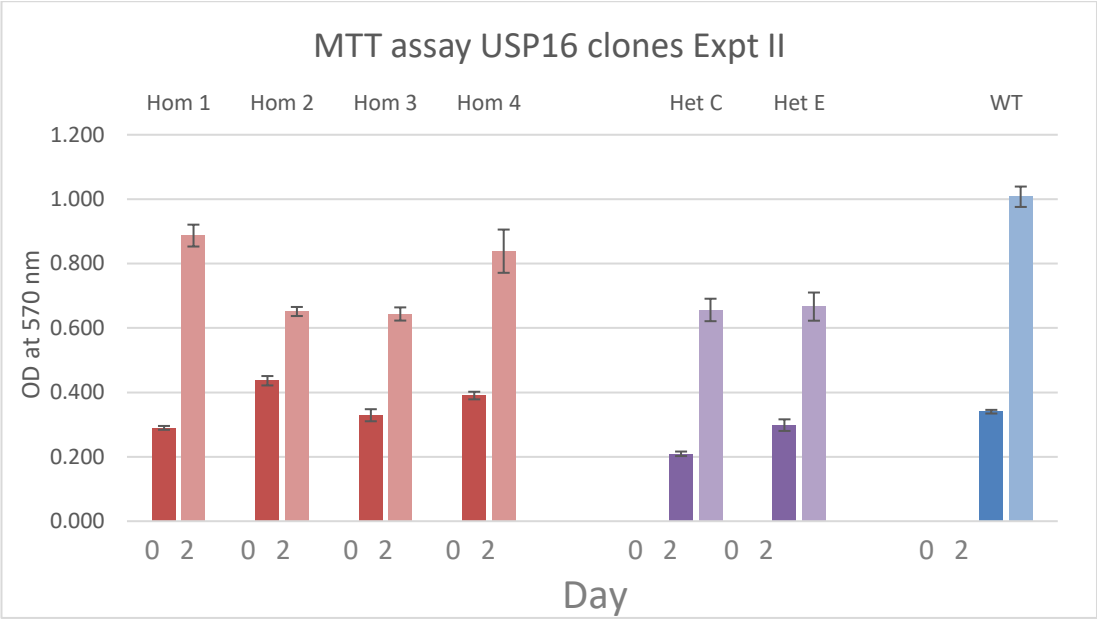

Figure 6-13 MTT assay USP16 knockout clones vs wildtype THP-1 cells. Experiment II, from the end of the project. Day 0 is symbolized by bolder colour, day 2 by paler colour. Hom = Homozygote for USP16 knockout are pictured in red, Het = Heterozygote for USP16 knockout are pictured in purple, and WT = wildtype THP-1 cells are pictured in blue. Error bars represent standard deviation of the readings.

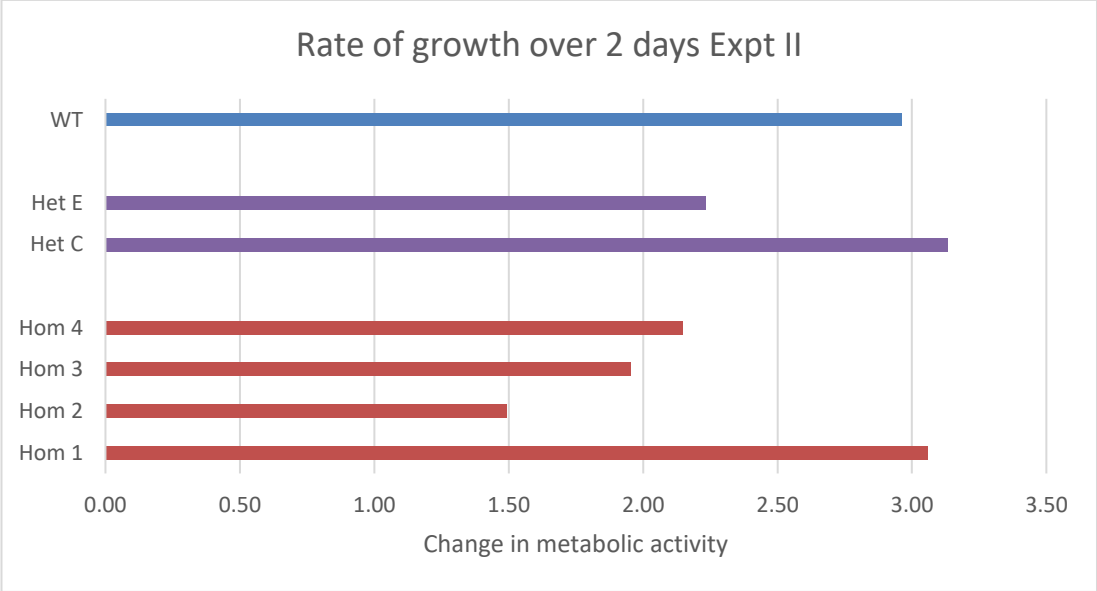

Figure 6-14 Rate of growth of USP16 clones calculated from MTT assays at the end of the project. Hom = Homozygote for USP16 knockout, Het = Heterozygote for USP16 knockout, WT = wildtype THP-1 cells.

Calculating statistical significance using two-sample t-test between wildtype and homozygous clones using data from both MTT assays show that the rate of growth of homozygous *USP16* clones over two days was significantly slower than that of wildtype (p-value = 0.004). However, the difference in rate of growth over 2 days for heterozygotes and wildtype was not significant (p-value = 0.211).

To find out whether the deficiency of USP16 protein in THP-1 cells has a role in cell cycle progression, *USP16* homozygous and heterozygous knockout clones, together with wildtype THP-1, were differentiated into macrophages with PMA, over the course of 3 days, checking the proportions of cells in different phases of the cell cycle every day by propidium iodide staining.

This experiment was repeated four times. Before differentiation, the *USP16* homozygous clones behaved similarly to the wildtype (Figure 6-15). Around 20% cells were in the S-phase, which number should decline after differentiation as proliferation stops. Indeed, as seen on day 2, the proportion of cells in S-phase was drastically reduced to around 10%, while gaining G1 phase cells in wildtype and heterozygotes (Figure 6-16). However, in 3 of the homozygous knockout clones (Hom 2, Hom 3 and Hom 4) there was no change in percentage of cells in S-phase, whereas the G2 proportion of cells had more than doubled. Hom 1 follows similar pattern to wildtype (WT) progression.

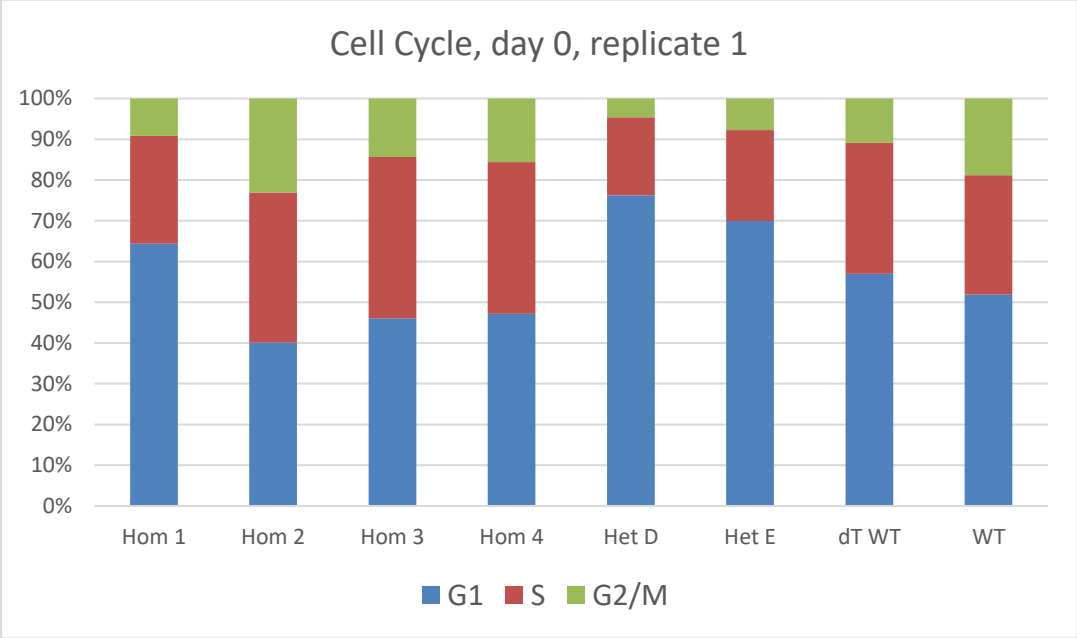

Figure 6-15 Cell cycle phases for non-differentiated THP-1 monocytes at day 0. DNA was stained by propidium iodide, detected by Flow Cytometry and cell cycle phases were assigned by FlowJo (See section 2.4.2). Homozygous clones for USP16 knockout are Hom 1, Hom 2, Hom 3 and Hom 4. Heterozygous clones for USP16 knockout are Het D and Het E. Double-targeted wildtype clone (which reverted to wildtype after the second targeting with USP16 g1) is dT WT. For wildtype THP-1 (WT), the passage number was 11.

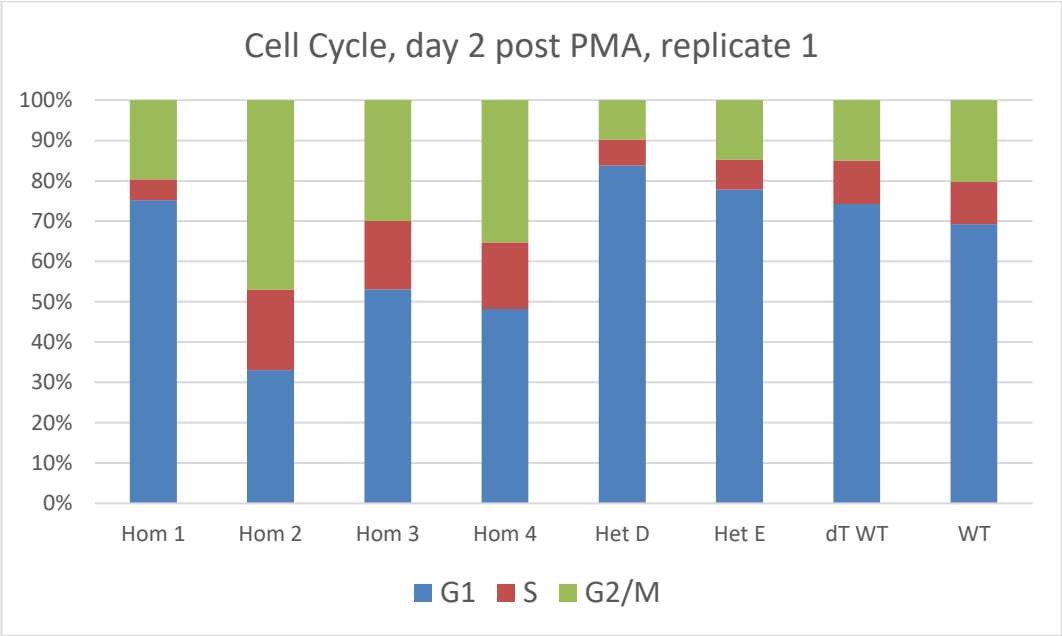

Figure 6-16 Cell cycle phases for differentiated THP-1 monocytes at day 2. DNA was stained by propidium iodide, detected by Flow Cytometry and cell cycle phases were assigned by FlowJo (See section 2.4.2). Homozygous clones for USP16 knockout are Hom 1, Hom 2, Hom 3 and Hom 4. Heterozygous clones for USP16 knockout are Het D and Het E. Double-targeted wildtype clone (which reverted to wildtype after the second targeting with USP16 g1) is dT WT. For wildtype THP-1 (WT), the passage number was 11.

These results were not replicated in a consistent manner in the next three experiments. All replicates had the four homozygous clones and wildtype present, but the heterozygous clones and double targeted wildtype clone were not always included. Looking at the average values for cell cycle phases at day 0 (Figure 6-17), the results show quite high standard deviation and standard error across S phase and G2 phase (Table 6-5). The percentage of cells which were in S-phase was around the same for all clones before the differentiation (day 0, around 30%). However, Hom 2 and WT had a higher number of cells in G2 phase than other clones; and also lower number of cells in G1 than the others.

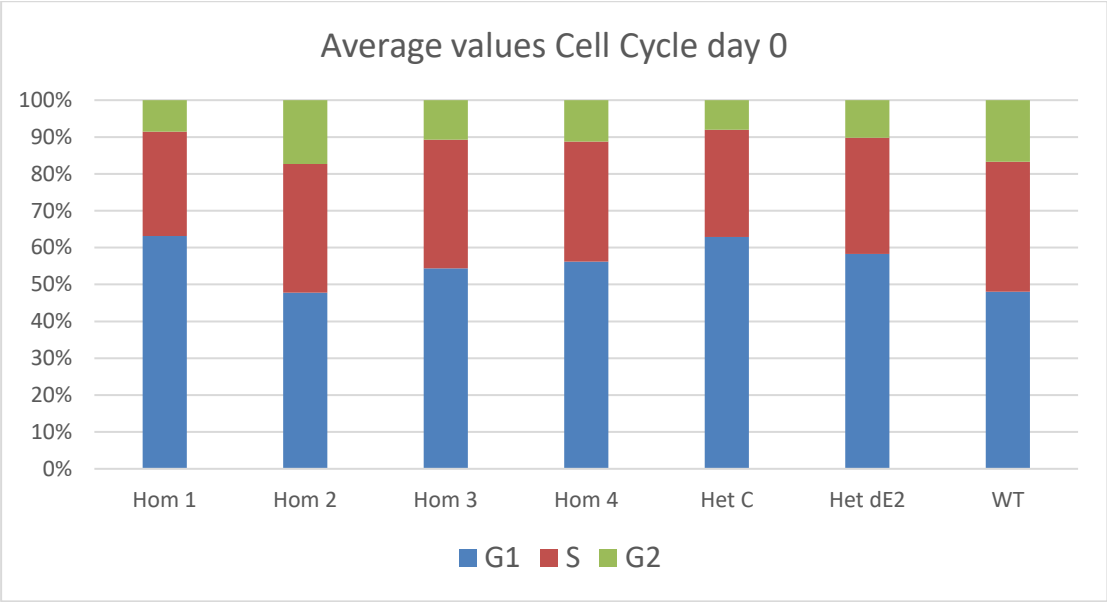

Figure 6-17 Average values at day 0 for each phase of the cell cycle for 4 USP16 homozygous clones (N=4 each), one parental heterozygous clone (Het C, N=3), one double targeted heterozygous clone (Het dE2, N=2) and wildtype (WT, N=4). Sum of three cell cycle phases G1, S, G2 equals 100%. SD and SE are given in Table 6-5.

Day 0

Standard deviation	Hom 1	Hom 2	Hom 3	Hom 4	Het C	Het dE2	WT
G1	0.06443	0.05404	0.07792	0.07551	0.02744	0.08212	0.03478
S	0.06059	0.02149	0.07317	0.05566	0.03247	0.08160	0.04885
G2	0.00777	0.04421	0.02335	0.02783	0.01952	0.00051	0.01691
Standard error	Hom 1	Hom 2	Hom 3	Hom 4	Het C	Het dE2	WT
G1	0.03221	0.03120	0.03896	0.03775	0.01585	0.05806	0.01739
S	0.03029	0.01241	0.03659	0.02783	0.01875	0.05770	0.02442
G2	0.00388	0.02552	0.01167	0.01391	0.01127	0.00036	0.00846

Table 6-5 Standard deviation and standard error calculated for Cell Cycle day 0 values, 1 = 100%.

As mentioned previously, Hom 1 seems to behave differently to other homozygous clones after differentiation. Average values for day 2 (Figure 6-18) show that Hom 1 has very similar profile to its parental clone, Het C. The results indicate that Hom 2 is the most different to other homozygous clones (and also wildtype), with the highest number of cells in S and G2 phase. The standard deviation and standard error for day 2 were also quite high (Table 6-6), which suggest that the results were not consistent between the biological replicates of this experiment.

Taking all biological replicates together, the difference in the proportion of cells in G2 or S phase between homozygotes and wildtype samples was not significantly different either on day 0 or on day 2 (two-sample t-test, p-value > 0.25 in all cases).

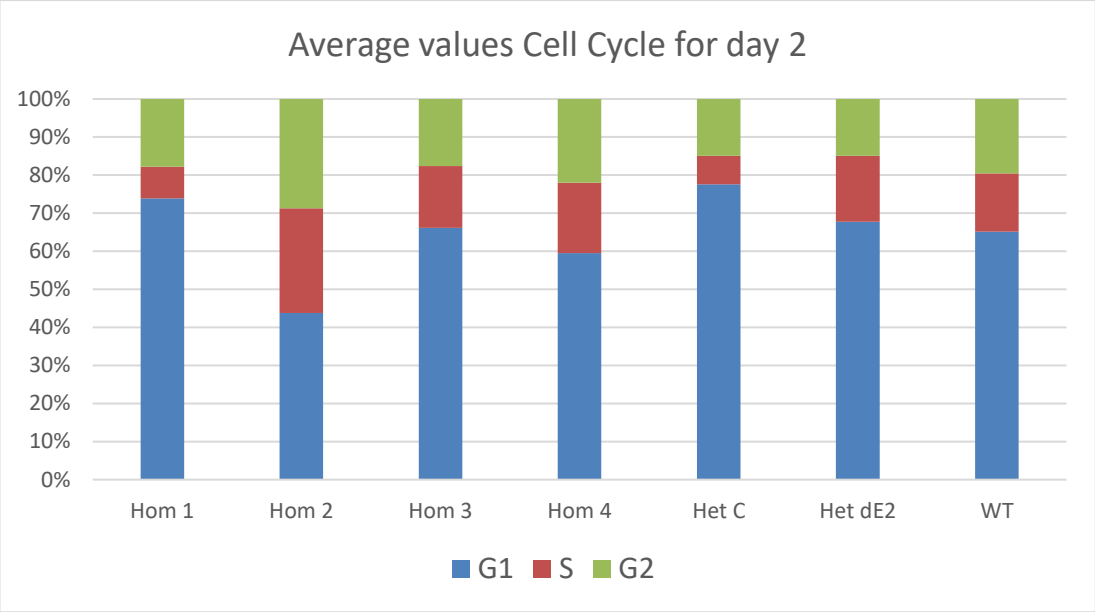

Figure 6-18 Average values at day 2 for each phase of the cell cycle for 4 USP16 homozygous clones (N=4 each), one parental heterozygous clone (Het C, N=3), one double targeted heterozygous clone (Het dE2, N=2) and wildtype (WT, N=4). Sum of three cell cycle phases G1, S, G2 equals 100%. SD and SE are given in Table 6-6.

Day 2

Standard deviation	Hom 1	Hom 2	Hom 3	Hom 4	Het C	Het dE2	WT
G1	0.0636	0.0846	0.0895	0.1097	0.0693	0.1155	0.0714
S	0.0330	0.0591	0.0255	0.0464	0.0230	0.0581	0.0773
G2	0.0398	0.1072	0.0799	0.0883	0.0631	0.0574	0.0377
Standard error	Hom 1	Hom 2	Hom 3	Hom 4	Het C	Het dE2	WT
G1	0.0318	0.0423	0.0447	0.0548	0.0400	0.0817	0.0357
S	0.0165	0.0295	0.0127	0.0232	0.0132	0.0410	0.0387
G2	0.0199	0.0536	0.0399	0.0441	0.0364	0.0406	0.0188

Table 6-6 Standard deviation and standard error calculated for Cell Cycle day 2 values, 1 = 100%.

In order to find out more about these USP16 homozygous and heterozygous clones, their RNA at day 0 and 24 hours post differentiation was selected for CAGE sequencing.

6.3.5 CAGE sequencing of *USP16* clones

As discussed in Chapter 3, all CAGE reads were mapped to hg38 and normalised between each other using power law distribution. The 22 CAGE samples included in the analysis for this chapter were all knockout samples (homozygotes and heterozygotes for *USP16*) and wildtype controls from THP-1 differentiation from Chapter 3 for 0 hours and 24 hours.

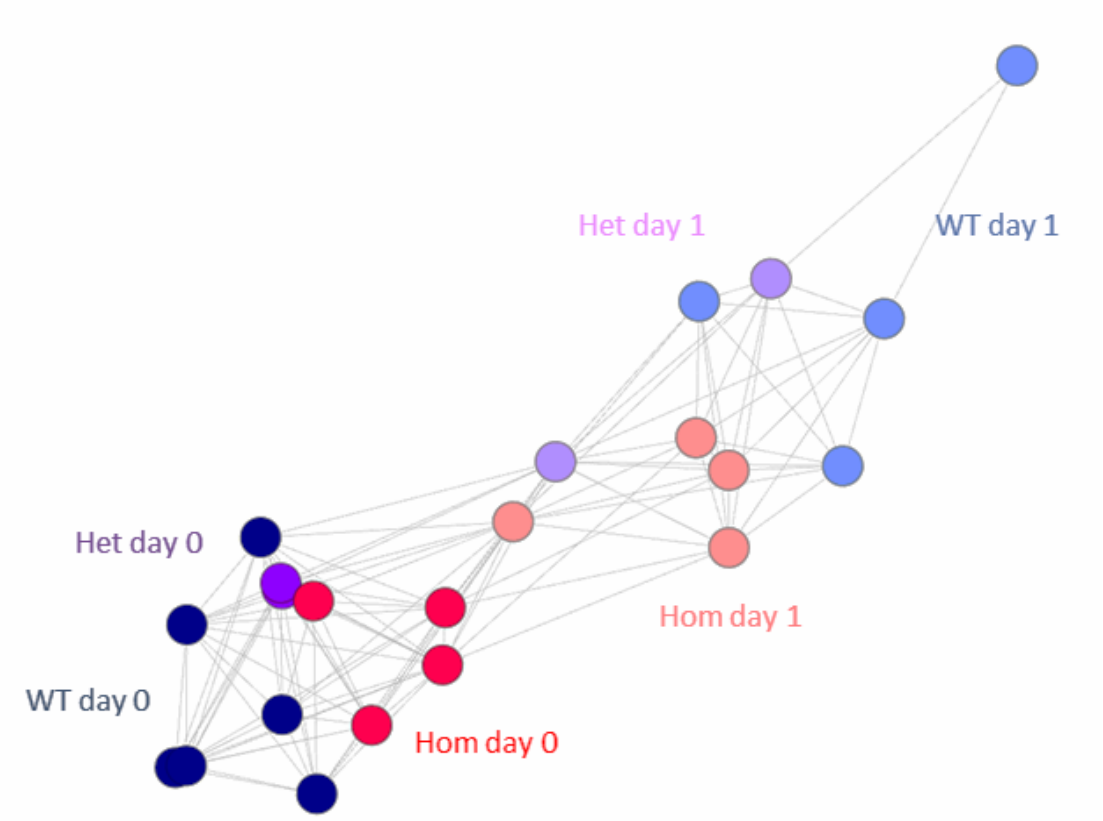

Figure 6-19 Sample-sample correlation of correlation coefficient of 0.88 derived from Miru programme. Samples are nodes and edges are connections between them with a correlation coefficient of 0.88 or greater, ie how closely correlated the expression patterns are to each other across all samples. Nodes which are close are highly correlated. Each sample type is different colour: wildtype – blue, USP16 heterozygotes – purple, USP16 homozygotes – red. Day 0 is bolder colour, Day 1 paler.

All samples were then correlated to each other with Pearson's correlation coefficient of 0.88 (Figure 6-19) using Miru software. From the Figure 6-19, it can be seen that

all samples changed in expression pattern between Day 0 and Day 1. All undifferentiated samples (Day 0, bolder colours) are on the left of the figure and their expression correlates between each other more than the differentiated samples (Day 1, paler colours), which is primarily on the right. The greatest difference between Day 0 and Day 1 was in the wildtype and one Day 1 sample (Dif2_23.2) showed the most extreme difference. The *USP16* heterozygotes were in general similar to wildtype but for the *USP16* homozygotes the Day 0 and Day 1 samples were closer to each other in the network, indicating a reduced transition following PMA treatment.

Next, the expression levels of the *USP16* gene were calculated using the Zenbu data viewer. Figure 6-20 shows the *USP16* expression based on CAGE sequencing in TPM. Homozygotes (Hom) had around half of *USP16* expression in heterozygotes (Het) or wildtypes (WT), which supports the data from qRT-PCR (Figure 6-6, Figure 6-7, Figure 6-8). After one day of treatment with PMA (day 1), homozygotes had significantly lower *USP16* expression than wildtype (two sample t-test, p-value 0.008). Heterozygotes had around the same *USP16* expression as WT. The highest TPM among homozygotes was in Hom 1 after differentiation, but still less than WT or its parental heterozygote Het C. This is in agreement with the notion that there is still a low level of *USP16* mRNA in the knockout clones (Figure 6-6, Figure 6-7, Figure 6-8), so some is present to be detected by CAGE.

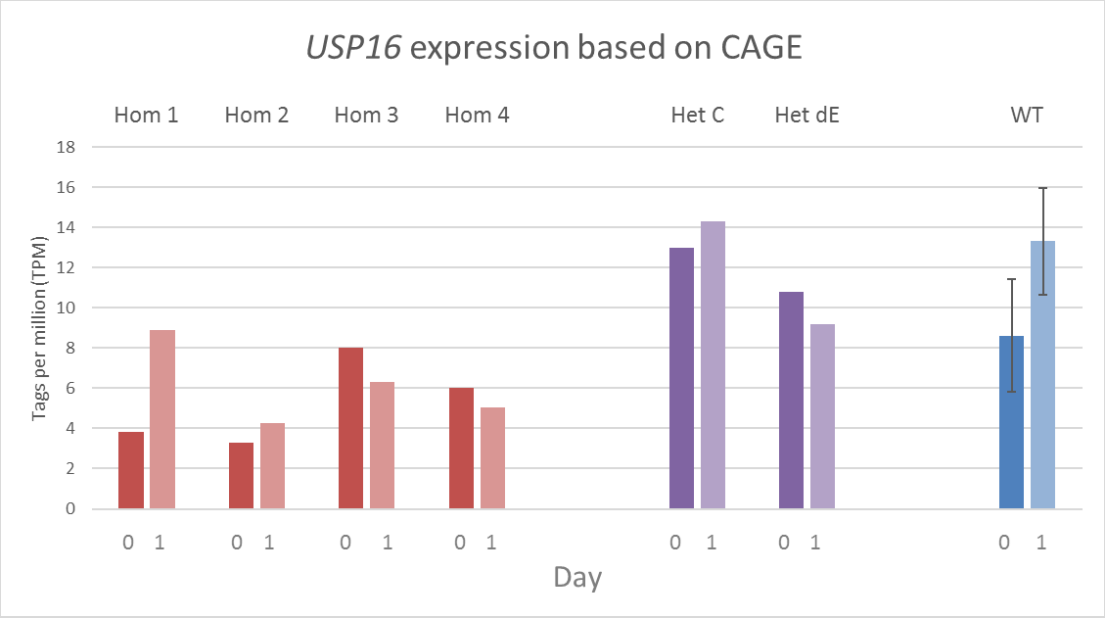

Figure 6-20 USP16 expression in USP16 knockout clones based on CAGE sequencing. Expression is normalised and is given in Tags per million. Day 0 is bolder colour; Day 1 is paler. Wildtype values were calculated using mean of the different biological replicates (6 for day 0, 4 for day 1). Error bars for wildtype show standard deviation.

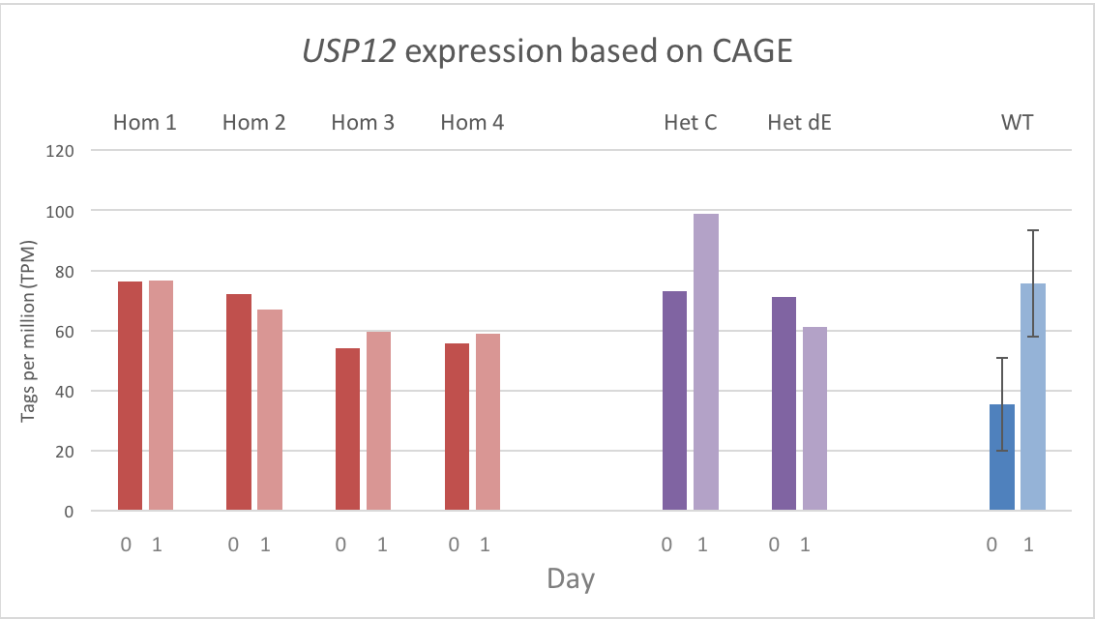

Figure 6-21 USP12 expression in USP16 knockout clones based on CAGE sequencing. Expression is normalised and is given in Tags per million. Day 0 is bolder colour; Day 1 is paler. All wildtype values were calculated using mean of the different biological replicates (6 for day 0, 4 for day 1). Error bars for wildtype show standard deviation.

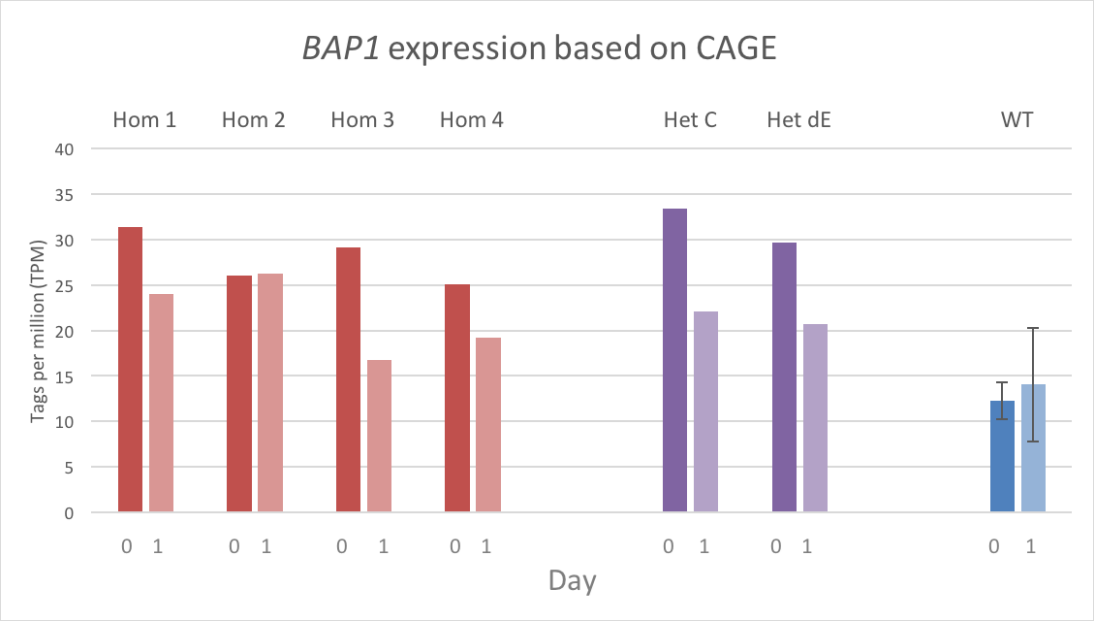

Figure 6-22 BAP1 expression in USP16 knockout clones based on CAGE sequencing. Expression is normalised and is given in Tags per million. Day 0 is bolder colour; Day 1 is paler. All wildtype values were calculated using mean of the different biological replicates (6 for day 0, 4 for day 1). Error bars for wildtype show standard deviation.

The other H2A deubiquitinases, *USP12* and *BAP1,* were upregulated in *USP16* homozygous and heterozygous clones (Figure 6-21, Figure 6-22). *USP12* was significantly upregulated at day 0 (p-value = 0.006, two-sample t-test) but not at day 1 (p-value = 0.170). *BAP1* was significantly upregulated at both data points (p-value = 3.67956E-06 for day 0 and 0.047 for day 1).

The results suggested that other genes of the deubiquitinase family might also be upregulated in *USP16* knockout samples. *MYSM1* was significantly higher than wildtype at day 0 only, whereas *USP21* was higher at both time points (Table 6-7). *USP22* was very slightly upregulated (not significant), but other enzymes like *USP3* and *USP46* were not (Table 6-7).

	MYSM1		***USP21***		***USP22***	
	Day 0	Day 1	Day 0	Day 1	Day 0	Day 1
HOM MEAN	61.841	68.952	32.324	32.081	84.169	72.064
SD	4.610	5.347	9.251	6.964	11.474	8.269
WT MEAN	43.256	72.300	20.159	24.931	102.609	91.453
SD	15.212	25.186	5.256	4.225	17.706	18.588
2-SAMPLE T TEST	0.024	0.352	0.014	0.065	0.053	0.053

Table 6-7 Average values for deubiquitinases in USP16 knockout homozygotes (Hom mean) with standard deviation (SD) vs average values for wildtype (WT mean) with standard deviation (SD). P-values for the difference between homozygotes and wildtype was calculated using 2 –sample t test.

To validate whether *USP16* knockout cells were at a similar level of macrophage differentiation, the macrophage marker *CSF1R* and the progenitor cell marker *MYB* were looked at. *CSF1R* was lower in Hom 2, Hom 3 and Hom 4 samples on the second day compared to wildtype, indicating that they had not differentiated towards macrophages. However, Hom 1 retained the same expression profile as Het C (Figure 6-23), consistent with results in section 6.3.4 showing that it had similar metabolic activity to the heterozygotes and wildtype. *MYB* was upregulated in *USP16* knockout samples at both time points compared to wildtype (Figure 6-24), indicating that they had remained in a less differentiated state, although the profile for Hom 1 was again similar to that for wildtype and heterozygotes.

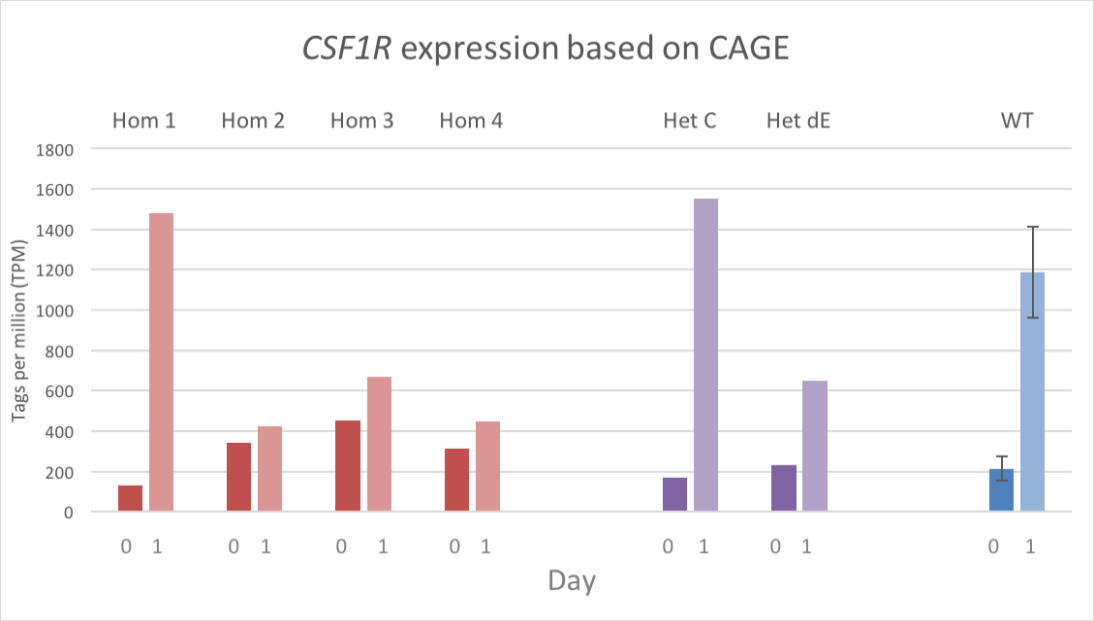

Figure 6-23 CSF1R expression in USP16 clones based on CAGE sequencing. Expression is normalised and is given in Tags per million. Day 0 is bolder colour; Day 1 is paler. All wildtype values were calculated using mean of the different biological replicates (6 for day 0, 4 for day 1). Error bars for wildtype show standard deviation.

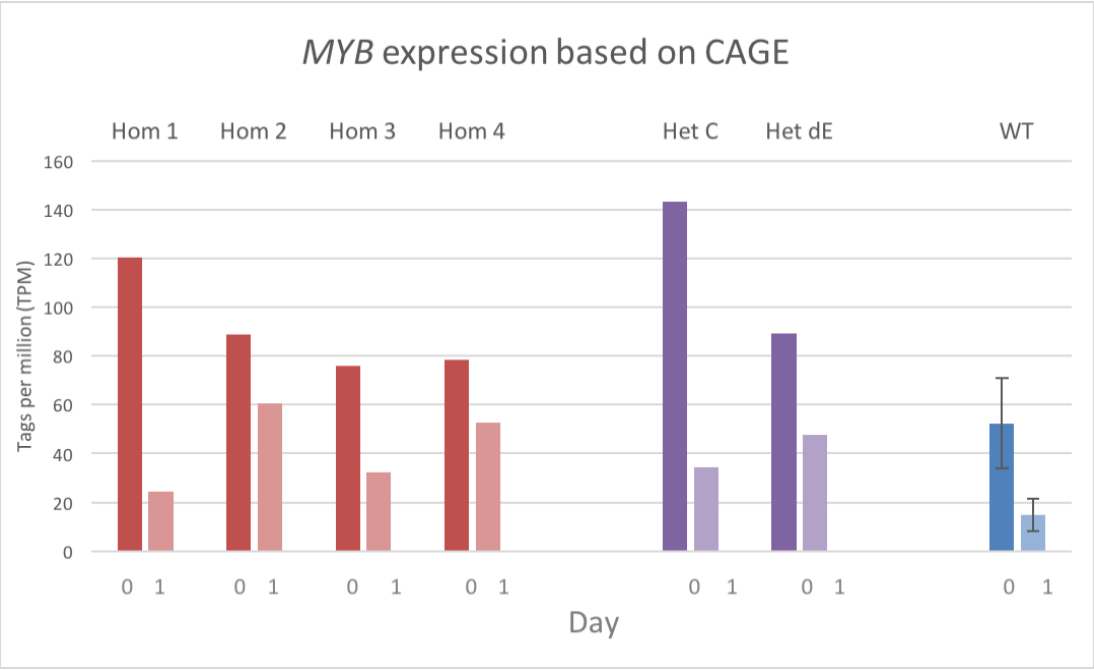

Figure 6-24 MYB expression in USP16 clones based on CAGE sequencing. Expression is normalised and is given in Tags per million. Day 0 is bolder colour; Day 1 is paler. All wildtype values were calculated using mean of the different biological replicates (6 for day 0, 4 for day 1). Error bars for wildtype show standard deviation.

The *USP16* knockout homozygous, heterozygous and wildtype clones were analysed using Miru. A gene-to-gene analysis was performed with Pearson's correlation coefficient of 0.92 (Figure 6-25). Using MCL clustering in Miru (see section 2.5.5) it was determined that there were various clusters of CTSS correlated with each other. Cluster lists are available in Appendix 8.4. The cluster with the most nodes (CTSS) was Cluster 01 with 850 nodes, which includes *CSF1R* and contains genes which were upregulated after differentiation (Figure 6-26). The second most abundant cluster was Cluster 02 with 286 nodes, which shows genes mostly expressed in monocytic THP-1 cells prior to differentiation (Figure 6-26). There was little indication of an effect of *USP16* knockout on genes in these clusters.

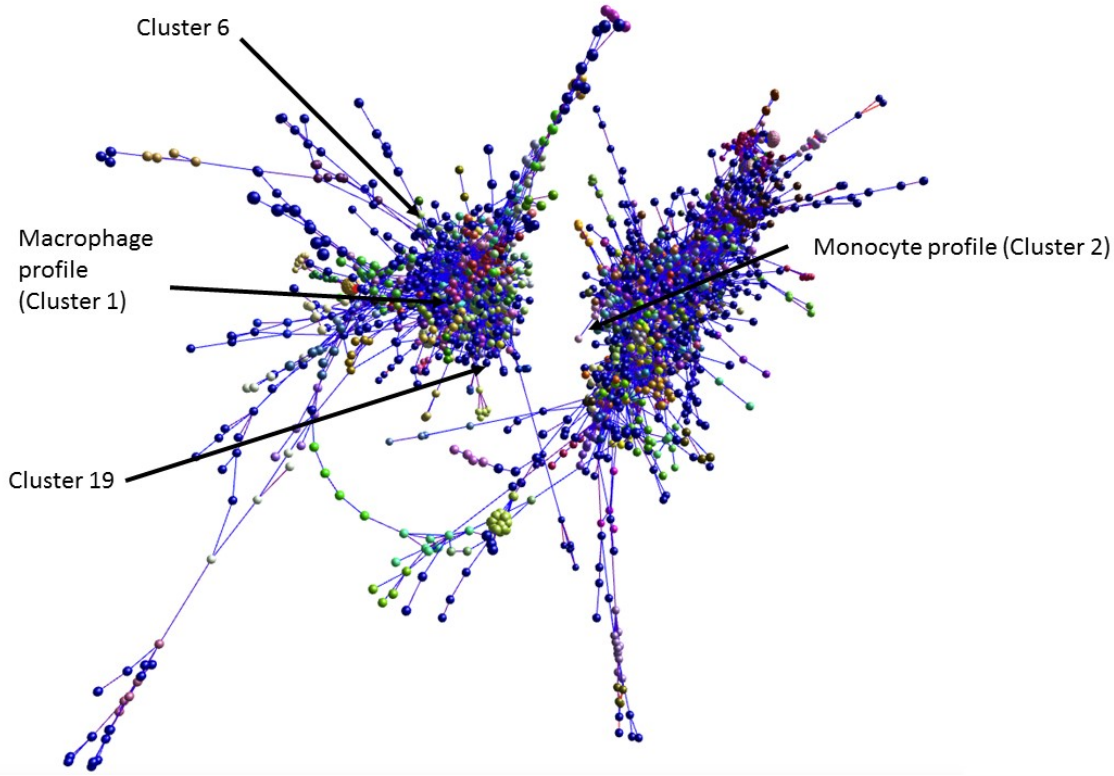

Figure 6-25 Image from Miru showing CTSS-CTSS correlation (nodes = CTSS). Including all USP16 knockout and wildtype samples analysed using a Pearson's correlation coefficient ≥ 0.92. Location of important clusters is annotated on the diagram.

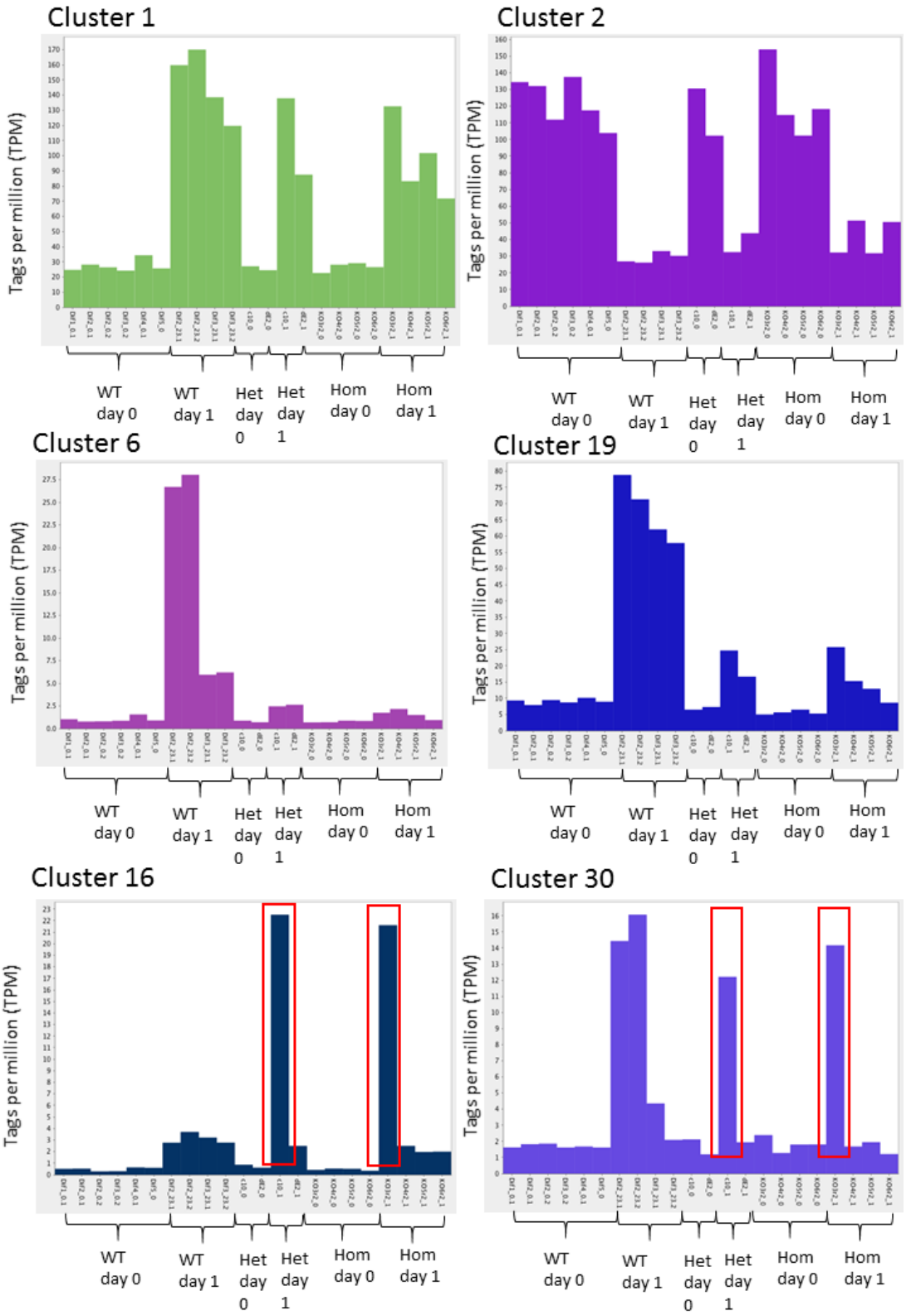

Figure 6-26 Average expression profiles of various clusters of co-expressed genes. USP16 knockout and wildtype samples only, with correlation coefficient ≥ 0.92. For clusters 16 and 30, red box denotes Hom 1 and its parental heterozygote, Het C.

Cluster 6 with 129 nodes and Cluster 19 with 44 nodes contained genes differently expressed between wildtype differentiated macrophages and *USP16* homozygous and heterozygous knockout macrophages (day 1) (Figure 6-26). However, two of the wildtype biological replicates in Cluster 6 were more highly expressed than the other two. Nevertheless, genes within these two clusters were annotated for GO terms (Table 6-8), and found to be enriched for terms involved with extracellular space, inflammatory response, intracellular signalling cascade and nucleoside binding.

Cluster 6		**GO terms**	**Group score**
	Group 1		1.98
		Extracellular space	
		Secreted	
	Group 2		1.97
		Inflammatory response	
		Wound healing	
		Chemotaxis	
	Group 3		1.01
		Positive regulation of response to external stimulus	
	Group 4		1.01
		Nucleoside binding	
Cluster 19			
	Group 1		0.9
		Intracellular signalling cascade	
	Group 2		0.78
		Cytoskeleton protein binding	

Table 6-8 GO annotation of clusters 6 and 19 by DAVID. Group score is an internal group score calculated by DAVID software, indicating the enrichment for related GO terms. A high score indicates high enrichment. GO terms with p-value smaller than 0.05 shown.

To attempt to understand the close similarity of cell cycle analysis and proliferation analysis between Hom 1 and its parental heterozygote Het C, Clusters 16 and 30 were chosen for further analysis (Figure 6-26), since these clusters contain CTSS which are exclusively expressed in differentiated Hom 1 and Het C cells (Cluster 16 with 52 nodes) or expressed only in differentiated wildtype and Hom 1 and Het C cells (Cluster 30 with 29 nodes). Wildtype biological replicates did not behave similarly in Cluster

30, as it was in the case of Cluster 6, where two replicates had different expression pattern to the other two. However, when taken together, these clusters were significantly enriched for GO terms associated with von Willebrand factor domain, integrin-mediated signalling pathway and innate immunity response (Table 6-9).

Cluster 16		**GO terms**	**Group score**
	Group 1		1.9
		von Willebrand factor (vWF) type A domain	
	Group 2		1.39
		Integrin-mediated signalling pathway	
	Group 3		0.7
		Cell surface	
Cluster 30			
	Group 1		1.7
		Innate immunity response	
	Group 2		1.79
		Extracellular exosome	
		Signal peptide	

Table 6-9 GO annotation of clusters 16 and 30 by DAVID. Group score is an internal group score by DAVID software, the higher the better. GO terms with p-value smaller than 0.05 shown.

To determine how many CTSS were differentially expressed in *USP16* knockouts compared with wildtype THP-1 cells, the EdgeR package for R was used. Table 6-10 shows the number of CTSS differentially expressed (either upregulated or downregulated) between different groups of samples, with p-value below 0.01. There were 390 CTSS differentially expressed in *USP16* homozygous monocytic THP-1s compared to wildtype, in comparison to 588 CTSS in *USP16* homozygous differentiated macrophage THP-1 versus wildtype. Overall, thousands of CTSS changed expression between differentiated and undifferentiated state in both *USP16* clones and wildtype. However, there was a higher number of differentially expressed CTSS in wildtype differentiation than in *USP16* homozygous knockouts' differentiation. Smaller numbers of CTSS with altered expression were found when comparing heterozygotes with both homozygotes and wildtype than for homozygotes

versus wildtype, indicating the possibility of dominance relationships for expression level for some genes and a potential dose response for others.

Comparison	Hom 0 - WT 0	Hom 1 - WT 1	Hom 0 - Hom 1	WT 0 - WT 1
-1	90	157	1819	2953
1	300	431	2287	6998
	Hom 0 - Het 0	**Hom 1 - Het 1**	**Het 0 - WT 0**	**Het 1 - WT 1**
-1	33	67	1	24
1	18	15	90	300

Table 6-10 Number of CTSS which were upregulated (1) or downregulated (-1) in comparison between USP16 homozygous knockouts (Hom), USP16 heterozygous knockouts (Het) or wildtype (WT) day 0 (0) or day 1 (1) with smaller p-value than 0.01 as calculated by EdgeR package.

Looking at lists of most differentially expressed CTSS (ordered by the smallest False Discovery Rate (FDR) – thus the smallest probability that the difference in expression is due to chance), THP-1 *USP16* knockout monocytes upregulated genes like *EPHA7*, *IRF2BP2* and *PGGHG* (Table 6-11). THP-1 *USP16* knockout macrophages downregulated genes like *COL6A1*, *COL6A2*, *MAGED1* and *EREG* (Table 6-12).

CTSS	logFC	logCPM	P-Value	FDR	Gene
chr6_93419269_93419683_-	5.555778	5.946459	7.05E-44	2.02E-37	EPHA7
chr6_170223475_170223593_+	2.909838	1.846875	7.10E-35	6.86E-29	?
chr8_47736696_47739149_-	1.252408	6.672056	7.20E-35	6.86E-29	CEBPD
chr20_46709535_46709742_+	-3.9987	2.345827	3.59E-32	2.57E-26	SLC2A10 (GLUT10)
chr1_234604364_234611111_-	1.148929	8.766413	3.74E-31	2.14E-25	IRF2BP2
chr11_289070_289568_+	2.333227	2.534418	2.69E-30	1.28E-24	PGGHG
chr1_145995971_145996793_-	1.372818	7.047642	5.38E-29	2.20E-23	TXNIP
chr4_15416345_15417173_-	1.939454	2.703491	5.36E-22	1.92E-16	inside C1QTNF7, probably AC098829.1 antisense transcript
chr6_170225206_170226024_-	1.969923	3.643444	2.38E-21	7.56E-16	?
chr1_146673933_146674113_+	2.000452	3.741057	7.69E-21	2.20E-15	inside HYDIN2, pseudogene

Table 6-11 EdgeR list of CTSS with greatest difference in TPM between USP16 homozygous clones and wildtype THP-1 clones in undifferentiated state at day 0 before PMA treatment. The list is ordered by the lowest False Discovery Rate (FDR) value. Genes were manually annotated. Questionmark denotes a CTSS which was not found near any gene.

CTSS	logFC	logCPM	P-Value	FDR	Gene
chrX_51802996_51803395_+	-5.7955455	1.602815	6.24E-38	1.78E-31	MAGED1
chr21_46003385_46005020_+	-3.8357612	2.801117	1.21E-34	1.73E-28	COL6A1
chr17_7251246_7252164_-	0.9849368	6.837028	1.40E-33	1.34E-27	CTDNEP1
chr5_98338656_98338731_+	2.8190013	2.374547	2.58E-32	1.85E-26	?
chr19_17377247_17377429_-	-4.6594492	1.191346	2.40E-26	1.38E-20	PLVAP
chr21_46098057_46098456_+	-2.9903802	2.984407	8.38E-25	4.00E-19	COL6A2
chr1_151458286_151459537_-	1.2490982	4.000285	1.43E-24	5.86E-19	POGZ
chr4_74364625_74365329_+	-8.0689984	1.135303	1.35E-23	4.83E-18	EREG
chr16_11585505_11587540_-	1.0172729	8.883237	1.01E-22	2.93E-17	LITAF
chr3_160224136_160227192_-	1.4789862	6.080353	1.03E-22	2.93E-17	C3orf80

Table 6-12 EdgeR list of CTSS with greatest difference in TPM between USP16 homozygous clones and Wildtype clones in differentiated state at day 1 post PMA treatment. The list is ordered by the lowest False Discovery Rate (FDR) value. Genes were manually annotated. Questionmark denotes a CTSS which was not found near any gene.

6.3.6 CRISPR editing of *BAP1, USP12, KDM6B* and *UTY* was associated with low viability after cell sorting and was not successful

Since genome editing of THP-1 cells was successful for *USP16*, attempts were then made to edit the other chromatin modifying genes. Nucleofection of THP-1 cells with low passage number was achieved each time, although with very low rate of transfection – around 2% of live cells were GFP positive. Consequently, a very small number of cells was collected each time using FACS, and these cells were single cell sorted to make clonal lineages. There were not enough cells to analyse pooled DNA to assess whether indels had been created.

Sorted single cells were cultured to create clonal lineages of THP-1 cells in the same manner as for the *USP16* knockout experiment to see if the guides created knockout cell lines. These experiments were repeated multiple times over the course of almost two years. However, when these four genes were targeted, THP-1 cells did not grow up at suitable rates after single cell sorting, and in each experiment only a few cells out of the whole plate survived and successfully created a clonal cell line.

BAP1 guide plasmids were electroporated into THP-1 cells a total of three times. The first sort had two 96-well plates and only 10 clones survived the sorting. The second time, three 96 well plates were used but none of the cells survived. The third time, again, three 96-well plates were used but only 2 clones lived. None of the surviving clones had the deletion of exon 4 (not pictured); all were wildtype.

USP12 guides were used in an attempt to knockout *USP12* in THP-1 cells total of four times over the course of 2 years. For the first two trials, only single guides were used. The first time, 6 plates were sorted with 35 clones sequenced, for the second time again 6 plates were used with 43 clones surviving. None of these clones had an indel as confirmed by Sanger sequencing. The third sort using 9 plates had no survivors, and the last sort with 6 plates had 19 clones surviving. Again, no clones had a deletion.

Both *UTY* and *KDM6B* potentially edited THP-1 cells were sorted in three batches over the course of a year. In sort 1, various combinations of *KDM6B* guides were

electroporated into THP-1 cells, and single GFP^+ cells were sorted into 96-well plates. In total, out of ten 96-well plates, only 18 single clones grew to cover the whole well. These clones were genotyped and none of the clones had an edit. For second batch, out of nine 96-well plates containing single THP-1 cells, none of the cells grew up. For the third batch, out of six 96-well plates 6 clones grew up, and none had an edited genotype. *UTY* clones had similar fate to *KDM6B* clones. For the first sort, only 16 clones survived out of ten plates and none had the edited genotype. For the second sort, no cells survived (out of 15 plates) and for the final sort only 7 clones survived, with no edits in the genome. It needs to be noted that no cells survived the second sort for either edited gene, probably due to the failure of the incubator during the week that the cells were sorted.

6.4 Discussion

6.4.1 Chapter overview

In this chapter, CRISPR-Cas9 system has been used to create mutations in five genes of chosen chromatin modifying enzymes in acute monocytic leukaemia THP-1 cells. Only one knockout, of the H2A deubiquitinase *USP16,* was successfully achieved. *USP16* knockout cells were able to differentiate into macrophages with the addition of PMA. Their ability to phagocytose particles was not disrupted in comparison with wildtype cells, but lack of USP16 did affect their cell proliferation. *USP16* knockout cells were analysed using CAGE (Cap Analysis of Gene Expression) technology. CAGE libraries were successfully created and the results show that *USP16* deficiency in THP-1 cells downregulates the levels of macrophage markers such as *CSF1R* at the same time point post PMA stimulation as wildtype. This is the first study of USP16 in monocytes and their differentiation into macrophages.

6.4.2 Efficiency and limitations of CRISPR-Cas9 system in THP-1 cells

Characterising functional roles of the three deubiquitinases and two demethylases in macrophage differentiation and leukaemia was attempted by creating knockout clones using CRISPR-Cas9 system. Bioinformatics analysis of these five genes was used to identify the important domains and gene structures and to design the guides (Chapters 4 and 5).

Following the successful creation of *USP16* knockout lines, CRISPR-Cas9 technology was used in an attempt to knockout the *USP12, BAP1, KDM6B* and *UTY* genes in THP-1 cells to see whether the absence of one or other of these genes and hence of functional protein affects cell survival and proliferation or monocyte to macrophage transition. However, no edited lines could be created for these genes.

One question is whether the guides were actually efficient in creating a double strand break. Since the percentage of GFP^+ cells in the population was very low, most of the cells were single cell sorted onto plates. Therefore, it was impossible to collect enough cells in bulk to extract DNA and do a T7 endonuclease assay or TIDE analysis

(Brinkman *et al*., 2014) both of which detect indel formation. One of the solutions, which was not completed due to time constrains of this project, would be to use an easy-to-transfect cell line to check the efficiency of the guides and select only the ones which are most likely to create a double strand break. As *UTY* is found on Y chromosome, the ideal cell line would contain Y chromosome. Human cancerous pulmonary epithelial cell line A549 was considered, as it was collected from a male patient (Giard *et al*., 1973). However, PCR of genomic regions of Y chromosome showed that the cell line found in the laboratory differed from ATCC specifications and did not contain a Y chromosome (data not shown). Therefore, it was not possible to use this line to check the guide for *UTY*. Also, it is very likely that the efficiency of guides in creating a double strand break would differ between cell types, so in the end validation of the chosen guides in THP-1 would still be needed.

It is possible that no indels were found among the small number of clonal lines that developed following CRISPR-Cas9 treatment because removal of functional enzymes was lethal to the cells. Since THP-1 cells lack *KDM6A*, the options for compensation using other demethylases are limited, and (as discussed in section 5.4.2) double knockout of *Kdm6a*$^{-}$*/Uty*$^{-}$ is embryonic lethal in mice while homozygous *Kdm6b* deletion is lethal in the perinatal period. For H2A deubiquitinases, as previously mentioned in section 4.1, knockout of *Bap1* in mice also results in embryonic death but *Usp12* knockout mice are viable. Given the severe phenotype of these whole animal knockouts, it may have been unrealistic to create edited lines. However, knockout of *Usp16* is also embryonic lethal, and yet it was possible to create a knockout cell line. This might suggest that USP16 is less important to survival of THP-1 cells than the other enzymes, although it is necessary for the survival of the whole organism. However, it clearly does have a role in the differentiation process as the transcriptomic profile of the *USP16* knockout cells was distinct from that of wildtype. Further exploration of the other chromatin modifying enzymes may require alternative approaches, including the use of conditional knockouts that could be activated at precise times during differentiation.

It was difficult to achieve a high transfection rate when THP-1 cells were transfected with a pX458 plasmid. The transfection efficiency was around 1-2% using optimized Nucleofection 4D system (Lonza), which was more efficient than using lipofection (data not shown). Using the smaller enhanced GFP plasmid included in the Nucleofection package, it was possible to achieve GFP+ rates of around 40% (not shown). Thus, the problem might be actually the size of the pX458 plasmid (9300 bp), which is quite large for transfection. One of the solutions would be using lentiviral approaches to deliver the plasmid into the cells, which has higher transfection efficiency overall. Another solution would be to nucleofect the Cas9 protein and sgRNA, and not the whole plasmid. Increasing the amount of DNA in nucleofection is not an option, as the higher amount of nucleic acid would result in higher cell death (personal email correspondence with a Lonza scientific advisor, 2016). Monocytes and THP-1 cells are able to sense cytoplasmic DNA, including the CRISPR-Cas9 plasmid through cGAS-STING axis (Paijo *et al*., 2016), and thus are sensitive to any attempts at transfection.

One of the reasons why the CRISPR-Cas9 system is less efficient in THP-1 cells is that monocytes have impaired double strand break repair, a key component in establishing an edited cell line (Bauer *et al*., 2011). Regardless of delivery system this is likely to reduce the chance of editing the genome of these cells. However, since the beginning of this project, two studies have been able to create a THP-1 knockout cell lines, for example using lentiviral approach (Goetze *et al*., 2017) or electroporation (Schmidt *et al*., 2016). Exploration of the techniques used may reveal ways of increasing the success rate to continue these experiments.

One of the biggest challenges in this project was to ensure the viability of THP-1 cells post sorting. As cell sorting at the Roslin Institute is only available in a not completely sterile environment, antibiotics must be used to prevent bacterial contamination, which might impede cell survival. As shown by Melanie Caruso, a student in the laboratory, for the mouse RAW264 macrophage cell line, monocyte and macrophage-like cell proliferation is reduced following single cell sorting (MSc thesis, Melanie Caruso, 2017) although cells grown in batches of at least ten seem to survive at normal rate.

This experiment was also repeated in this project with THP-1 cells, with similar outcomes (not shown). Conditioned media and increased percentage of FBS in the media (up to 30%) was also tried at the beginning of this project, with similar rate of cell survival as THP-1 media (not shown). However, it seems that survival of cells was better at the beginning of the project than at the end. For example, sort number 2 was devastating for all sorted cells, edited or wildtype (section 6.3.6) probably due to incubator failure few days after sorting. The number of cells surviving the later sorts was much lower than the number of cells surviving for the *USP16* or *USP12* sorts (section 6.3.6). These discrepancies cannot be accounted by anything other than the date of the sort, as the passage number of cells, nucleofection and cell sorting procedure and variety of different plasmids were kept constant. Cells during sorting were also stained by propidium iodide to prevent sorting of dead cells into the wells. The buffer which was used for resuspending the cells prior to cell sorting also had increased amount of FBS to ensure the cells' survival.

Usually, all GFP positive cells would be collected together, and their DNA would be analysed by PCR to see if the CRISPR-Cas9 guides created a double strand break. However, this was not possible with THP-1 cells due to the low cell numbers.

Therefore, further investigation of THP-1 survival post sorting and increasing the efficiency of nucleofection of the CRISPR-Cas9 parts would be necessary to properly create more edited clonal THP-1 cell lines.

6.4.3 *USP16* knockout clones in THP-1 clones

6.4.3.1 *USP16* deficiency does not prevent macrophage phagocytosis function, but it does affect cell proliferation

USP16 homozygous clones were created with one CRISPR guide targeting twice and were passaged a few times before being used for experiments. All *USP16* knockout homozygous or heterozygous cells changed their morphology and adhered to the surface with the addition of PMA. One hour treatment with zymosan-coated beads showed that *USP16* knockout macrophages phagocytosed around the same number of beads like wildtype. However, the THP-1 cells seemed to be saturated with 100 beads

per cell quite quickly. It is possible that phagocytosis might be slower in the *USP16* knockout cells than in the wildtype. A time course experiment should be carried out to confirm this.

During initial passaging, it had been noticed that the *USP16* clones seemed to grow at a slower rate than wildtype THP-1 cells. This has thus been tested experimentally by a colorimetric MTT assay, which measures metabolic activity of cells. For the MTT assay, a precise number of cells per well is required and as counting cells with haemocytometer is not precise, metabolic activity was assayed on day 0 and day 2 of the same concentration of cells and then the rate of growth over two days was calculated. Other ways of measuring cell proliferation were considered, but they either did not work properly (Cell Trace system from Thermo Scientific) or were too expensive (BrdU kit). Results showed that three out of four *USP16* clones with identical mutation in *USP16* gene had significantly smaller cell growth than wildtype cells, but not the first one (Hom 1). Even with the Hom 1, overall the rate of growth of homozygous cells was significantly slower than wildtype. The test was repeated at a later date, with higher passage number, and the homozygous clones had again significantly slower rate of proliferation than wildtype, which is in concordance with results of other cells treated with *USP16* RNAi (Joo *et al*., 2007).

As stated in the introduction, *USP16* has an important role in G2/M phase cell cycle progression (Xu *et al*., 2013) and knockdown in HeLa cells induced a cell cycle arrest resulting in fewer cells progressing to G2/M phase (Joo *et al*., 2007). To characterise the effect of *USP16* deficiency in macrophage differentiation, percentage of cells in separate cell cycle phases were determined at undifferentiated state and two days post differentiation. As mentioned in Chapter 3 section 3.2.2, THP-1 cells stop proliferating after PMA activation and most cells stay in G1 phase, with almost no cells in S-phase. *USP16* clones show variable results at different stages of their passage number with high standard deviation (Table 6-5, Table 6-6). However, three out of four homozygous clones (Hom 2, Hom 3 and Hom 4) did not have change in proportion of cells in S-phase after PMA stimulation, and stayed at around 20%, even though for wildtype and other cells this number decreased. The same happened for G2/M phase.

As discussed in Chapter 3, PMA stops the THP-1 cells at the G1 phase check point and cells do not progress into S-phase post-differentiation. These results raise an interesting question whether the cells lacking *USP16* in the experiment either did not have enough time to go through the cell cycle and stop at G1 due to slow proliferation or the lack of USP16 either does not allow full differentiation into macrophages in the set time or USP16 also has a role in G1 check point.

As with the rate of proliferation, homozygous clone 1 (Hom 1) and its parental heterozygote (Het C) had almost identical results to wildtype cells. The difference between homozygote 1 and other ones will be discussed in section 6.4.3.3.

6.4.3.2 *USP16* deficiency affects the level of differentiation markers in THP-1 macrophages

The effect of USP16 on the cell cycle is independent to USP16 regulating function in activating gene expression (Xu *et al*., 2013). Thus, it was interesting to know whether *USP16* knockout clones have different global gene expression levels to that of wildtype. Overall, hundreds of genes were differentially expressed in *USP16* homozygotes compared to wildtype samples in both day 0 and day 1, according to EdgeR differential expression analysis.

In non-differentiated THP-1 cells, one of the differentially expressed genes was *EPHA7*, which was significantly upregulated in *USP16* knockout homozygotes and one heterozygote (HetdE), but all wildtype samples had almost zero expression. *EPHA7* is also expressed by *USP16* knockout macrophages, but less than *USP16* monocytes. *EPHA7* was found to be downregulated in maturation process of mouse peripheral bone marrow monocytes (Mukai *et al*., 2017), and also not expressed in human monocytes, but present in human hematopoietic stem cells (Nguyen *et al*., 2017). Another gene significantly downregulated in *USP16* knockout THP-1 monocytes was *GLUT10* (also known as *SLC2A10*), a glucose transporter. This gene was downregulated in all homozygous clones, but not wildtype or heterozygotes. GLUT10 transports the oxidised form of Vitamin C into mitochondria and protects the cell against oxidative stress (Lee *et al*., 2010). Mutations in GLUT10 are also

associated with perturbations in TGF-β signalling (Willaert *et al*., 2012). Together, these results indicate that USP16 deficiency causes less mature monocyte phenotype and potentially increased cell stress in THP-1 monocytes.

USP16 knockout macrophages downregulate the expression of genes such as *MAGED1*, *COL6A1, COL6A2*, *PLVAP* and *EREG*. *MAGED1* was also not expressed at all in USP16 knockout monocytes, but it was expressed in one heterozygote. MAGED1 is required for cell cycle inhibition from G1 to G0 and proper muscle differentiation (Nguyen *et al*., 2010). *COL6A1* and *COL6A2* code for collagen normally required for macrophage function of modulation cell-cell interactions (Schnoor *et al*., 2008). *Plvap*-deficient mice do not produce any fetal liver monocyte-derived macrophages and is also needed for iron homeostasis (Rantakari *et al*., 2016). *EREG* is a strongly induced TLR-dependent gene, which is upregulated during macrophage stimulation with pathogens (Thuong *et al*., 2012). Together with GO terms found in Miru Clusters 6 and 19, it seems that *USP16* deficiency results in incomplete differentiation into THP-1 macrophages. This is probably due to global gene expression changes which USP16 cannot facilitate properly by deubiquitinating H2A histones so that genes which should be turned on are not.

With *USP16* knockout, *BAP1, USP12* and other deubiquitinases were found to be upregulated compared to wildtype. This raises an interesting question of whether H2A deubiquitinases can compensate for each other and thus the effect of USP16 depletion was not as severe as it would without a compensation mechanism from other deubiquitinases.

6.4.3.3 Clonal variability in *USP16* knockout cells

It is interesting to note that despite the fact all *USP16* homozygotes having the same genomic mutation, homozygous *USP16* clone 1 (Hom 1) had similar cell cycle, proliferation and gene expression profile to its parental heterozygote Het C but not to other *USP16* homozygous clones. Hom 1 had similar western blot profile to the other homozygotes, and there was no indication of producing any functional USP16 protein, so this anomaly cannot be explained by a leaky phenotype. It could be argued that it

was caused by an off-target mutation, but the *USP16* guide used in this CRISPR study had a high score based on the CRISPR guide selection tool (section 6.2.1) with 89 total possible off-targets according to the tool, with only 7 in genes. BLAST on Ensembl database was used to find other possible genomic matches of the guide, but the search only yielded one result, the *USP16* site. The highest score with a gene hit (0.9) was *ENPP5*, which is only marginally expressed in THP-1 wildtype cells and there was no difference in the gene expression between Hom 1 or Het C and other cells. The highest score with non-gene hit (2.3) is on chromosome 5 and does not appear to have any gene regulation features or SNP variants according to Ensembl.

Miru software recognised the difference in transcriptomic signature between these two related clones and the others through two clusters – Cluster 16 with 52 CTSS and Cluster 30 with 29 CTSS. Significant GO terms for these clusters included von Willebrand factor (vWF) type A domain, which is found in various glycoproteins such as collagens. GO terms also included transcriptional regulation, which was not significant (eg transcription factors SOX13 or ZNF467). This could mean that the difference in clones might be down to few genes, as differential expression of transcription factors might change expression of a vast number of genes downstream.

This difference in gene expression profile might be due to the nature of leukemic cell line; these cells are clonal and might drift from the parental cells' gene signature. Even biological replicates in FANTOM4 study were not always correlated (Figure 1-3, Figure 1-4). Mutations might arise not only by the CRISPR-Cas9 system that was used but also by the adaptation of the cancer cell line to the harsh conditions of electroporation, FACS or post-sorting single cell growth. *USP16* is a known tumour suppressor gene, which has a role in regulation of DNA damage (Zhang *et al*., 2014). It is possible that USP16 depletion in the Het C clone might have indirectly caused further accumulation of mutations. The other possibility is that the parental heterozygote of Hom 2, 3 and 4 (not included in the analysis) was the one that accumulated the mutations. The answer to this question would be answered by whole genome sequencing of these *USP16* knockout cell lines and comparing them to the wildtype. Other way how to solve this might be to look at the genes which are

differentially expressed only in all homozygotes but not heterozygotes (as for example *GLUT10* above).

6.4.4 Conclusions and future directions

Knockout of *USP16* in THP-1 cells resulted in differential expression of multiple genes, especially in macrophage function. *USP16* knockout THP-1 cells seemed to have lower levels of differentiation markers (e.g. *CSF1R*), and thus it might be hypothesized that *USP16* deficiency results in slower or in non-complete differentiation in THP-1 macrophages. To further test this, subsequent LPS stimulation of these macrophages (THP-1 monocytes which would be activated by PMA for 1 or 2 or multiple days) could be carried out, and their gene expression signature could be analysed to find out whether these *USP16* knockout macrophages are able to fully perform their function.

In several instances in the *USP16* knockouts, when the levels of one deubiquitinase was lower, gene expression of other deubiquitinases was upregulated. This has been documented previously by literature – for example *USP12* downregulation results in *USP46* upregulation (Joo *et al*., 2011). As documented in Chapter 4.4.3 there are various deubiquitinases that might share redundant function between each other. Future knockdowns or knockouts of multiple deubiquitinases might answer this question. In cases where constitutive knockouts might lead to phenotypes affecting cell viability or proliferation, novel conditional or inducible gene editing tools might still allow to dissect mechanisms of gene/protein functions (Dai *et al*., 2017, Braun *et al*., 2017).

In summary, this chapter explored the function of USP16, a deubiquitinase that has received very little study in the past. It presented the first analysis of USP16 function in the differentiation of monocytes to macrophages, as modelled by the THP-1 cell line.

Chapter 7: General discussion and future studies

7.1 Summary of findings

The main objective of this thesis was to assess the importance of chromatin modifying enzymes during cellular differentiation. It was hypothesised that enzymes removing the repressive epigenetic modifications from chromatin would be regulated, in order to open the chromatin around genes that must be expressed to trigger a transition in cellular state. Using a genome-wide technique which quantifies transcription during differentiation of a human cell line can reveal the differences in expression of these genes and help elucidate their function. In the present study, a time course of THP-1 monocytic cells differentiating into macrophage-like cells was used as a model for transition in cellular state and the changes in the transcriptome during differentiation were assessed using a technique that quantifies transcription start site usage.

The first experimental chapter (**Chapter 3**) investigated transcriptomic changes during THP-1 differentiation using CAGE technology. The study provided 18 different time points, with multiple biological replicates, which makes it the most detailed study of THP-1 differentiation to date. It was an extension of an earlier study by the FANTOM4 consortium (Suzuki *et al*., 2009), and full comparison with this study and other data from the FANTOM5 consortium still needs to be completed (Suzuki *et al*., 2009, Forrest *et al*., 2014). This vast and detailed resource of transcriptional dynamics is currently undergoing detailed analysis to describe the full scale of changes in the differentiating cells, especially to detect enhancer activation followed by target gene transcription (Arner *et al*., 2015). The analysis presented in this thesis shows examples manifesting the scope of the transcriptional activity captured by deepCAGE sequencing, including quantifying expression levels at the genome scale, detection of enhancers and identifying use of alternate promoters as differentiation progresses.

The next two chapters (**Chapter 4** and **Chapter 5**) took the information about six chromatin modifying enzymes provided by the transcriptomic analysis of THP-1 differentiation, and looked in detail at their gene architecture and other information from publicly available sources. The enzymes studied were histone deubiquitinases

BAP1, USP12 and USP16 and histone demethylases KDM6A, KDM6B and UTY. It has been more than three years since H3K27me3 demethylase activity of UTY was reported (Walport *et al*., 2014), but the role of UTY has still not been widely recognised in research. The bioinformatics analysis and literature review of KDM6A/UTY publications presented in **Chapter 5** emphasized UTY as an important enzyme with a definite role in human development.

Following the bioinformatics analysis, both chapters then used RNAi technology to reduce the expression of these genes in differentiating cells. Unfortunately, this approach did not produce acceptable results, as either the levels of the genes were not knocked down to less than 50% of original transcription (Chapter 5 section 5.3.3) or the biological replicates were variable so significant differences between untreated and treated samples could not be detected (Chapter 4 section 4.3.2).

Variable results from Chapters 4 and 5 led to a different approach to reducing the level of expression of the genes of interest. **Chapter 6** presents the use of CRISPR-Cas9 gene editing to knockout the genes of chromatin modifying enzymes. *USP16* knockout was successfully created in THP-1 cells, one of the few examples of genetic editing in THP-1 cells, which has proved to be difficult in several laboratories (Schmidt *et al*., 2016, Laugel *et al*., 2016, Goetze *et al*., 2017). This study shows the first investigation of role of USP16 in macrophages known to date.

Overall the study presented in this thesis has generated a large resource of transcriptomic data for the process of differentiation in THP-1 cells, and used these results to investigate the role of one group of enzymes during this process. The data will now be made available for other researchers interested in cell state transitions.

7.2 Methods of transcriptome-wide analysis of gene expression

The method used for genome-wide analysis of gene expression in this study was CAGE. As mentioned in Chapter 1 section 1.4.1, CAGE allows for precisely monitoring the start of the transcription (Zhao *et al*., 2011) as well as quantifying the

level of expression at the promoter and gene level. However, there are various other techniques that also analyse genome wide gene transcription.

The oldest method still in wide use is the expression microarray. Initially this technology used competitive two colour hybridisation to detect differential expression between two samples (Allison *et al*., 2006, Tarca *et al*., 2006). This technology was at best semiquatitative and has been superseded by more recent platforms where a single colour is used and many samples can be compared on the basis of expression levels. These platforms have greater sensitivity and dynamic range (Chen *et al*., 2017, Roy *et al*., 2011, Bradford *et al*., 2010) but are still based on hybridisation, in this case to oligonucleotide sequences. Hybridisation might produce large background levels which limit the accuracy for transcripts with low levels of expression (Marioni *et al*., 2008, Allison *et al*., 2006). However, the major disadvantage of expression microarrays is that they can only detect already known gene structures and do not allow for novel features (Tarca *et al*., 2006). Although sequence-based methods are now increasing in use, microarrays are still cost-effective and have the advantage of well-established robust analysis platforms such as the Affymetrix Expression and Transcriptome Consoles (ThermoFisher, 2017).

A number of approaches take advantage of high through-put sequencing. These include RNA sequencing (RNAseq) and CAGE. RNAseq was designed to sequence randomly fragmented pieces of RNA, which are then mapped to a reference genome and quantified to identify new transcript structures and determine gene expression levels (Marioni *et al*., 2008). As discussed in earlier chapters, CAGE involves sequencing short fragments from the beginning of a transcript, which are then mapped to a reference genome. Gene expression levels are comparable between RNAseq and CAGE (Kawaji *et al*., 2014). In general, CAGEseq and RNAseq are considered to be complementary technologies, with CAGEseq predominantly used for analysis of TSS positions while RNAseq is used to detect alternatively spliced transcripts. Strategies based on sequencing have the disadvantage that the analysis is computationally demanding. Although pipelines exist for the preparation and analysis of the data, such as CAGEr (Haberle *et al*., 2015), they are not yet as easy to use as the platforms for

microarrays. They also depend on the quality of the reference genome to which the reads are mapped.

Another method which can analyse TSS positions is called DeepRACE, which couples 5'RACE (rapid amplification of cDNA ends) with high-throughput sequencing (Olivarius *et al*., 2009). RACE can specifically amplify 5' ends of the transcripts (5' RACE) and thus, like CAGE, analyse TSS (Olivarius *et al*., 2009). The difference is that this technique uses gene specific primers rather than random priming. It was used to examine TSS diversity for as many as 17 genes (Olivarius *et al*., 2009). RACE-seq is another method, which combines RACE with long read RNA sequencing. Again this method is targeted to specific genes and has been used to determine the start, end and splice forms of long noncoding RNAs (Lagarde *et al*., 2016). Neither of these methods provides genome-wide transcript information, but they can be used to obtain further information about TSS and processing of transcripts of interest.

In the present study, CAGE sequencing was used for transcriptomic analysis. However, there are a number of potential issues with this technology, as outlined below.

Most current transcriptomic techniques require a high level of starting RNA. The CAGE protocol used in this study required 1-5 µg of RNA (Takahashi *et al*., 2012). This is not always available, especially if primary cells are being analysed. New protocols could be developed in this case. For example, Poulain *et al*. (2017) have published a CAGE technique with a minimal starting amount of RNA (50 ng) called nanoCAGE. Other techniques, such as RNAseq from a single cell are being rapidly developed and widely used in immune cells (Papalexi and Satija, 2017). This method, however, does not detect non-poly adenylated transcripts and might capture cell-specific anomalies which are not consistent with the features of the population (Papalexi and Satija, 2017).

Another disadvantage of CAGE is the PCR amplification step, where a high number of cycles might introduce DNA polymerase bias (Takahashi *et al*., 2012). Most CAGE

libraries in this project were amplified for 10 cycles, which was well below the recommended maximum number of 15 cycles for mammalian cells (Takahashi *et al*., 2012). The FANTOM5 project initially used the Helicos single molecule sequencing technology to develop HeliScope CAGE, which did not require an amplification step (Kanamori-Katayama *et al*., 2011). Unfortunately, due to the failure of the company Helicon Biosciences Corporation, this technology is no longer available and more recent CAGE analysis has been performed using the Illumina platform, as in this thesis. In future, a different method of CAGE library preparation without the PCR amplification step could be used, for example nAnT-iCAGE (Murata *et al*., 2014).

CAGE libraries are known to add a G nucleotide at the 5'end of the transcript due to the first reverse transcriptase activity (Figure 2-1) (Zhao *et al*., 2011). This G addition bias was corrected post-mapping stage using CAGEr (Haberle *et al*., 2015). However, efforts could be made to remove this G addition before mapping step in the future, in case it created an imprecise mapping of the sequenced tags.

In summary, CAGE sequencing was chosen as the best method for analysis of 94 samples for least cost. Using this approach enabled comparison with previous FANTOM data and TSS information. Every possible precaution was taken to minimise the biases known to happen in CAGE library preparation and sequencing.

7.3 Chromatin modifying enzymes in THP-1 differentiation

To assess the importance of the different chromatin modifying enzymes during differentiation, attempts were made to perturb the expression of the enzymes of interest, but showed disappointing results. The experiments described in Chapters 4 and 5 showed variability in the level of siRNA knock down of the chromatin modifying enzymes and in the response to the inhibitors. Similarly, the use of gene editing technology to create inactivating mutations in the genes produced only one successful knockout, of *USP16* (Chapter 6).

One of the possible reasons why the results were so variable is that the THP-1 cells had been subjected to genetic drift. Cells in culture for multiple passages can exhibit

reduction in key functions (Hughes *et al*., 2007). To minimise this problem, THP-1 cells in this thesis were not used for more than 7 passages, and cells for quantitative experiments were always at low passage number (from the time of single cell cloning to produce Clone 5), usually p10 or p11. THP-1 is a tumour cell line, which may have changes in mutation detection and DNA repair systems, known to be associated with tumourigenesis. Although the karyotype is relatively normal, there are a number of known mutations which may predispose to further genetic change after time in culture.

As mentioned in section 6.4, the failure to produce null mutations for most of the enzymes using CRISPR-Cas9 editing may indicate that lack of these enzymes was lethal to cells, so that the genome edited clones failed to survive. Single cells after cell sorting are subjected to selective pressure, which might also explain the differences in phenotype of the *USP16* clones (Hughes *et al*., 2007). The clonogenic potential of THP-1 cells is controlled at least in part by members of the polycomb complexes (Huang *et al*., 2014), and knockout of these genes encoding enzymes that reverse polycomb marks may have ablated their ability to form colonies from single cells.

In spite of the difficulties in experimentally altering the levels of the enzymes of interest, this study provides some data that elucidates the role of chromatin modification in this model system of differentiation. The results in this study show that there are at least three different transcription patterns found in these chromatin modifying enzymes. Firstly, mRNA for two genes, *USP12* and *KDM6B*, increased fairly early in the differentiation, with peaks of expression few hours after PMA stimulation. Secondly, there were cyclical patterns of expression. For example, *BAP1* mRNA peaked every 3-4 hours during differentiation. Thirdly, there were genes which were constitutively expressed with little fluctuation during differentiation (*USP16* and to some extent, *UTY*).

It could be speculated that enzymes such as KDM6B have specific gene targets which need to be activated first. These could include genes in Cluster 6 (section 3.3.3) which increase a couple of hours after the *KDM6B* peak. This group included genes involved in transcriptional regulation (Table 3-1), suggesting that KDM6B may trigger a wave of transcription factor production which would in turn activate other genes involved

in the differentiation process. *USP12*, which increases slightly later, may even be a target of KDM6B. USP12 may then have gene targets which are transcribed later in the differentiation process. These could include genes in Cluster 3, which were associated with lysosomes and endoplasmic reticulum.

Oscillating enzymes such as BAP1 might be required periodically, as waves of repression and activation of genes progress during cell differentiation. Both Cluster 15 (containing *BAP1*) and Cluster 36 (containing *USP12*) show a cyclical pattern, although this was so not evident in the expression of *USP12* itself (Figure 3-32). Both clusters were associated with terms relating to ubiquitin, suggesting that adding and removing ubiquitin are important regulated functions during this differentiation.

Then there were constitutively expressed enzymes (such as USP16) which might be required for basal expression of a wide range of genes and potentially function to keep the chromatin poised for transcription of key genes. Interestingly, *USP16* was the only gene which was successfully knocked out by CRISPR-Cas9 technology. The cells lacking a functional copy of this enzyme showed reduced proliferation and an altered transcriptome after differentiation, including reduction in macrophage markers and intra and extracellular signalling (as discussed in Chapter 6 section 6.4.3.2), indicating possible targets for this enzyme.

7.4 Additional areas of further investigation

This study focussed on transcription as the driver for phenotypic change, but it is also important to ascertain whether the protein products are made and activated. Various post-translational modifications alter protein function. For example, as mentioned in section 1.2.2.2, USP12 alters the quantity of NOTCH receptor by deubiquitination (Moretti *et al*., 2012), which cannot be detected by mRNA-level gene expression studies. Further proteomic analysis in THP-1 cells should be carried out and compared to, for example THP-1 kinome analysis of differentiation (Richter *et al*., 2016) or THP-1 cells' response to LPS using mass spectrometry (Tarasova *et al*., 2015). Additionally, protein levels of the chromatin modifying enzymes should be assessed by semi-quantitative western blotting and protein modifications by mass spectrometry to ensure the levels of the proteins in the cell increase in parallel with the mRNA during

differentiation and whether there are specific post-translational modifications potentially preventing their function.

As mentioned in the Introduction (section 1.4.1), the FANTOM4 study was repeated here with the same conditions in order to be able to compare the datasets. In particular, the concentration of PMA that triggered the THP-1 response was the same, as altering the PMA concentration might have resulted in different dynamics of differentiation (Daigneault *et al*., 2010, Suzuki *et al*., 2009). Other culture conditions such as high or low confluence of the cells can also alter the way THP-1 cells respond to PMA (Aldo *et al*., 2013). This highlights the need for detailed description of the experimental techniques in biological research, especially working with cell lines that have been in culture for decades, to increase the reproducibility of scientific data. This anomaly in THP-1 cells could be further assessed by repeating the experiments in this thesis with changed culture conditions and different PMA concentrations.

As discussed in Chapter 3 (section 3.4.6), this thesis did not completely address the macrophage markers and macrophage phenotype in the THP-1 differentiation study. Further investigation is needed, for example looking at networks of transcription factor up- or downregulation and subsequent changes in expression of their target genes. Comparing publicly available chromatin data (for example from the ENCODE project) with this dataset would be a non-expensive way to link chromatin modifying enzymes with their target modification. Comparing epigenetic marks of enhancers with the bidirectional promoter data in this study is beneficial for finding new important regulatory sequences which might cause leukaemia. For example, Zhang *et al*. (2016) identified copy number gains of noncoding regions including super enhancers near *USP12* in colorectal carcinoma, which probably represents a common mechanism to increase tumorigenicity in cells.

The THP-1 cell line is widely used in studying macrophage inflammatory reactions against various stimulants, such as pathogens and injury (Kim *et al*., 2016, Ma *et al*., 2015), but also in inflammatory diseases, for example in vasculature (Qin, 2012, Borst *et al*., 2015). In the present study, the response of the cells to pathogens (or the

pathogen associated molecule LPS) or to other inflammatory signals was not tested in the normal or knockout cells. The *USP16* knockout cell line is now a resource for further exploration of the intersection between chromatin modification and response to stimulation. Similarly, the extensive CAGE data generated by this project could be used to compare the THP-1 derived macrophage-like cells with tissue and bone marrow derived macrophages. Focusing on the macrophage-specific gene signatures in the CAGE dataset and their analysis might be of interest to a wide macrophage scientific community.

This thesis primarily outlines the role of a limited number of chromatin modifying enzymes. However, other enzymes, for example enzymes of PRC1 or PRC2 such as EZH2 or RING1B, have also been implicated in cancer and other diseases (Brazel and Vernimmen, 2016). As these enzymes directly facilitate the creation of the marks that BAP1, USP12, USP16 or KDM6A, KDM6B and UTY remove, the result of inhibiting their function could be opposite. However, as a large number of genes are downregulated in THP-1 differentiation (section 3.3.3, Cluster 1), creating repressive chromatin marks is also required for cell differentiation. Further research into other components of PRC1 and PRC2 could show similar results of reduced differentiation potential similar to the THP-1 *USP16* clones. A wide variety of inhibitors of these enzymes such as EZH2 inhibitor, which had been developed as a possible treatment for cancer, could be used (Stazi *et al*., 2017, Brazel and Vernimmen, 2016).

As mentioned in the introduction section 1.2.2, other USPs deubiquitinating H2A have also been implicated in cancer. Since the beginning of the project, MYSM1 was found to be essential for foetal liver and adult bone marrow haematopoiesis (Forster *et al*., 2015) and USP22 was found to stimulate breast cancer growth by positively regulating cMYC (Kim *et al*., 2017). It is highly probable that these two proteins also affect the THP-1 differentiation and their effect could be investigated in future studies.

The *USP16* knockout cells have a defective cell cycle arrest during differentiation, and reduced proliferation (section 6.3.4). The THP-1 line used here was a clone selected for high differentiation potential (Chapter 1 section 1.3.2) (Suzuki *et al*., 2009). The results for the *USP16* knockout in THP-1 Clone 5 could be compared the stock THP-

1 line available from the ATCC. The ATCC cell line could also be recloned to get clones with low and high differentiation response measured by adherence, and which could be compared with Clone 5 and the *USP16* knockout clones. Such clones could also be used for further exploration of the role of other chromatin modifying enzymes in proliferation.

This knowledge that at least one of the enzymes is involved in regulating cell division could be exploited for cancer treatment, where the goal is to reduce cell proliferation. There are as yet no inhibitors specifically against USP16, which could be potentially used in cancer therapy. CAGE technology can also be used to analyse functional features of the genome, such as enhancers (as discussed in section 3.4.7), and it would be beneficial to fully explore the differential transcriptome in USP16 knockout cells. As chromatin state affects complex processes such as promoter shifting and enhancer usage, these features could have also been affected in these cells, and may indicate targets for drug design.

These chromatin modifying enzymes affect the chromatin marks, but no assessment of these marks was carried out yet in this work. ChIPseq of various chromatin marks like H3K27me3, H3K4me3, H2Aub and H2Bub could be carried out throughout the differentiation time course, (such as in Prasad *et al*. (2014)). CAGEseq data from this study could be further validated by DNaseI sequencing, which identifies open chromatin regions (Boyle *et al*., 2008).

Interestingly, tumour suppressor genes that escape X inactivation like *KDM6A* contribute to cancer sex bias, as men carry higher risk of cancer than woman (Dunford *et al*., 2017). Other chromatin modifying enzymes, such as KDM5C and USP9X with genes located on the X chromosome, also have Y paralogues - *KDM5D* and *USP9Y* (section 5.4.2.3) (Arseneault *et al*., 2017, Hall *et al*., 2003). Full investigation into all XY chromatin modifying pairs might be beneficial for future differential gender-dependent cancer treatments, as they are all implicated in tumorigenesis (Dunford *et al*., 2017).

A full transcriptomic time course of primary cells, separately for healthy female and male donors, could suggest whether THP-1 is the right model to study the differentiation. Primary monocytes from donors could also be targeted by CRISPR-Cas9 lentivirus to validate the *USP16* knockout data in differentiating macrophages. With the rise of personalised medicine, deep transcriptomic sequencing such as that presented in this study could in future become the norm for patients with leukaemia. For now, other leukaemia cell lines (such as U-937 and HL-60) and samples from leukaemia patients could be used to develop precise TSS monitoring in comparison with THP-1 and wildtype cells to determine the best model for the analysis of monocytic leukaemia and development of potential therapies.

7.5 Conclusions

We live in an era of big data, where the cost of sequencing is still going down and genome-wide transcriptional datasets are appearing at high rate (Schmidt and Hildebrandt, 2017). Transcriptional analysis is not only performed in human or mouse, as the main models for disease, but also other animals such as sheep, pig and dog (Freeman *et al*., 2012, Clark *et al*., 2017, Markby *et al*., 2017). New genome assemblies for various animals are available with more and more paralogues of human genes annotated, through which we can observe evolutionary constraints over many years (Yates *et al*., 2016, Lesch *et al*., 2016). The question is now whether we are able to interpret the data into a physiologically significant meaning. Studying potential functional roles of genes, such as was shown in this thesis, is one of the many ways to achieve this.

In summary, results presented in this thesis provide a resource to study transcriptional changes in cell differentiation, particularly in macrophages. Throughout the whole thesis an acute monocytic leukaemia cell line was used as an experimental model, which might help in future therapeutic treatments of cancer. This study highlights the difficulties of certain experimental techniques (e.g. siRNA mRNA knock down and CRISPR-Cas9 genome editing) used in a haematopoietic cell line, but also reveals new data about chromatin modifying enzymes in THP-1 differentiation. The initial hypothesis that chromatin modifying enzymes are important for the differentiation of

THP-1 cells was supported, at least for the deubiquitinase USP16, and further studies using similar models and primary cells should extend these results and indicate potential treatments for acute monocytic leukaemia.

Chapter 8: Appendices

Appendices 8.1- 8.9 are in files contained in CD on the inside back cover of this thesis.

8.1 List of oligonucleotide sequences in CAGE

8.2 Split_library.pl programme used in sorting sequences

8.3 CAGE samples summary

8.4 Miru clusters gene names for Chapter 3

8.5 Miru clusters gene names for Chapter 4

8.6 FANTOM5 data for *KDM6A* and *UTY* for Chapter 5

Data are in TPM (Tags per million) for all samples in FANTOM5.

8.7 Ensembl KDM6A, KDM6B and UTY homology for Chapter 5

8.8 Transcription factors and protein-protein interactions of KDM6A, KDM6B and UTY for Chapter 5

8.9 Miru clusters gene names for Chapter 6

8.10 Instructions for accessing normalised data in Zenbu

Normalised results were deposited into Zenbu web browser (Severin *et al*., 2014). There is a collaboration project called Thesis, which contains two views – one is called Chapter 3 and the other called Chapter4_6 containing data for both siRNA treatment and KO samples, together with relevant samples from Chapter 3. Views are accessible from 'data explorer' tab when selecting Thesis collaboration.

References

ABDEL-RAHMAN, M. H., PILARSKI, R., CEBULLA, C. M., MASSENGILL, J. B., CHRISTOPHER, B. N., BORU, G., HOVLAND, P. & DAVIDORF, F. H. 2011. Germline BAP1 mutation predisposes to uveal melanoma, lung adenocarcinoma, meningioma, and other cancers. *J Med Genet,* 48**,** 856-9.

ABDEL-WAHAB, O., ADLI, M., LAFAVE, L. M., GAO, J., HRICIK, T., SHIH, A. H., PANDEY, S., PATEL, J. P., CHUNG, Y. R., KOCHE, R., PERNA, F., ZHAO, X., TAYLOR, J. E., PARK, C. Y., CARROLL, M., MELNICK, A., NIMER, S. D., JAFFE, J. D., AIFANTIS, I., BERNSTEIN, B. E. & LEVINE, R. L. 2012. ASXL1 mutations promote myeloid transformation through loss of PRC2-mediated gene repression. *Cancer Cell,* 22**,** 180-93.

ABDEL-WAHAB, O. & DEY, A. 2013. The ASXL-BAP1 axis: new factors in myelopoiesis, cancer and epigenetics. *Leukemia,* 27**,** 10-5.

ADATI, N., HUANG, M. C., SUZUKI, T., SUZUKI, H. & KOJIMA, T. 2009. High-resolution analysis of aberrant regions in autosomal chromosomes in human leukemia THP-1 cell line. *BMC Res Notes,* 2**,** 153.

ADORNO, M., SIKANDAR, S., MITRA, S. S., KUO, A., NICOLIS DI ROBILANT, B., HARO-ACOSTA, V., OUADAH, Y., QUARTA, M., RODRIGUEZ, J., QIAN, D., REDDY, V. M., CHESHIER, S., GARNER, C. C. & CLARKE, M. F. 2013. Usp16 contributes to somatic stem-cell defects in Down's syndrome. *Nature,* 501**,** 380-4.

AHN, J., KIM, K. H., PARK, S., AHN, Y. H., KIM, H. Y., YOON, H., LEE, J. H., BANG, D. & LEE, D. H. 2016. Target sequencing and CRISPR/Cas editing reveal simultaneous loss of UTX and UTY in urothelial bladder cancer. *Oncotarget,* 7**,** 63252-63260.

ALBERTS, B., BRAY, D., HOPKIN, K., JOHNSON, A., LEWIS, J., RAFF, M., ROBERTS, K. & WALTER, P. 2013. *Essential Cell Biology, Fourth Edition*, Taylor & Francis Group.

ALBIG, W., KIOSCHIS, P., POUSTKA, A., MEERGANS, K. & DOENECKE, D. 1997. Human histone gene organization: nonregular arrangement within a large cluster. *Genomics,* 40**,** 314-22.

ALDO, P. B., CRAVEIRO, V., GULLER, S. & MOR, G. 2013. Effect of culture conditions on the phenotype of THP-1 monocyte cell line. *Am J Reprod Immunol,* 70**,** 80-6.

ALLIS, C. D., BERGER, S. L., COTE, J., DENT, S., JENUWIEN, T., KOUZARIDES, T., PILLUS, L., REINBERG, D., SHI, Y., SHIEKHATTAR, R., SHILATIFARD, A., WORKMAN, J. & ZHANG, Y. 2007. New nomenclature for chromatin-modifying enzymes. *Cell,* 131**,** 633-6.

ALLISON, D. B., CUI, X., PAGE, G. P. & SABRIPOUR, M. 2006. Microarray data analysis: from disarray to consolidation and consensus. *Nat Rev Genet,* 7**,** 55-65.

ALTIERI, D. C. & EDGINGTON, T. S. 1988. The saturable high affinity association of factor X to ADP-stimulated monocytes defines a novel function of the Mac-1 receptor. *J Biol Chem,* 263**,** 7007-15.

ANDERSSON, R., GEBHARD, C., MIGUEL-ESCALADA, I., HOOF, I., BORNHOLDT, J., BOYD, M., CHEN, Y., ZHAO, X., SCHMIDL, C., SUZUKI, T., NTINI, E., ARNER, E., VALEN, E., LI, K., SCHWARZFISCHER, L., GLATZ, D., RAITHEL, J., LILJE, B., RAPIN, N., BAGGER, F. O., JORGENSEN, M., ANDERSEN, P. R., BERTIN, N., RACKHAM, O., BURROUGHS, A. M., BAILLIE, J. K., ISHIZU, Y., SHIMIZU, Y., FURUHATA, E., MAEDA, S., NEGISHI, Y., MUNGALL, C. J., MEEHAN, T. F., LASSMANN, T., ITOH, M., KAWAJI, H., KONDO, N., KAWAI, J., LENNARTSSON, A., DAUB, C. O., HEUTINK, P., HUME, D. A., JENSEN, T. H., SUZUKI, H., HAYASHIZAKI, Y., MULLER, F., FORREST, A. R. R., CARNINCI,

P., REHLI, M. & SANDELIN, A. 2014. An atlas of active enhancers across human cell types and tissues. *Nature,* 507**,** 455-461.

ANDREWS, S. 2010. *FastQC: a quality control tool for high throughput sequence data* [Online]. Available: http://www.bioinformatics.babraham.ac.uk/projects/fastqc [Accessed].

ANZAI, N., LEE, Y., YOUN, B. S., FUKUDA, S., KIM, Y. J., MANTEL, C., AKASHI, M. & BROXMEYER, H. E. 2002. C-kit associated with the transmembrane 4 superfamily proteins constitutes a functionally distinct subunit in human hematopoietic progenitors. *Blood,* 99**,** 4413-21.

ARNER, E., DAUB, C. O., VITTING-SEERUP, K., ANDERSSON, R., LILJE, B., DRABLOS, F., LENNARTSSON, A., RONNERBLAD, M., HRYDZIUSZKO, O., VITEZIC, M., FREEMAN, T. C., ALHENDI, A. M., ARNER, P., AXTON, R., BAILLIE, J. K., BECKHOUSE, A., BODEGA, B., BRIGGS, J., BROMBACHER, F., DAVIS, M., DETMAR, M., EHRLUND, A., ENDOH, M., ESLAMI, A., FAGIOLINI, M., FAIRBAIRN, L., FAULKNER, G. J., FERRAI, C., FISHER, M. E., FORRESTER, L., GOLDOWITZ, D., GULER, R., HA, T., HARA, M., HERLYN, M., IKAWA, T., KAI, C., KAWAMOTO, H., KHACHIGIAN, L. M., KLINKEN, S. P., KOJIMA, S., KOSEKI, H., KLEIN, S., MEJHERT, N., MIYAGUCHI, K., MIZUNO, Y., MORIMOTO, M., MORRIS, K. J., MUMMERY, C., NAKACHI, Y., OGISHIMA, S., OKADA-HATAKEYAMA, M., OKAZAKI, Y., ORLANDO, V., OVCHINNIKOV, D., PASSIER, R., PATRIKAKIS, M., POMBO, A., QIN, X. Y., ROY, S., SATO, H., SAVVI, S., SAXENA, A., SCHWEGMANN, A., SUGIYAMA, D., SWOBODA, R., TANAKA, H., TOMOIU, A., WINTERINGHAM, L. N., WOLVETANG, E., YANAGI-MIZUOCHI, C., YONEDA, M., ZABIEROWSKI, S., ZHANG, P., ABUGESSAISA, I., BERTIN, N., DIEHL, A. D., FUKUDA, S., FURUNO, M., HARSHBARGER, J., HASEGAWA, A., HORI, F., ISHIKAWA-KATO, S., ISHIZU, Y., ITOH, M., KAWASHIMA, T., KOJIMA, M., KONDO, N., LIZIO, M., MEEHAN, T. F., MUNGALL, C. J., MURATA, M., NISHIYORI-SUEKI, H., SAHIN, S., NAGAO-SATO, S., SEVERIN, J., DE HOON, M. J., KAWAI, J., KASUKAWA, T., LASSMANN, T., *et al*. 2015. Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells. *Science,* 347**,** 1010-4.

ARSENEAULT, M., MONLONG, J., VASUDEV, N. S., LASKAR, R. S., SAFISAMGHABADI, M., HARNDEN, P., EGEVAD, L., NOURBEHESHT, N., PANICHNANTAKUL, P., HOLCATOVA, I., BRISUDA, A., JANOUT, V., KOLLAROVA, H., FORETOVA, L., NAVRATILOVA, M., MATES, D., JINGA, V., ZARIDZE, D., MUKERIA, A., JANDAGHI, P., BRENNAN, P., BRAZMA, A., TOST, J., SCELO, G., BANKS, R. E., LATHROP, M., BOURQUE, G. & RIAZALHOSSEINI, Y. 2017. Loss of chromosome Y leads to down regulation of KDM5D and KDM6C epigenetic modifiers in clear cell renal cell carcinoma. *Sci Rep,* 7**,** 44876.

AYOUBI, T. A. & VAN DE VEN, W. J. 1996. Regulation of gene expression by alternative promoters. *FASEB J,* 10**,** 453-60.

AYUSO, P., BLANCA, M., CORNEJO-GARCIA, J. A., TORRES, M. J., DONA, I., SALAS, M., BLANCA-LOPEZ, N., CANTO, G., RONDON, C., CAMPO, P., LAGUNA, J. J., FERNANDEZ, J., MARTINEZ, C., AGUNDEZ, J. A. & GARCIA-MARTIN, E. 2013. Variability in histamine receptor genes HRH1, HRH2 and HRH4 in patients with hypersensitivity to NSAIDs. *Pharmacogenomics,* 14**,** 1871-8.

BAILLIE, J. K., ARNER, E., DAUB, C., DE HOON, M., ITOH, M., KAWAJI, H., LASSMANN, T., CARNINCI, P., FORREST, A. R., HAYASHIZAKI, Y., CONSORTIUM, F., FAULKNER, G. J., WELLS, C. A., REHLI, M., PAVLI, P., SUMMERS, K. M. & HUME, D. A. 2017. Analysis of the human monocyte-derived

macrophage transcriptome and response to lipopolysaccharide provides new insights into genetic aetiology of inflammatory bowel disease. *PLoS Genet,* 13**,** e1006641.
BANNISTER, A. J. & KOUZARIDES, T. 2011. Regulation of chromatin by histone modifications. *Cell Res,* 21**,** 381-95.
BATTAGLIA, A. 2014. The Importance of Multidisciplinary Approach in Early Detection of BAP1 Tumor Predisposition Syndrome: Clinical Management and Risk Assessment. *Clin Med Insights Oncol,* 8**,** 37-47.
BAUER, M., GOLDSTEIN, M., CHRISTMANN, M., BECKER, H., HEYLMANN, D. & KAINA, B. 2011. Human monocytes are severely impaired in base and DNA double-strand break repair that renders them vulnerable to oxidative stress. *Proc Natl Acad Sci U S A,* 108**,** 21105-10.
BAZIL, V., HOREJSI, V., HILGERT, I. & MCMICHAEL, M. J. 1982. Leucocyte Typing III. White Cell Differentiation Antigens. The Workshop: Myeloid Panel Antibodies Recognizing the 53-kDa Molecular Weight Monocyte Antigen (CD14). *Oxford. Oxford University Press***,** 611–613.
BELLE, J. I. & NIJNIK, A. 2014. H2A-DUBbing the mammalian epigenome: expanding frontiers for histone H2A deubiquitinating enzymes in cell biology and physiology. *Int J Biochem Cell Biol,* 50**,** 161-74.
BELLOTT, D. W., HUGHES, J. F., SKALETSKY, H., BROWN, L. G., PYNTIKOVA, T., CHO, T. J., KOUTSEVA, N., ZAGHLUL, S., GRAVES, T., ROCK, S., KREMITZKI, C., FULTON, R. S., DUGAN, S., DING, Y., MORTON, D., KHAN, Z., LEWIS, L., BUHAY, C., WANG, Q., WATT, J., HOLDER, M., LEE, S., NAZARETH, L., ALFOLDI, J., ROZEN, S., MUZNY, D. M., WARREN, W. C., GIBBS, R. A., WILSON, R. K. & PAGE, D. C. 2014. Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators. *Nature,* 508**,** 494-9.
BERGER, S. L., KOUZARIDES, T., SHIEKHATTAR, R. & SHILATIFARD, A. 2009. An operational definition of epigenetics. *Genes Dev,* 23**,** 781-3.
BILANGES, B., VARRAULT, A., MAZUMDAR, A., PANTALONI, C., HOFFMANN, A., BOCKAERT, J., SPENGLER, D. & JOURNOT, L. 2001. Alternative splicing of the imprinted candidate tumor suppressor gene ZAC regulates its antiproliferative and DNA binding activities. *Oncogene,* 20**,** 1246-53.
BLACKLEDGE, N. P., FARCAS, A. M., KONDO, T., KING, H. W., MCGOURAN, J. F., HANSSEN, L. L., ITO, S., COOPER, S., KONDO, K., KOSEKI, Y., ISHIKURA, T., LONG, H. K., SHEAHAN, T. W., BROCKDORFF, N., KESSLER, B. M., KOSEKI, H. & KLOSE, R. J. 2014. Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation. *Cell,* 157**,** 1445-59.
BOGERSHAUSEN, N., GATINOIS, V., RIEHMER, V., KAYSERILI, H., BECKER, J., THOENES, M., SIMSEK-KIPER, P. O., BARAT-HOUARI, M., ELCIOGLU, N. H., WIECZOREK, D., TINSCHERT, S., SARRABAY, G., STROM, T. M., FABRE, A., BAYNAM, G., SANCHEZ, E., NURNBERG, G., ALTUNOGLU, U., CAPRI, Y., ISIDOR, B., LACOMBE, D., CORSINI, C., CORMIER-DAIRE, V., SANLAVILLE, D., GIULIANO, F., LE QUAN SANG, K. H., KAYIRANGWA, H., NURNBERG, P., MEITINGER, T., BODUROGLU, K., ZOLL, B., LYONNET, S., TZSCHACH, A., VERLOES, A., DI DONATO, N., TOUITOU, I., NETZER, C., LI, Y., GENEVIEVE, D., YIGIT, G. & WOLLNIK, B. 2016. Mutation Update for Kabuki Syndrome Genes KMT2D and KDM6A and Further Delineation of X-Linked Kabuki Syndrome Subtype 2. *Hum Mutat,* 37**,** 847-64.
BOJJIREDDY, N., GUZMAN-HERNANDEZ, M. L., REINHARD, N. R., JOVIC, M. & BALLA, T. 2015. EFR3s are palmitoylated plasma membrane proteins that control responsiveness to G-protein-coupled receptors. *J Cell Sci,* 128**,** 118-28.
BORST, O., SCHAUB, M., WALKER, B., SCHMID, E., MUNZER, P., VOELKL, J., ALESUTAN, I., RODRIGUEZ, J. M., VOGEL, S., SCHOENBERGER, T., METZGER, K., RATH, D., UMBACH, A., KUHL, D., MULLER, II, SEIZER, P.,

GEISLER, T., GAWAZ, M. & LANG, F. 2015. Pivotal role of serum- and glucocorticoid-inducible kinase 1 in vascular inflammation and atherogenesis. *Arterioscler Thromb Vasc Biol,* 35**,** 547-57.

BOSSHART, H. & HEINZELMANN, M. 2016. THP-1 cells as a model for human monocytes. *Ann Transl Med,* 4**,** 438.

BOYLE, A. P., DAVIS, S., SHULHA, H. P., MELTZER, P., MARGULIES, E. H., WENG, Z., FUREY, T. S. & CRAWFORD, G. E. 2008. High-resolution mapping and characterization of open chromatin across the genome. *Cell,* 132**,** 311-22.

BRADFORD, J. R., HEY, Y., YATES, T., LI, Y., PEPPER, S. D. & MILLER, C. J. 2010. A comparison of massively parallel nucleotide sequencing with oligonucleotide microarrays for global transcription profiling. *BMC Genomics,* 11**,** 282.

BRAUN, S. M. G., KIRKLAND, J. G., CHORY, E. J., HUSMANN, D., CALARCO, J. P. & CRABTREE, G. R. 2017. Rapid and reversible epigenome editing by endogenous chromatin regulators. *Nat Commun,* 8**,** 560.

BRAZEL, A. J. & VERNIMMEN, D. 2016. The complexity of epigenetic diseases. *J Pathol,* 238**,** 333-44.

BRINKMAN, E. K., CHEN, T., AMENDOLA, M. & VAN STEENSEL, B. 2014. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res,* 42**,** e168.

BURGOLD, T., VOITURON, N., CAGANOVA, M., TRIPATHI, P. P., MENUET, C., TUSI, B. K., SPREAFICO, F., BEVENGUT, M., GESTREAU, C., BUONTEMPO, S., SIMEONE, A., KRUIDENIER, L., NATOLI, G., CASOLA, S., HILAIRE, G. & TESTA, G. 2012. The H3K27 demethylase JMJD3 is required for maintenance of the embryonic respiratory neuronal network, neonatal breathing, and survival. *Cell Rep,* 2**,** 1244-58.

BURSKA, U. L., HARLE, V. J., COFFEY, K., DARBY, S., RAMSEY, H., O'NEILL, D., LOGAN, I. R., GAUGHAN, L. & ROBSON, C. N. 2013. Deubiquitinating enzyme Usp12 is a novel co-activator of the androgen receptor. *J Biol Chem,* 288**,** 32641-50.

CAI, S. Y., BABBITT, R. W. & MARCHESI, V. T. 1999. A mutant deubiquitinating enzyme (Ubp-M) associates with mitotic chromosomes and blocks cell division. *Proc Natl Acad Sci U S A,* 96**,** 2828-33.

CALO, E. & WYSOCKA, J. 2013. Modification of enhancer chromatin: what, how, and why? *Mol Cell,* 49**,** 825-37.

CARNINCI, P., KVAM, C., KITAMURA, A., OHSUMI, T., OKAZAKI, Y., ITOH, M., KAMIYA, M., SHIBATA, K., SASAKI, N., IZAWA, M., MURAMATSU, M., HAYASHIZAKI, Y. & SCHNEIDER, C. 1996. High-efficiency full-length cDNA cloning by biotinylated CAP trapper. *Genomics,* 37**,** 327-36.

CARNINCI, P., SANDELIN, A., LENHARD, B., KATAYAMA, S., SHIMOKAWA, K., PONJAVIC, J., SEMPLE, C. A., TAYLOR, M. S., ENGSTROM, P. G., FRITH, M. C., FORREST, A. R., ALKEMA, W. B., TAN, S. L., PLESSY, C., KODZIUS, R., RAVASI, T., KASUKAWA, T., FUKUDA, S., KANAMORI-KATAYAMA, M., KITAZUME, Y., KAWAJI, H., KAI, C., NAKAMURA, M., KONNO, H., NAKANO, K., MOTTAGUI-TABAR, S., ARNER, P., CHESI, A., GUSTINCICH, S., PERSICHETTI, F., SUZUKI, H., GRIMMOND, S. M., WELLS, C. A., ORLANDO, V., WAHLESTEDT, C., LIU, E. T., HARBERS, M., KAWAI, J., BAJIC, V. B., HUME, D. A. & HAYASHIZAKI, Y. 2006. Genome-wide analysis of mammalian promoter architecture and evolution. *Nat Genet,* 38**,** 626-35.

CHEN, L., SUN, F., YANG, X., JIN, Y., SHI, M., WANG, L., SHI, Y., ZHAN, C. & WANG, Q. 2017. Correlation between RNA-Seq and microarrays results using TCGA data. *Gene,* 628**,** 200-204.

CHO, Y. W., HONG, T., HONG, S., GUO, H., YU, H., KIM, D., GUSZCZYNSKI, T., DRESSLER, G. R., COPELAND, T. D., KALKUM, M. & GE, K. 2007. PTIP

associates with MLL3- and MLL4-containing histone H3 lysine 4 methyltransferase complex. *J Biol Chem,* 282**,** 20395-406.

CIZKOVA, A., STRANECKY, V., MAYR, J. A., TESAROVA, M., HAVLICKOVA, V., PAUL, J., IVANEK, R., KUSS, A. W., HANSIKOVA, H., KAPLANOVA, V., VRBACKY, M., HARTMANNOVA, H., NOSKOVA, L., HONZIK, T., DRAHOTA, Z., MAGNER, M., HEJZLAROVA, K., SPERL, W., ZEMAN, J., HOUSTEK, J. & KMOCH, S. 2008. TMEM70 mutations cause isolated ATP synthase deficiency and neonatal mitochondrial encephalocardiomyopathy. *Nat Genet,* 40**,** 1288-90.

CLARK, E. L., BUSH, S. J., MCCULLOCH, M. E. B., FARQUHAR, I. L., YOUNG, R., LEFEVRE, L., PRIDANS, C., TSANG, H. G., WU, C., AFRASIABI, C., WATSON, M., WHITELAW, C. B., FREEMAN, T. C., SUMMERS, K. M., ARCHIBALD, A. L. & HUME, D. A. 2017. A high resolution atlas of gene expression in the domestic sheep (Ovis aries). *PLoS Genet,* 13**,** e1006997.

CORTEZ, D., MARIN, R., TOLEDO-FLORES, D., FROIDEVAUX, L., LIECHTI, A., WATERS, P. D., GRUTZNER, F. & KAESSMANN, H. 2014. Origins and functional evolution of Y chromosomes across mammals. *Nature,* 508**,** 488-93.

DAI, X., CHEN, X., FANG, Q., LI, J. & BAI, Z. 2017. Inducible CRISPR genome-editing tool: classifications and future trends. *Crit Rev Biotechnol***,** 1-14.

DAIGNEAULT, M., PRESTON, J. A., MARRIOTT, H. M., WHYTE, M. K. & DOCKRELL, D. H. 2010. The identification of markers of macrophage differentiation in PMA-stimulated THP-1 cells and monocyte-derived macrophages. *PLoS One,* 5**,** e8668.

DAVULURI, R. V., SUZUKI, Y., SUGANO, S., PLASS, C. & HUANG, T. H. 2008. The functional consequences of alternative promoter use in mammalian genomes. *Trends Genet,* 24**,** 167-77.

DE CASTRO, E., SIGRIST, C. J., GATTIKER, A., BULLIARD, V., LANGENDIJK-GENEVAUX, P. S., GASTEIGER, E., BAIROCH, A. & HULO, N. 2006. ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Res,* 34**,** W362-5.

DE RIE, D., ABUGESSAISA, I., ALAM, T., ARNER, E., ARNER, P., ASHOOR, H., ASTROM, G., BABINA, M., BERTIN, N., BURROUGHS, A. M., CARLISLE, A. J., DAUB, C. O., DETMAR, M., DEVIATIIAROV, R., FORT, A., GEBHARD, C., GOLDOWITZ, D., GUHL, S., HA, T. J., HARSHBARGER, J., HASEGAWA, A., HASHIMOTO, K., HERLYN, M., HEUTINK, P., HITCHENS, K. J., HON, C. C., HUANG, E., ISHIZU, Y., KAI, C., KASUKAWA, T., KLINKEN, P., LASSMANN, T., LECELLIER, C. H., LEE, W., LIZIO, M., MAKEEV, V., MATHELIER, A., MEDVEDEVA, Y. A., MEJHERT, N., MUNGALL, C. J., NOMA, S., OHSHIMA, M., OKADA-HATAKEYAMA, M., PERSSON, H., RIZZU, P., ROUDNICKY, F., SAETROM, P., SATO, H., SEVERIN, J., SHIN, J. W., SWOBODA, R. K., TARUI, H., TOYODA, H., VITTING-SEERUP, K., WINTERINGHAM, L., YAMAGUCHI, Y., YASUZAWA, K., YONEDA, M., YUMOTO, N., ZABIEROWSKI, S., ZHANG, P. G., WELLS, C. A., SUMMERS, K. M., KAWAJI, H., SANDELIN, A., REHLI, M., CONSORTIUM, F., HAYASHIZAKI, Y., CARNINCI, P., FORREST, A. R. R. & DE HOON, M. J. L. 2017. An integrated expression atlas of miRNAs and their promoters in human and mouse. *Nat Biotechnol,* 35**,** 872-878.

DE SANTA, F., NARANG, V., YAP, Z. H., TUSI, B. K., BURGOLD, T., AUSTENAA, L., BUCCI, G., CAGANOVA, M., NOTARBARTOLO, S., CASOLA, S., TESTA, G., SUNG, W. K., WEI, C. L. & NATOLI, G. 2009. Jmjd3 contributes to the control of gene expression in LPS-activated macrophages. *EMBO J,* 28**,** 3341-52.

DE SANTA, F., TOTARO, M. G., PROSPERINI, E., NOTARBARTOLO, S., TESTA, G. & NATOLI, G. 2007. The histone H3 lysine-27 demethylase Jmjd3 links inflammation to inhibition of polycomb-mediated gene silencing. *Cell,* 130**,** 1083-94.

DEATON, A. M. & BIRD, A. 2011. CpG islands and the regulation of transcription. *Genes Dev,* 25**,** 1010-22.

DEVILBISS, A. W., BOYER, M. E. & BRESNICK, E. H. 2013. Establishing a hematopoietic genetic network through locus-specific integration of chromatin regulators. *Proc Natl Acad Sci U S A,* 110**,** E3398-407.

DEY, A., SESHASAYEE, D., NOUBADE, R., FRENCH, D. M., LIU, J., CHAURUSHIYA, M. S., KIRKPATRICK, D. S., PHAM, V. C., LILL, J. R., BAKALARSKI, C. E., WU, J., PHU, L., KATAVOLOS, P., LAFAVE, L. M., ABDEL-WAHAB, O., MODRUSAN, Z., SESHAGIRI, S., DONG, K., LIN, Z., BALAZS, M., SURIBEN, R., NEWTON, K., HYMOWITZ, S., GARCIA-MANERO, G., MARTIN, F., LEVINE, R. L. & DIXIT, V. M. 2012. Loss of the tumor suppressor BAP1 causes myeloid transformation. *Science,* 337**,** 1541-6.

DHAR, S. S., LEE, S. H., CHEN, K., ZHU, G., OH, W., ALLTON, K., GAFNI, O., KIM, Y. Z., TOMOIGA, A. S., BARTON, M. C., HANNA, J. H., WANG, Z., LI, W. & LEE, M. G. 2016. An essential role for UTX in resolution and activation of bivalent promoters. *Nucleic Acids Res,* 44**,** 3659-74.

DKHISSI, F., AGGOUNE, D., PONTIS, J., SOREL, N., PICCIRILLI, N., LECORF, A., GUILHOT, F., CHOMEL, J. C., AIT-SI-ALI, S. & TURHAN, A. G. 2015. The downregulation of BAP1 expression by BCR-ABL reduces the stability of BRCA1 in chronic myeloid leukemia. *Exp Hematol,* 43**,** 775-80.

DOUDNA, J. A. & CHARPENTIER, E. 2014. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science,* 346**,** 1258096.

DUNFORD, A., WEINSTOCK, D. M., SAVOVA, V., SCHUMACHER, S. E., CLEARY, J. P., YODA, A., SULLIVAN, T. J., HESS, J. M., GIMELBRANT, A. A., BEROUKHIM, R., LAWRENCE, M. S., GETZ, G. & LANE, A. A. 2017. Tumor-suppressor genes that escape from X-inactivation contribute to cancer sex bias. *Nat Genet,* 49**,** 10-16.

ELHAMAMSY, A. R. 2017. Role of DNA methylation in imprinting disorders: an updated review. *J Assist Reprod Genet,* 34**,** 549-562.

ENCODE 2004. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science,* 306**,** 636-40.

ENCODE 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature,* 489**,** 57-74.

ESSAFI-BENKHADIR, K., GROSSO, S., PUISSANT, A., ROBERT, G., ESSAFI, M., DECKERT, M., CHAMOREY, E., DASSONVILLE, O., MILANO, G., AUBERGER, P. & PAGES, G. 2009. Dual role of Sp3 transcription factor as an inducer of apoptosis and a marker of tumour aggressiveness. *PLoS One,* 4**,** e4478.

EWELS, P., MAGNUSSON, M., LUNDIN, S. & KALLER, M. 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics,* 32**,** 3047-8.

FERNANDEZ, P., CARRETERO, J., MEDINA, P. P., JIMENEZ, A. I., RODRIGUEZ-PERALES, S., PAZ, M. F., CIGUDOSA, J. C., ESTELLER, M., LOMBARDIA, L., MORENTE, M., SANCHEZ-VERDE, L., SOTELO, T. & SANCHEZ-CESPEDES, M. 2004. Distinctive gene expression of human lung adenocarcinomas carrying LKB1 mutations. *Oncogene,* 23**,** 5084-91.

FISZER-KIERZKOWSKA, A., VYDRA, N., WYSOCKA-WYCISK, A., KRONEKOVA, Z., JARZAB, M., LISOWSKA, K. M. & KRAWCZYK, Z. 2011. Liposome-based DNA carriers may induce cellular stress response and change gene expression pattern in transfected cells. *BMC Mol Biol,* 12**,** 27.

FORREST, A. R., KAWAJI, H., REHLI, M., BAILLIE, J. K., DE HOON, M. J., HABERLE, V., LASSMANN, T., KULAKOVSKIY, I. V., LIZIO, M., ITOH, M., ANDERSSON, R., MUNGALL, C. J., MEEHAN, T. F., SCHMEIER, S., BERTIN, N., JORGENSEN, M., DIMONT, E., ARNER, E., SCHMIDL, C., SCHAEFER, U., MEDVEDEVA, Y. A., PLESSY, C., VITEZIC, M., SEVERIN, J., SEMPLE, C., ISHIZU, Y., YOUNG, R. S., FRANCESCATTO, M., ALAM, I., ALBANESE, D.,

ALTSCHULER, G. M., ARAKAWA, T., ARCHER, J. A., ARNER, P., BABINA, M., RENNIE, S., BALWIERZ, P. J., BECKHOUSE, A. G., PRADHAN-BHATT, S., BLAKE, J. A., BLUMENTHAL, A., BODEGA, B., BONETTI, A., BRIGGS, J., BROMBACHER, F., BURROUGHS, A. M., CALIFANO, A., CANNISTRACI, C. V., CARBAJO, D., CHEN, Y., CHIERICI, M., CIANI, Y., CLEVERS, H. C., DALLA, E., DAVIS, C. A., DETMAR, M., DIEHL, A. D., DOHI, T., DRABLOS, F., EDGE, A. S., EDINGER, M., EKWALL, K., ENDOH, M., ENOMOTO, H., FAGIOLINI, M., FAIRBAIRN, L., FANG, H., FARACH-CARSON, M. C., FAULKNER, G. J., FAVOROV, A. V., FISHER, M. E., FRITH, M. C., FUJITA, R., FUKUDA, S., FURLANELLO, C., FURINO, M., FURUSAWA, J., GEIJTENBEEK, T. B., GIBSON, A. P., GINGERAS, T., GOLDOWITZ, D., GOUGH, J., GUHL, S., GULER, R., GUSTINCICH, S., HA, T. J., HAMAGUCHI, M., HARA, M., HARBERS, M., HARSHBARGER, J., HASEGAWA, A., HASEGAWA, Y., HASHIMOTO, T., HERLYN, M., HITCHENS, K. J., HO SUI, S. J., HOFMANN, O. M., HOOF, I., HORI, F., HUMINIECKI, L., *et al*. 2014. A promoter-level mammalian expression atlas. *Nature,* 507**,** 462-70.

FORSTER, M., BELLE, J. I., PETROV, J. C., RYDER, E. J., CLARE, S. & NIJNIK, A. 2015. Deubiquitinase MYSM1 Is Essential for Normal Fetal Liver Hematopoiesis and for the Maintenance of Hematopoietic Stem Cells in Adult Bone Marrow. *Stem Cells Dev,* 24**,** 1865-77.

FRANGINI, A., SJOBERG, M., ROMAN-TRUFERO, M., DHARMALINGAM, G., HABERLE, V., BARTKE, T., LENHARD, B., MALUMBRES, M., VIDAL, M. & DILLON, N. 2013. The aurora B kinase and the polycomb protein ring1B combine to regulate active promoters in quiescent lymphocytes. *Mol Cell,* 51**,** 647-61.

FREEMAN, T. C., IVENS, A., BAILLIE, J. K., BERALDI, D., BARNETT, M. W., DORWARD, D., DOWNING, A., FAIRBAIRN, L., KAPETANOVIC, R., RAZA, S., TOMOIU, A., ALBERIO, R., WU, C., SU, A. I., SUMMERS, K. M., TUGGLE, C. K., ARCHIBALD, A. L. & HUME, D. A. 2012. A gene expression atlas of the domestic pig. *BMC Biol,* 10**,** 90.

FUKAYA, T., LIM, B. & LEVINE, M. 2016. Enhancer Control of Transcriptional Bursting. *Cell,* 166**,** 358-368.

GALLAGHER, R., COLLINS, S., TRUJILLO, J., MCCREDIE, K., AHEARN, M., TSAI, S., METZGAR, R., AULAKH, G., TING, R., RUSCETTI, F. & GALLO, R. 1979. Characterization of the continuous, differentiating myeloid cell line (HL-60) from a patient with acute promyelocytic leukemia. *Blood,* 54**,** 713-33.

GANGULA, N. R. & MADDIKA, S. 2013. WD repeat protein WDR48 in complex with deubiquitinase USP12 suppresses Akt-dependent cell survival signaling by stabilizing PH domain leucine-rich repeat protein phosphatase 1 (PHLPP1). *J Biol Chem,* 288**,** 34545-54.

GATES, L. A., FOULDS, C. E. & O'MALLEY, B. W. 2017. Histone Marks in the 'Driver's Seat': Functional Roles in Steering the Transcription Cycle. *Trends Biochem Sci,* 42**,** 977-989.

GELSI-BOYER, V., TROUPLIN, V., ADELAIDE, J., ACETO, N., REMY, V., PINSON, S., HOUDAYER, C., ARNOULET, C., SAINTY, D., BENTIRES-ALJ, M., OLSCHWANG, S., VEY, N., MOZZICONACCI, M. J., BIRNBAUM, D. & CHAFFANET, M. 2008. Genome profiling of chronic myelomonocytic leukemia: frequent alterations of RAS and RUNX1 genes. *BMC Cancer,* 8**,** 299.

GERRARD, D. T. & FILATOV, D. A. 2005. Positive and negative selection on mammalian Y chromosomes. *Mol Biol Evol,* 22**,** 1423-32.

GERSTEIN, M. B., BRUCE, C., ROZOWSKY, J. S., ZHENG, D., DU, J., KORBEL, J. O., EMANUELSSON, O., ZHANG, Z. D., WEISSMAN, S. & SNYDER, M. 2007. What is a gene, post-ENCODE? History and updated definition. *Genome Res,* 17**,** 669-81.

GEZGIN, G., DOGRUSOZ, M., VAN ESSEN, T. H., KROES, W. G. M., LUYTEN, G. P. M., VAN DER VELDEN, P. A., WALTER, V., VERDIJK, R. M., VAN HALL, T., VAN DER BURG, S. H. & JAGER, M. J. 2017. Genetic evolution of uveal melanoma guides the development of an inflammatory microenvironment. *Cancer Immunol Immunother,* 66**,** 903-912.

GIARD, D. J., AARONSON, S. A., TODARO, G. J., ARNSTEIN, P., KERSEY, J. H., DOSIK, H. & PARKS, W. P. 1973. In vitro cultivation of human tumors: establishment of cell lines derived from a series of solid tumors. *J Natl Cancer Inst,* 51**,** 1417-23.

GOETZE, R. W., KIM, D. H., SCHINAZI, R. F. & KIM, B. 2017. A CRISPR/Cas9 approach reveals that the polymerase activity of DNA polymerase beta is dispensable for HIV-1 infection in dividing and nondividing cells. *J Biol Chem,* 292**,** 14016-14025.

GORDON, S. & TAYLOR, P. R. 2005. Monocyte and macrophage heterogeneity. *Nat Rev Immunol,* 5**,** 953-64.

GREENFIELD, A., CARREL, L., PENNISI, D., PHILIPPE, C., QUADERI, N., SIGGERS, P., STEINER, K., TAM, P. P., MONACO, A. P., WILLARD, H. F. & KOOPMAN, P. 1998. The UTX gene escapes X inactivation in mice and humans. *Hum Mol Genet,* 7**,** 737-42.

GU, Y., JONES, A. E., YANG, W., LIU, S., DAI, Q., LIU, Y., SWINDLE, C. S., ZHOU, D., ZHANG, Z., RYAN, T. M., TOWNES, T. M., KLUG, C. A., CHEN, D. & WANG, H. 2016. The histone H2A deubiquitinase Usp16 regulates hematopoiesis and hematopoietic stem cell function. *Proc Natl Acad Sci U S A,* 113**,** E51-60.

HA, S. D., CHO, W. & KIM, S. O. 2017. HDAC8 Prevents Anthrax Lethal Toxin-induced Cell Cycle Arrest through Silencing PTEN in Human Monocytic THP-1 Cells. *Toxins (Basel),* 9.

HABERLE, V., FORREST, A. R., HAYASHIZAKI, Y., CARNINCI, P. & LENHARD, B. 2015. CAGEr: precise TSS data retrieval and high-resolution promoterome mining for integrative analyses. *Nucleic Acids Res,* 43**,** e51.

HABERLE, V., LI, N., HADZHIEV, Y., PLESSY, C., PREVITI, C., NEPAL, C., GEHRIG, J., DONG, X., AKALIN, A., SUZUKI, A. M., VAN, I. W. F. J., ARMANT, O., FERG, M., STRAHLE, U., CARNINCI, P., MULLER, F. & LENHARD, B. 2014. Two independent transcription initiation codes overlap on vertebrate core promoters. *Nature,* 507**,** 381-385.

HAH, N., BENNER, C., CHONG, L. W., YU, R. T., DOWNES, M. & EVANS, R. M. 2015. Inflammation-sensitive super enhancers form domains of coordinately regulated enhancer RNAs. *Proc Natl Acad Sci U S A,* 112**,** E297-302.

HALL, N. M., BROWN, G. M., FURLONG, R. A., SARGENT, C. A., MITCHELL, M., ROCHA, D. & AFFARA, N. A. 2003. Usp9y (ubiquitin-specific protease 9 gene on the Y) is associated with a functional promoter and encodes an intact open reading frame homologous to Usp9x that is under selective constraint. *Mamm Genome,* 14**,** 437-47.

HARD, R. L., LIU, J., SHEN, J., ZHOU, P. & PEI, D. 2010. HDAC6 and Ubp-M BUZ domains recognize specific C-terminal sequences of proteins. *Biochemistry,* 49**,** 10737-46.

HARRIS, P. E., RALPH, P., LITCOFSKY, P. & MOORE, M. A. 1985. Distinct activities of interferon-gamma, lymphokine and cytokine differentiation-inducing factors acting on the human monoblastic leukemia cell line U937. *Cancer Res,* 45**,** 9-13.

HEINEMANN, B., NIELSEN, J. M., HUDLEBUSCH, H. R., LEES, M. J., LARSEN, D. V., BOESEN, T., LABELLE, M., GERLACH, L. O., BIRK, P. & HELIN, K. 2014. Inhibition of demethylases by GSK-J1/J4. *Nature,* 514**,** E1-2.

HNISZ, D., ABRAHAM, B. J., LEE, T. I., LAU, A., SAINT-ANDRE, V., SIGOVA, A. A., HOKE, H. A. & YOUNG, R. A. 2013. Super-enhancers in the control of cell identity and disease. *Cell,* 155**,** 934-47.

HOEKSEMA, M. A. & DE WINTHER, M. P. 2016. Epigenetic Regulation of Monocyte and Macrophage Function. *Antioxid Redox Signal,* 25**,** 758-774.

HON, C. C., RAMILOWSKI, J. A., HARSHBARGER, J., BERTIN, N., RACKHAM, O. J., GOUGH, J., DENISENKO, E., SCHMEIER, S., POULSEN, T. M., SEVERIN, J., LIZIO, M., KAWAJI, H., KASUKAWA, T., ITOH, M., BURROUGHS, A. M., NOMA, S., DJEBALI, S., ALAM, T., MEDVEDEVA, Y. A., TESTA, A. C., LIPOVICH, L., YIP, C. W., ABUGESSAISA, I., MENDEZ, M., HASEGAWA, A., TANG, D., LASSMANN, T., HEUTINK, P., BABINA, M., WELLS, C. A., KOJIMA, S., NAKAMURA, Y., SUZUKI, H., DAUB, C. O., DE HOON, M. J., ARNER, E., HAYASHIZAKI, Y., CARNINCI, P. & FORREST, A. R. 2017. An atlas of human long non-coding RNAs with accurate 5' ends. *Nature,* 543**,** 199-204.

HONG, S., CHO, Y. W., YU, L. R., YU, H., VEENSTRA, T. D. & GE, K. 2007. Identification of JmjC domain-containing UTX and JMJD3 as histone H3 lysine 27 demethylases. *Proc Natl Acad Sci U S A,* 104**,** 18439-44.

HOOK, E. B. & WARBURTON, D. 2014. Turner syndrome revisited: review of new data supports the hypothesis that all viable 45,X cases are cryptic mosaics with a rescue cell line, implying an origin by mitotic loss. *Hum Genet,* 133**,** 417-24.

HUANG DA, W., SHERMAN, B. T. & LEMPICKI, R. A. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc,* 4**,** 44-57.

HUANG, X., SPENCER, G. J., LYNCH, J. T., CICERI, F., SOMERVILLE, T. D. & SOMERVAILLE, T. C. 2014. Enhancers of Polycomb EPC1 and EPC2 sustain the oncogenic potential of MLL leukemia stem cells. *Leukemia,* 28**,** 1081-91.

HUGHES, P., MARSHALL, D., REID, Y., PARKES, H. & GELBER, C. 2007. The costs of using unauthenticated, over-passaged cell lines: how much more data do we need? *Biotechniques,* 43**,** 575, 577-8, 581-2 passim.

HUME, D. A. & FREEMAN, T. C. 2014. Transcriptomic analysis of mononuclear phagocyte differentiation and activation. *Immunol Rev,* 262**,** 74-84.

HUME, D. A., SUMMERS, K. M. & REHLI, M. 2016. Transcriptional Regulation and Macrophage Differentiation. *Microbiol Spectr,* 4.

IGLESIAS, M. J., REILLY, S. J., EMANUELSSON, O., SENNBLAD, B., PIRMORADIAN NAJAFABADI, M., FOLKERSEN, L., MALARSTIG, A., LAGERGREN, J., ERIKSSON, P., HAMSTEN, A. & ODEBERG, J. 2012. Combined chromatin and expression analysis reveals specific regulatory mechanisms within cytokine genes in the macrophage early immune response. *PLoS One,* 7**,** e32306.

IYER, V., SHEN, B., ZHANG, W., HODGKINS, A., KEANE, T., HUANG, X. & SKARNES, W. C. 2015. Off-target mutations are rare in Cas9-modified mice. *Nat Methods,* 12**,** 479.

JAHAN, A. S., LESTRA, M., SWEE, L. K., FAN, Y., LAMERS, M. M., TAFESSE, F. G., THEILE, C. S., SPOONER, E., BRUZZONE, R., PLOEGH, H. L. & SANYAL, S. 2016. Usp12 stabilizes the T-cell receptor complex at the cell surface during signaling. *Proc Natl Acad Sci U S A,* 113**,** E705-14.

JENSEN, D. E., PROCTOR, M., MARQUIS, S. T., GARDNER, H. P., HA, S. I., CHODOSH, L. A., ISHOV, A. M., TOMMERUP, N., VISSING, H., SEKIDO, Y., MINNA, J., BORODOVSKY, A., SCHULTZ, D. C., WILKINSON, K. D., MAUL, G. G., BARLEV, N., BERGER, S. L., PRENDERGAST, G. C. & RAUSCHER, F. J., 3RD 1998. BAP1: a novel ubiquitin hydrolase which binds to the BRCA1 RING finger and enhances BRCA1-mediated cell growth suppression. *Oncogene,* 16**,** 1097-112.

JI, Z., MOHAMMED, H., WEBBER, A., RIDSDALE, J., HAN, N., CARROLL, J. S. & SHARROCKS, A. D. 2014. The forkhead transcription factor FOXK2 acts as a chromatin targeting factor for the BAP1-containing histone deubiquitinase complex. *Nucleic Acids Res,* 42**,** 6232-42.

JIANG, X. X., NGUYEN, Q., CHOU, Y., WANG, T., NANDAKUMAR, V., YATES, P., JONES, L., WANG, L., WON, H., LEE, H. R., JUNG, J. U., MUSCHEN, M., HUANG, X. F. & CHEN, S. Y. 2011. Control of B cell development by the histone H2A deubiquitinase MYSM1. *Immunity,* 35**,** 883-96.

JONKERS, I. & LIS, J. T. 2015. Getting up to speed with transcription elongation by RNA polymerase II. *Nat Rev Mol Cell Biol,* 16**,** 167-77.

JOO, H. Y., JONES, A., YANG, C., ZHAI, L., SMITH, A. D. T., ZHANG, Z., CHANDRASEKHARAN, M. B., SUN, Z. W., RENFROW, M. B., WANG, Y., CHANG, C. & WANG, H. 2011. Regulation of histone H2A and H2B deubiquitination and Xenopus development by USP12 and USP46. *J Biol Chem,* 286**,** 7190-201.

JOO, H. Y., ZHAI, L., YANG, C., NIE, S., ERDJUMENT-BROMAGE, H., TEMPST, P., CHANG, C. & WANG, H. 2007. Regulation of cell cycle progression and gene expression by H2A deubiquitination. *Nature,* 449**,** 1068-72.

JOSEPH, R. W., KAPUR, P., SERIE, D. J., ECKEL-PASSOW, J. E., PARASRAMKA, M., HO, T., CHEVILLE, J. C., FRENKEL, E., RAKHEJA, D., BRUGAROLAS, J. & PARKER, A. 2014. Loss of BAP1 protein expression is an independent marker of poor prognosis in patients with low-risk clear cell renal cell carcinoma. *Cancer,* 120**,** 1059-67.

KAIKKONEN, M. U., SPANN, N. J., HEINZ, S., ROMANOSKI, C. E., ALLISON, K. A., STENDER, J. D., CHUN, H. B., TOUGH, D. F., PRINJHA, R. K., BENNER, C. & GLASS, C. K. 2013. Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer transcription. *Mol Cell,* 51**,** 310-25.

KAMIYA, M., JUDSON, H., OKAZAKI, Y., KUSAKABE, M., MURAMATSU, M., TAKADA, S., TAKAGI, N., ARIMA, T., WAKE, N., KAMIMURA, K., SATOMURA, K., HERMANN, R., BONTHRON, D. T. & HAYASHIZAKI, Y. 2000. The cell cycle control gene ZAC/PLAGL1 is imprinted--a strong candidate gene for transient neonatal diabetes. *Hum Mol Genet,* 9**,** 453-60.

KANAMORI-KATAYAMA, M., ITOH, M., KAWAJI, H., LASSMANN, T., KATAYAMA, S., KOJIMA, M., BERTIN, N., KAIHO, A., NINOMIYA, N., DAUB, C. O., CARNINCI, P., FORREST, A. R. & HAYASHIZAKI, Y. 2011. Unamplified cap analysis of gene expression on a single-molecule sequencer. *Genome Res,* 21**,** 1150-9.

KAWAJI, H., LIZIO, M., ITOH, M., KANAMORI-KATAYAMA, M., KAIHO, A., NISHIYORI-SUEKI, H., SHIN, J. W., KOJIMA-ISHIYAMA, M., KAWANO, M., MURATA, M., NINOMIYA-FUKUDA, N., ISHIKAWA-KATO, S., NAGAO-SATO, S., NOMA, S., HAYASHIZAKI, Y., FORREST, A. R., CARNINCI, P. & CONSORTIUM, F. 2014. Comparison of CAGE and RNA-seq transcriptome profiling using clonally amplified and single-molecule next-generation sequencing. *Genome Res,* 24**,** 708-17.

KAWANE, K., FUKUYAMA, H., KONDOH, G., TAKEDA, J., OHSAWA, Y., UCHIYAMA, Y. & NAGATA, S. 2001. Requirement of DNase II for definitive erythropoiesis in the mouse fetal liver. *Science,* 292**,** 1546-9.

KIM, D., HONG, A., PARK, H. I., SHIN, W. H., YOO, L., JEON, S. J. & CHUNG, K. C. 2017. Deubiquitinating enzyme USP22 positively regulates c-Myc stability and tumorigenic activity in mammalian and breast cancer cells. *J Cell Physiol,* 232**,** 3664-3676.

KIM, J. H., SOHN, H. J., YOO, J. K., KANG, H., SEONG, G. S., CHWAE, Y. J., KIM, K., PARK, S. & SHIN, H. J. 2016. NLRP3 Inflammasome Activation in THP-1 Target Cells Triggered by Pathogenic Naegleria fowleri. *Infect Immun,* 84**,** 2422-8.

KIM, T. K., HEMBERG, M., GRAY, J. M., COSTA, A. M., BEAR, D. M., WU, J., HARMIN, D. A., LAPTEWICZ, M., BARBARA-HALEY, K., KUERSTEN, S., MARKENSCOFF-PAPADIMITRIOU, E., KUHL, D., BITO, H., WORLEY, P. F.,

KREIMAN, G. & GREENBERG, M. E. 2010. Widespread transcription at neuronal activity-regulated enhancers. *Nature,* 465**,** 182-7.

KLOSE, R. J., KALLIN, E. M. & ZHANG, Y. 2006. JmjC-domain-containing proteins and histone demethylation. *Nat Rev Genet,* 7**,** 715-27.

KOBUNE, F., AMI, Y., KATAYAMA, M., TAKAHASHI, M., TUUL, R., KORUKLUOGLU, G., KIYOHARA, T., MIURA, R., SATO, H., YONEDA, M. & KAI, C. 2007. A novel monolayer cell line derived from human umbilical cord blood cells shows high sensitivity to measles virus. *J Gen Virol,* 88**,** 1565-7.

KOMANDER, D., CLAGUE, M. J. & URBE, S. 2009. Breaking the chains: structure and function of the deubiquitinases. *Nat Rev Mol Cell Biol,* 10**,** 550-63.

KRUIDENIER, L., CHUNG, C. W., CHENG, Z., LIDDLE, J., CHE, K., JOBERTY, G., BANTSCHEFF, M., BOUNTRA, C., BRIDGES, A., DIALLO, H., EBERHARD, D., HUTCHINSON, S., JONES, E., KATSO, R., LEVERIDGE, M., MANDER, P. K., MOSLEY, J., RAMIREZ-MOLINA, C., ROWLAND, P., SCHOFIELD, C. J., SHEPPARD, R. J., SMITH, J. E., SWALES, C., TANNER, R., THOMAS, P., TUMBER, A., DREWES, G., OPPERMANN, U., PATEL, D. J., LEE, K. & WILSON, D. M. 2012. A selective jumonji H3K27 demethylase inhibitor modulates the proinflammatory macrophage response. *Nature,* 488**,** 404-8.

KU, M., KOCHE, R. P., RHEINBAY, E., MENDENHALL, E. M., ENDOH, M., MIKKELSEN, T. S., PRESSER, A., NUSBAUM, C., XIE, X., CHI, A. S., ADLI, M., KASIF, S., PTASZEK, L. M., COWAN, C. A., LANDER, E. S., KOSEKI, H. & BERNSTEIN, B. E. 2008. Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genet,* 4**,** e1000242.

KUROKI, Y., SUZUKI, Y., CHYO, H., HATA, A. & MATSUI, I. 1981. A new malformation syndrome of long palpebral fissures, large ears, depressed nasal tip, and skeletal anomalies associated with postnatal dwarfism and mental retardation. *J Pediatr,* 99**,** 570-3.

LAASER, I., THEIS, F. J., DE ANGELIS, M. H., KOLB, H. J. & ADAMSKI, J. 2011. Huge splicing frequency in human Y chromosomal UTY gene. *OMICS,* 15**,** 141-54.

LACHMANN, A., XU, H., KRISHNAN, J., BERGER, S. I., MAZLOOM, A. R. & MA'AYAN, A. 2010. ChEA: transcription factor regulation inferred from integrating genome-wide ChIP-X experiments. *Bioinformatics,* 26**,** 2438-44.

LAFAVE, L. M., BEGUELIN, W., KOCHE, R., TEATER, M., SPITZER, B., CHRAMIEC, A., PAPALEXI, E., KELLER, M. D., HRICIK, T., KONSTANTINOFF, K., MICOL, J. B., DURHAM, B., KNUTSON, S. K., CAMPBELL, J. E., BLUM, G., SHI, X., DOUD, E. H., KRIVTSOV, A. V., CHUNG, Y. R., KHODOS, I., DE STANCHINA, E., OUERFELLI, O., ADUSUMILLI, P. S., THOMAS, P. M., KELLEHER, N. L., LUO, M., KEILHACK, H., ABDEL-WAHAB, O., MELNICK, A., ARMSTRONG, S. A. & LEVINE, R. L. 2015. Loss of BAP1 function leads to EZH2-dependent transformation. *Nat Med,* 21**,** 1344-9.

LAGARDE, J., USZCZYNSKA-RATAJCZAK, B., SANTOYO-LOPEZ, J., GONZALEZ, J. M., TAPANARI, E., MUDGE, J. M., STEWARD, C. A., WILMING, L., TANZER, A., HOWALD, C., CHRAST, J., VELA-BOZA, A., RUEDA, A., LOPEZ-DOMINGO, F. J., DOPAZO, J., REYMOND, A., GUIGO, R. & HARROW, J. 2016. Extension of human lncRNA transcripts by RACE coupled with long-read high-throughput sequencing (RACE-Seq). *Nat Commun,* 7**,** 12339.

LAHN, B. T. & PAGE, D. C. 1999. Four evolutionary strata on the human X chromosome. *Science,* 286**,** 964-7.

LANZUOLO, C. & ORLANDO, V. 2012. Memories from the polycomb group proteins. *Annu Rev Genet,* 46**,** 561-89.

LAUGEL, B., LLOYD, A., MEERMEIER, E. W., CROWTHER, M. D., CONNOR, T. R., DOLTON, G., MILES, J. J., BURROWS, S. R., GOLD, M. C., LEWINSOHN, D. M. & SEWELL, A. K. 2016. Engineering of Isogenic Cells Deficient for MR1 with a

CRISPR/Cas9 Lentiviral System: Tools To Study Microbial Antigen Processing and Presentation to Human MR1-Restricted T Cells. *J Immunol,* 197**,** 971-82.
LAVIN, Y., WINTER, D., BLECHER-GONEN, R., DAVID, E., KEREN-SHAUL, H., MERAD, M., JUNG, S. & AMIT, I. 2014. Tissue-resident macrophage enhancer landscapes are shaped by the local microenvironment. *Cell,* 159**,** 1312-26.
LEDERER, D., GRISART, B., DIGILIO, M. C., BENOIT, V., CRESPIN, M., GHARIANI, S. C., MAYSTADT, I., DALLAPICCOLA, B. & VERELLEN-DUMOULIN, C. 2012. Deletion of KDM6A, a histone demethylase interacting with MLL2, in three patients with Kabuki syndrome. *Am J Hum Genet,* 90**,** 119-24.
LEE, J. Y., LEE, C. H., SHIM, S. H., SEO, H. K., KYHM, J. H., CHO, S. & CHO, Y. H. 2002. Molecular cytogenetic analysis of the monoblastic cell line U937. karyotype clarification by G-banding, whole chromosome painting, microdissection and reverse painting, and comparative genomic hybridization. *Cancer Genet Cytogenet,* 137**,** 124-32.
LEE, Y. C., HUANG, H. Y., CHANG, C. J., CHENG, C. H. & CHEN, Y. T. 2010. Mitochondrial GLUT10 facilitates dehydroascorbic acid import and protects cells against oxidative stress: mechanistic insight into arterial tortuosity syndrome. *Hum Mol Genet,* 19**,** 3721-33.
LEON-PONTE, M., KIRCHHOF, M. G., SUN, T., STEPHENS, T., SINGH, B., SANDHU, S. & MADRENAS, J. 2005. Polycationic lipids inhibit the pro-inflammatory response to LPS. *Immunol Lett,* 96**,** 73-83.
LER, L. D., GHOSH, S., CHAI, X., THIKE, A. A., HENG, H. L., SIEW, E. Y., DEY, S., KOH, L. K., LIM, J. Q., LIM, W. K., MYINT, S. S., LOH, J. L., ONG, P., SAM, X. X., HUANG, D., LIM, T., TAN, P. H., NAGARAJAN, S., CHENG, C. W., HO, H., NG, L. G., YUEN, J., LIN, P. H., CHUANG, C. K., CHANG, Y. H., WENG, W. H., ROZEN, S. G., TAN, P., CREASY, C. L., PANG, S. T., MCCABE, M. T., POON, S. L. & TEH, B. T. 2017. Loss of tumor suppressor KDM6A amplifies PRC2-regulated transcriptional repression in bladder cancer and can be targeted through inhibition of EZH2. *Sci Transl Med,* 9.
LESCH, B. J., SILBER, S. J., MCCARREY, J. R. & PAGE, D. C. 2016. Parallel evolution of male germline epigenetic poising and somatic development in animals. *Nat Genet,* 48**,** 888-94.
LI, G., DAVIS, B. W., RAUDSEPP, T., PEARKS WILKERSON, A. J., MASON, V. C., FERGUSON-SMITH, M., O'BRIEN, P. C., WATERS, P. D. & MURPHY, W. J. 2013a. Comparative analysis of mammalian Y chromosomes illuminates ancestral structure and lineage-specific evolution. *Genome Res,* 23**,** 1486-95.
LI, H. & DURBIN, R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics,* 25**,** 1754-60.
LI, H., HANDSAKER, B., WYSOKER, A., FENNELL, T., RUAN, J., HOMER, N., MARTH, G., ABECASIS, G., DURBIN, R. & GENOME PROJECT DATA PROCESSING, S. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics,* 25**,** 2078-9.
LI, W., NOTANI, D., MA, Q., TANASA, B., NUNEZ, E., CHEN, A. Y., MERKURJEV, D., ZHANG, J., OHGI, K., SONG, X., OH, S., KIM, H. S., GLASS, C. K. & ROSENFELD, M. G. 2013b. Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature,* 498**,** 516-20.
LI, X., ZHANG, Q., SHI, Q., LIU, Y., ZHAO, K., SHEN, Q., SHI, Y., LIU, X., WANG, C., LI, N., MA, Y. & CAO, X. 2017. Demethylase Kdm6a epigenetically promotes IL-6 and IFN-beta production in macrophages. *J Autoimmun,* 80**,** 85-94.
LI, Y., LU, Y., WANG, S., HAN, Z., ZHU, F., NI, Y., LIANG, R., ZHANG, Y., LENG, Q., WEI, G., SHI, G., ZHU, R., LI, D., WANG, H., ZHENG, S. G., XU, H., TSUN, A. & LI, B. 2016. USP21 prevents the generation of T-helper-1-like Treg cells. *Nat Commun,* 7**,** 13559.

LIFESCIENCES, D. G. 2017. *Effective controls for RNA Interference (RNAi) experiments using siRNA* [Online]. Available: http://dharmacon.gelifesciences.com/uploadedFiles/Resources/effective-sirna-controls-technote.pdf [Accessed 27/09 2017].

LIN, M., SUTHERLAND, D. R., HORSFALL, W., TOTTY, N., YEO, E., NAYAR, R., WU, X. F. & SCHUH, A. C. 2002. Cell surface antigen CD109 is a novel member of the alpha(2) macroglobulin/C3, C4, C5 family of thioester-containing proteins. *Blood,* 99**,** 1683-91.

LONG, H. K., KING, H. W., PATIENT, R. K., ODOM, D. T. & KLOSE, R. J. 2016. Protection of CpG islands from DNA methylation is DNA-encoded and evolutionarily conserved. *Nucleic Acids Res,* 44**,** 6693-706.

LUND, M. E., TO, J., O'BRIEN, B. A. & DONNELLY, S. 2016. The choice of phorbol 12-myristate 13-acetate differentiation protocol influences the response of THP-1 macrophages to a pro-inflammatory stimulus. *J Immunol Methods,* 430**,** 64-70.

MA, T. T., LI, X. F., LI, W. X., YANG, Y., HUANG, C., MENG, X. M., ZHANG, L. & LI, J. 2015. Geniposide alleviates inflammation by suppressing MeCP2 in mice with carbon tetrachloride-induced acute liver injury and LPS-treated THP-1 cells. *Int Immunopharmacol,* 29**,** 739-747.

MAESS, M. B., BUERS, I., ROBENEK, H. & LORKOWSKI, S. 2011. Improved protocol for efficient nonviral transfection of premature THP-1 macrophages. *Cold Spring Harb Protoc,* 2011**,** pdb prot5612.

MAESS, M. B., SENDELBACH, S. & LORKOWSKI, S. 2010. Selection of reliable reference genes during THP-1 monocyte differentiation into macrophages. *BMC Mol Biol,* 11**,** 90.

MANNA, S., KIM, J. K., BAUGE, C., CAM, M., ZHAO, Y., SHETTY, J., VACCHIO, M. S., CASTRO, E., TRAN, B., TESSAROLLO, L. & BOSSELUT, R. 2015. Histone H3 Lysine 27 demethylases Jmjd3 and Utx are required for T-cell differentiation. *Nat Commun,* 6**,** 8152.

MARIONI, J. C., MASON, C. E., MANE, S. M., STEPHENS, M. & GILAD, Y. 2008. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res,* 18**,** 1509-17.

MARKBY, G. R., SUMMERS, K. M., MACRAE, V. E. & CORCORAN, B. M. 2017. Comparative Transcriptomic Profiling and Gene Expression for Myxomatous Mitral Valve Disease in the Dog and Human. *Vet Sci,* 4.

MARTENS, J. H. & STUNNENBERG, H. G. 2013. BLUEPRINT: mapping human blood cell epigenomes. *Haematologica,* 98**,** 1487-9.

MARTIN, M. 2011. Cutadapt Removes Adapter Sequences From High-Throughput Sequencing Reads. *EMBnet,* 17**,** 10-12.

MARUI, T., FUNATOGAWA, I., KOISHI, S., YAMAMOTO, K., MATSUMOTO, H., HASHIMOTO, O., JINDE, S., NISHIDA, H., SUGIYAMA, T., KASAI, K., WATANABE, K., KANO, Y. & KATO, N. 2011. The NADH-ubiquinone oxidoreductase 1 alpha subcomplex 5 (NDUFA5) gene variants are associated with autism. *Acta Psychiatr Scand,* 123**,** 118-24.

MATEOS, M. K., BARBARIC, D., BYATT, S. A., SUTTON, R. & MARSHALL, G. M. 2015. Down syndrome and leukemia: insights into leukemogenesis and translational targets. *Transl Pediatr,* 4**,** 76-92.

MATHELIER, A., ZHAO, X., ZHANG, A. W., PARCY, F., WORSLEY-HUNT, R., ARENILLAS, D. J., BUCHMAN, S., CHEN, C. Y., CHOU, A., IENASESCU, H., LIM, J., SHYR, C., TAN, G., ZHOU, M., LENHARD, B., SANDELIN, A. & WASSERMAN, W. W. 2014. JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. *Nucleic Acids Res,* 42**,** D142-7.

MATYS, V., FRICKE, E., GEFFERS, R., GOSSLING, E., HAUBROCK, M., HEHL, R., HORNISCHER, K., KARAS, D., KEL, A. E., KEL-MARGOULIS, O. V., KLOOS, D. U., LAND, S., LEWICKI-POTAPOV, B., MICHAEL, H., MUNCH, R., REUTER, I., ROTERT, S., SAXEL, H., SCHEER, M., THIELE, S. & WINGENDER, E. 2003. TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res,* 31**,** 374-8.

MATYS, V., KEL-MARGOULIS, O. V., FRICKE, E., LIEBICH, I., LAND, S., BARRE-DIRRIE, A., REUTER, I., CHEKMENEV, D., KRULL, M., HORNISCHER, K., VOSS, N., STEGMAIER, P., LEWICKI-POTAPOV, B., SAXEL, H., KEL, A. E. & WINGENDER, E. 2006. TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res,* 34**,** D108-10.

MCCARTHY, D. J., CHEN, Y. & SMYTH, G. K. 2012. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res,* 40**,** 4288-97.

MCCLURG, U. L., HARLE, V. J., NABBI, A., BATALHA-PEREIRA, A., WALKER, S., COFFEY, K., GAUGHAN, L., MCCRACKEN, S. R. & ROBSON, C. N. 2015. Ubiquitin-specific protease 12 interacting partners Uaf-1 and WDR20 are potential therapeutic targets in prostate cancer. *Oncotarget,* 6**,** 37724-36.

MCCLURG, U. L., SUMMERSCALES, E. E., HARLE, V. J., GAUGHAN, L. & ROBSON, C. N. 2014. Deubiquitinating enzyme Usp12 regulates the interaction between the androgen receptor and the Akt pathway. *Oncotarget,* 5**,** 7081-92.

MILLER, S. A., MOHN, S. E. & WEINMANN, A. S. 2010. Jmjd3 and UTX play a demethylase-independent role in chromatin remodeling to regulate T-box family member-dependent gene expression. *Mol Cell,* 40**,** 594-605.

MIYAKE, N., MIZUNO, S., OKAMOTO, N., OHASHI, H., SHIINA, M., OGATA, K., TSURUSAKI, Y., NAKASHIMA, M., SAITSU, H., NIIKAWA, N. & MATSUMOTO, N. 2013. KDM6A point mutations cause Kabuki syndrome. *Hum Mutat,* 34**,** 108-10.

MORALES TORRES, C., LAUGESEN, A. & HELIN, K. 2013. Utx is required for proper induction of ectoderm and mesoderm during differentiation of embryonic stem cells. *PLoS One,* 8**,** e60020.

MORETTI, J., CHASTAGNER, P., LIANG, C. C., COHN, M. A., ISRAEL, A. & BROU, C. 2012. The ubiquitin-specific protease 12 (USP12) is a negative regulator of notch signaling acting on notch receptor trafficking toward degradation. *J Biol Chem,* 287**,** 29429-41.

MOSBECH, A., LUKAS, C., BEKKER-JENSEN, S. & MAILAND, N. 2013. The deubiquitylating enzyme USP44 counteracts the DNA double-strand break response mediated by the RNF8 and RNF168 ubiquitin ligases. *J Biol Chem,* 288**,** 16579-87.

MUKAI, M., SURUGA, N., SAEKI, N. & OGAWA, K. 2017. EphA receptors and ephrin-A ligands are upregulated by monocytic differentiation/maturation and promote cell adhesion and protrusion formation in HL60 monocytes. *BMC Cell Biol,* 18**,** 28.

MURATA, M., NISHIYORI-SUEKI, H., KOJIMA-ISHIYAMA, M., CARNINCI, P., HAYASHIZAKI, Y. & ITOH, M. 2014. Detecting expressed genes using CAGE. *Methods Mol Biol,* 1164**,** 67-85.

NAJI, L., PACHOLSKY, D. & ASPENSTROM, P. 2011. ARHGAP30 is a Wrch-1-interacting protein involved in actin dynamics and cell adhesion. *Biochem Biophys Res Commun,* 409**,** 96-102.

NAYAK, T. K., ALAMURU-YELLAPRAGADA, N. P. & PARSA, K. V. 2017. Deubiquitinase USP12 promotes LPS induced macrophage responses through inhibition of IkappaBalpha. *Biochem Biophys Res Commun,* 483**,** 69-74.

NEME, A., NURMINEN, V., SEUTER, S. & CARLBERG, C. 2016. The vitamin D-dependent transcriptome of human monocytes. *J Steroid Biochem Mol Biol,* 164**,** 180-187.

NGUYEN, T. H., BERTRAND, M. J., STERPIN, C., ACHOURI, Y. & DE BACKER, O. R. 2010. Maged1, a new regulator of skeletal myogenic differentiation and muscle regeneration. *BMC Cell Biol,* 11**,** 57.

NGUYEN, T. M., ARTHUR, A., ZANNETTINO, A. C. & GRONTHOS, S. 2017. EphA5 and EphA7 forward signaling enhances human hematopoietic stem and progenitor cell maintenance, migration, and adhesion via Rac1 activation. *Exp Hematol,* 48**,** 72-78.

NIIKAWA, N., MATSUURA, N., FUKUSHIMA, Y., OHSAWA, T. & KAJII, T. 1981. Kabuki make-up syndrome: a syndrome of mental retardation, unusual facies, large and protruding ears, and postnatal growth deficiency. *J Pediatr,* 99**,** 565-9.

NTZIACHRISTOS, P., TSIRIGOS, A., WELSTEAD, G. G., TRIMARCHI, T., BAKOGIANNI, S., XU, L., LOIZOU, E., HOLMFELDT, L., STRIKOUDIS, A., KING, B., MULLENDERS, J., BECKSFORT, J., NEDJIC, J., PAIETTA, E., TALLMAN, M. S., ROWE, J. M., TONON, G., SATOH, T., KRUIDENIER, L., PRINJHA, R., AKIRA, S., VAN VLIERBERGHE, P., FERRANDO, A. A., JAENISCH, R., MULLIGHAN, C. G. & AIFANTIS, I. 2014. Contrasting roles of histone 3 lysine 27 demethylases in acute lymphoblastic leukaemia. *Nature,* 514**,** 513-7.

OLIVARIUS, S., PLESSY, C. & CARNINCI, P. 2009. High-throughput verification of transcriptional starting sites by Deep-RACE. *Biotechniques,* 46**,** 130-2.

ONG, C. T. & CORCES, V. G. 2011. Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nat Rev Genet,* 12**,** 283-93.

OSTUNI, R., PICCOLO, V., BAROZZI, I., POLLETTI, S., TERMANINI, A., BONIFACIO, S., CURINA, A., PROSPERINI, E., GHISLETTI, S. & NATOLI, G. 2013. Latent enhancers activated by stimulation in differentiated cells. *Cell,* 152**,** 157-71.

PAI, M. T., TZENG, S. R., KOVACS, J. J., KEATON, M. A., LI, S. S., YAO, T. P. & ZHOU, P. 2007. Solution structure of the Ubp-M BUZ domain, a highly specific protein module that recognizes the C-terminal tail of free ubiquitin. *J Mol Biol,* 370**,** 290-302.

PAIJO, J., DORING, M., SPANIER, J., GRABSKI, E., NOORUZZAMAN, M., SCHMIDT, T., WITTE, G., MESSERLE, M., HORNUNG, V., KAEVER, V. & KALINKE, U. 2016. cGAS Senses Human Cytomegalovirus and Induces Type I Interferon Responses in Human Monocyte-Derived Cells. *PLoS Pathog,* 12**,** e1005546.

PAPALEXI, E. & SATIJA, R. 2017. Single-cell RNA sequencing to explore immune cell heterogeneity. *Nat Rev Immunol*.

PARK, E. K., JUNG, H. S., YANG, H. I., YOO, M. C., KIM, C. & KIM, K. S. 2007. Optimized THP-1 differentiation is required for the detection of responses to weak stimuli. *Inflamm Res,* 56**,** 45-50.

PENNACCHIO, L. A., BICKMORE, W., DEAN, A., NOBREGA, M. A. & BEJERANO, G. 2013. Enhancers: five essential questions. *Nat Rev Genet,* 14**,** 288-95.

POPOVA, T., HEBERT, L., JACQUEMIN, V., GAD, S., CAUX-MONCOUTIER, V., DUBOIS-D'ENGHIEN, C., RICHAUDEAU, B., RENAUDIN, X., SELLERS, J., NICOLAS, A., SASTRE-GARAU, X., DESJARDINS, L., GYAPAY, G., RAYNAL, V., SINILNIKOVA, O. M., ANDRIEU, N., MANIE, E., DE PAUW, A., GESTA, P., BONADONA, V., MAUGARD, C. M., PENET, C., AVRIL, M. F., BARILLOT, E., CABARET, O., DELATTRE, O., RICHARD, S., CARON, O., BENFODDA, M., HU, H. H., SOUFIR, N., BRESSAC-DE PAILLERETS, B., STOPPA-LYONNET, D. & STERN, M. H. 2013. Germline BAP1 mutations predispose to renal cell carcinomas. *Am J Hum Genet,* 92**,** 974-80.

POULAIN, S., KATO, S., ARNAUD, O., MORLIGHEM, J. E., SUZUKI, M., PLESSY, C. & HARBERS, M. 2017. NanoCAGE: A Method for the Analysis of Coding and Noncoding 5'-Capped Transcriptomes. *Methods Mol Biol,* 1543**,** 57-109.

PRASAD, P., RONNERBLAD, M., ARNER, E., ITOH, M., KAWAJI, H., LASSMANN, T., DAUB, C. O., FORREST, A. R., LENNARTSSON, A., EKWALL, K. &

CONSORTIUM, F. 2014. High-throughput transcription profiling identifies putative epigenetic regulators of hematopoiesis. *Blood,* 123**,** e46-57.
PRZYBYLSKI, S., GASCH, M., MARSCHNER, A., EBERT, M., EWE, A., HELMIG, G., HILGER, N., FRICKE, S., RUDZOK, S., AIGNER, A. & BURKHARDT, J. 2017. Influence of nanoparticle-mediated transfection on proliferation of primary immune cells in vitro and in vivo. *PLoS One,* 12**,** e0176517.
QIAN, Y., WANG, B., MA, A., ZHANG, L., XU, G., DING, Q., JING, T., WU, L., LIU, Y., YANG, Z. & LIU, Y. 2016. USP16 Downregulation by Carboxyl-terminal Truncated HBx Promotes the Growth of Hepatocellular Carcinoma Cells. *Sci Rep,* 6**,** 33039.
QIN, Z. 2012. The use of THP-1 cells as a model for mimicking the function and regulation of monocytes and macrophages in the vasculature. *Atherosclerosis,* 221**,** 2-11.
RAI, K., PILARSKI, R., CEBULLA, C. M. & ABDEL-RAHMAN, M. H. 2016. Comprehensive review of BAP1 tumor predisposition syndrome with report of two new cases. *Clin Genet,* 89**,** 285-94.
RAMACHANDRAN, S., HADDAD, D., LI, C., LE, M. X., LING, A. K., SO, C. C., NEPAL, R. M., GOMMERMAN, J. L., YU, K., KETELA, T., MOFFAT, J. & MARTIN, A. 2016. The SAGA Deubiquitination Module Promotes DNA Repair and Class Switch Recombination through ATM and DNAPK-Mediated gammaH2AX Formation. *Cell Rep,* 15**,** 1554-1565.
RAN, F. A., HSU, P. D., WRIGHT, J., AGARWALA, V., SCOTT, D. A. & ZHANG, F. 2013. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc,* 8**,** 2281-2308.
RANTAKARI, P., JAPPINEN, N., LOKKA, E., MOKKALA, E., GERKE, H., PEUHU, E., IVASKA, J., ELIMA, K., AUVINEN, K. & SALMI, M. 2016. Fetal liver endothelium regulates the seeding of tissue-resident macrophages. *Nature,* 538**,** 392-396.
RICHMOND, T. J. & DAVEY, C. A. 2003. The structure of DNA in the nucleosome core. *Nature,* 423**,** 145-50.
RICHTER, E., VENTZ, K., HARMS, M., MOSTERTZ, J. & HOCHGRAFE, F. 2016. Induction of Macrophage Function in Human THP-1 Cells Is Associated with Rewiring of MAPK Signaling and Activation of MAP3K7 (TAK1) Protein Kinase. *Front Cell Dev Biol,* 4**,** 21.
ROBINSON, M. D., MCCARTHY, D. J. & SMYTH, G. K. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics,* 26**,** 139-40.
RODRIGUEZ, J. M., MAIETTA, P., EZKURDIA, I., PIETRELLI, A., WESSELINK, J. J., LOPEZ, G., VALENCIA, A. & TRESS, M. L. 2013. APPRIS: annotation of principal and alternative splice isoforms. *Nucleic Acids Res,* 41**,** D110-7.
RODRIGUEZ, L., DIEGO-ALVAREZ, D., LORDA-SANCHEZ, I., GALLARDO, F. L., MARTINEZ-FERNANDEZ, M. L., ARROYO-MUNOZ, M. E. & MARTINEZ-FRIAS, M. L. 2008. A small and active ring X chromosome in a female with features of Kabuki syndrome. *Am J Med Genet A,* 146A**,** 2816-21.
ROJO, R., PRIDANS, C., LANGLAIS, D. & HUME, D. A. 2017. Transcriptional mechanisms that control expression of the macrophage colony-stimulating factor receptor locus. *Clin Sci (Lond),* 131**,** 2161-2182.
ROUILLARD, A. D., GUNDERSEN, G. W., FERNANDEZ, N. F., WANG, Z., MONTEIRO, C. D., MCDERMOTT, M. G. & MA'AYAN, A. 2016. The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins. *Database (Oxford),* 2016.
ROY, A. L., CARRUTHERS, C., GUTJAHR, T. & ROEDER, R. G. 1993. Direct role for Myc in transcription initiation mediated by interactions with TFII-I. *Nature,* 365**,** 359-61.
ROY, N. C., ALTERMANN, E., PARK, Z. A. & MCNABB, W. C. 2011. A comparison of analog and Next-Generation transcriptomic tools for mammalian studies. *Brief Funct Genomics,* 10**,** 135-50.

SAEED, S., QUINTIN, J., KERSTENS, H. H., RAO, N. A., AGHAJANIREFAH, A., MATARESE, F., CHENG, S. C., RATTER, J., BERENTSEN, K., VAN DER ENT, M. A., SHARIFI, N., JANSSEN-MEGENS, E. M., TER HUURNE, M., MANDOLI, A., VAN SCHAIK, T., NG, A., BURDEN, F., DOWNES, K., FRONTINI, M., KUMAR, V., GIAMARELLOS-BOURBOULIS, E. J., OUWEHAND, W. H., VAN DER MEER, J. W., JOOSTEN, L. A., WIJMENGA, C., MARTENS, J. H., XAVIER, R. J., LOGIE, C., NETEA, M. G. & STUNNENBERG, H. G. 2014. Epigenetic programming of monocyte-to-macrophage differentiation and trained innate immunity. *Science,* 345**,** 1251086.

SANDELIN, A., ALKEMA, W., ENGSTROM, P., WASSERMAN, W. W. & LENHARD, B. 2004. JASPAR: an open-access database for eukaryotic transcription factor binding profiles. *Nucleic Acids Res,* 32**,** D91-4.

SAXONOV, S., BERG, P. & BRUTLAG, D. L. 2006. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc Natl Acad Sci U S A,* 103**,** 1412-7.

SCHAEFER, K. A., WU, W. H., COLGAN, D. F., TSANG, S. H., BASSUK, A. G. & MAHAJAN, V. B. 2017. Unexpected mutations after CRISPR-Cas9 editing in vivo. *Nat Methods,* 14**,** 547-548.

SCHEUERMANN, J. C., DE AYALA ALONSO, A. G., OKTABA, K., LY-HARTIG, N., MCGINTY, R. K., FRATERMAN, S., WILM, M., MUIR, T. W. & MULLER, J. 2010. Histone H2A deubiquitinase activity of the Polycomb repressive complex PR-DUB. *Nature,* 465**,** 243-7.

SCHMIDL, C., RENNER, K., PETER, K., EDER, R., LASSMANN, T., BALWIERZ, P. J., ITOH, M., NAGAO-SATO, S., KAWAJI, H., CARNINCI, P., SUZUKI, H., HAYASHIZAKI, Y., ANDREESEN, R., HUME, D. A., HOFFMANN, P., FORREST, A. R., KREUTZ, M. P., EDINGER, M., REHLI, M. & CONSORTIUM, F. 2014. Transcription and enhancer profiling in human monocyte subsets. *Blood,* 123**,** e90-9.

SCHMIDT, B. & HILDEBRANDT, A. 2017. Next-generation sequencing: big data meets high performance computing. *Drug Discov Today,* 22**,** 712-717.

SCHMIDT, T., SCHMID-BURGK, J. L., EBERT, T. S., GAIDT, M. M. & HORNUNG, V. 2016. Designer Nuclease-Mediated Generation of Knockout THP1 Cells. *Methods Mol Biol,* 1338**,** 261-72.

SCHNOOR, M., CULLEN, P., LORKOWSKI, J., STOLLE, K., ROBENEK, H., TROYER, D., RAUTERBERG, J. & LORKOWSKI, S. 2008. Production of type VI collagen by human macrophages: a new dimension in macrophage functional heterogeneity. *J Immunol,* 180**,** 5707-19.

SCHOUMACHER, M., LE CORRE, S., HOUY, A., MULUGETA, E., STERN, M. H., ROMAN-ROMAN, S. & MARGUERON, R. 2016. Uveal melanoma cells are resistant to EZH2 inhibition regardless of BAP1 status. *Nat Med,* 22**,** 577-8.

SCHRECENGOST, R. S., DEAN, J. L., GOODWIN, J. F., SCHIEWER, M. J., URBAN, M. W., STANEK, T. J., SUSSMAN, R. T., HICKS, J. L., BIRBE, R. C., DRAGANOVA-TACHEVA, R. A., VISAKORPI, T., DEMARZO, A. M., MCMAHON, S. B. & KNUDSEN, K. E. 2014. USP22 regulates oncogenic signaling pathways to drive lethal cancer progression. *Cancer Res,* 74**,** 272-86.

SCHROEDER, A., MUELLER, O., STOCKER, S., SALOWSKY, R., LEIBER, M., GASSMANN, M., LIGHTFOOT, S., MENZEL, W., GRANZOW, M. & RAGG, T. 2006. The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Mol Biol,* 7**,** 3.

SCHUETTENGRUBER, B., CHOURROUT, D., VERVOORT, M., LEBLANC, B. & CAVALLI, G. 2007. Genome regulation by polycomb and trithorax proteins. *Cell,* 128**,** 735-45.

SCHWENDE, H., FITZKE, E., AMBS, P. & DIETER, P. 1996. Differences in the state of differentiation of THP-1 cells induced by phorbol ester and 1,25-dihydroxyvitamin D3. *J Leukoc Biol,* 59**,** 555-61.

SEVERIN, J., LIZIO, M., HARSHBARGER, J., KAWAJI, H., DAUB, C. O., HAYASHIZAKI, Y., CONSORTIUM, F., BERTIN, N. & FORREST, A. R. 2014. Interactive visualization and analysis of large-scale sequencing datasets using ZENBU. *Nat Biotechnol,* 32**,** 217-9.

SHARMA, N., ZHU, Q., WANI, G., HE, J., WANG, Q. E. & WANI, A. A. 2014. USP3 counteracts RNF168 via deubiquitinating H2A and gammaH2AX at lysine 13 and 15. *Cell Cycle,* 13**,** 106-14.

SHARP, A. J., STATHAKI, E., MIGLIAVACCA, E., BRAHMACHARY, M., MONTGOMERY, S. B., DUPRE, Y. & ANTONARAKIS, S. E. 2011. DNA methylation profiles of human active and inactive X chromosomes. *Genome Res,* 21**,** 1592-600.

SHI, C., SAKUMA, M., MOOROKA, T., LISCOE, A., GAO, H., CROCE, K. J., SHARMA, A., KAPLAN, D., GREAVES, D. R., WANG, Y. & SIMON, D. I. 2008. Down-regulation of the forkhead transcription factor Foxp1 is required for monocyte differentiation and macrophage function. *Blood,* 112**,** 4699-711.

SHPARGEL, K. B., SENGOKU, T., YOKOYAMA, S. & MAGNUSON, T. 2012. UTX and UTY demonstrate histone demethylase-independent function in mouse embryonic development. *PLoS Genet,* 8**,** e1002964.

SMITH, C. E. L., ALEXANDRAKI, A., CORDERY, S. F., PARMAR, R., BONTHRON, D. T. & VALLELEY, E. M. A. 2017. A tissue-specific promoter derived from a SINE retrotransposon drives biallelic expression of PLAGL1 in human lymphocytes. *PLoS One,* 12**,** e0185678.

SPENGLER, D., VILLALBA, M., HOFFMANN, A., PANTALONI, C., HOUSSAMI, S., BOCKAERT, J. & JOURNOT, L. 1997. Regulation of apoptosis and cell cycle arrest by Zac1, a novel zinc finger protein expressed in the pituitary gland and the brain. *EMBO J,* 16**,** 2814-25.

SPITZ, F. & FURLONG, E. E. 2012. Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet,* 13**,** 613-26.

STAZI, G., ZWERGEL, C., MAI, A. & VALENTE, S. 2017. EZH2 inhibitors: a patent review (2014-2016). *Expert Opin Ther Pat,* 27**,** 797-813.

STOCK, J. K., GIADROSSI, S., CASANOVA, M., BROOKES, E., VIDAL, M., KOSEKI, H., BROCKDORFF, N., FISHER, A. G. & POMBO, A. 2007. Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. *Nat Cell Biol,* 9**,** 1428-35.

STRUHL, K. & SEGAL, E. 2013. Determinants of nucleosome positioning. *Nat Struct Mol Biol,* 20**,** 267-73.

SUGIMOTO, K., TOYOSHIMA, H., SAKAI, R., MIYAGAWA, K., HAGIWARA, K., ISHIKAWA, F., TAKAKU, F., YAZAKI, Y. & HIRAI, H. 1992. Frequent mutations in the p53 gene in human myeloid leukemia cell lines. *Blood,* 79**,** 2378-83.

SUNDSTROM, C. & NILSSON, K. 1976. Establishment and characterization of a human histiocytic lymphoma cell line (U-937). *Int J Cancer,* 17**,** 565-77.

SUZUKI, H., FORREST, A. R., VAN NIMWEGEN, E., DAUB, C. O., BALWIERZ, P. J., IRVINE, K. M., LASSMANN, T., RAVASI, T., HASEGAWA, Y., DE HOON, M. J., KATAYAMA, S., SCHRODER, K., CARNINCI, P., TOMARU, Y., KANAMORI-KATAYAMA, M., KUBOSAKI, A., AKALIN, A., ANDO, Y., ARNER, E., ASADA, M., ASAHARA, H., BAILEY, T., BAJIC, V. B., BAUER, D., BECKHOUSE, A. G., BERTIN, N., BJORKEGREN, J., BROMBACHER, F., BULGER, E., CHALK, A. M., CHIBA, J., CLOONAN, N., DAWE, A., DOSTIE, J., ENGSTROM, P. G., ESSACK, M., FAULKNER, G. J., FINK, J. L., FREDMAN, D., FUJIMORI, K., FURUNO, M., GOJOBORI, T., GOUGH, J., GRIMMOND, S. M.,

GUSTAFSSON, M., HASHIMOTO, M., HASHIMOTO, T., HATAKEYAMA, M., HEINZEL, S., HIDE, W., HOFMANN, O., HORNQUIST, M., HUMINIECKI, L., IKEO, K., IMAMOTO, N., INOUE, S., INOUE, Y., ISHIHARA, R., IWAYANAGI, T., JACOBSEN, A., KAUR, M., KAWAJI, H., KERR, M. C., KIMURA, R., KIMURA, S., KIMURA, Y., KITANO, H., KOGA, H., KOJIMA, T., KONDO, S., KONNO, T., KROGH, A., KRUGER, A., KUMAR, A., LENHARD, B., LENNARTSSON, A., LINDOW, M., LIZIO, M., MACPHERSON, C., MAEDA, N., MAHER, C. A., MAQUNGO, M., MAR, J., MATIGIAN, N. A., MATSUDA, H., MATTICK, J. S., MEIER, S., MIYAMOTO, S., MIYAMOTO-SATO, E., NAKABAYASHI, K., NAKACHI, Y., NAKANO, M., NYGAARD, S., OKAYAMA, T., OKAZAKI, Y., OKUDA-YABUKAMI, H., ORLANDO, V., OTOMO, J., PACHKOV, M., PETROVSKY, N., *et al*. 2009. The transcriptional network that controls growth arrest and differentiation in a human myeloid leukemia cell line. *Nat Genet,* 41**,** 553-62.

TAKAHASHI, H., LASSMANN, T., MURATA, M. & CARNINCI, P. 2012. 5′ end–centered expression profiling using cap-analysis
gene expression and next-generation sequencing. *Nature Protocols,* 7**,** 542-561.

TANAKA, K., TANAKA, T., OGAWA, S., KUROKAWA, M., MITANI, K., YAZAKI, Y. & HIRAI, H. 1995. Increased expression of AML1 during retinoic-acid-induced differentiation of U937 cells. *Biochem Biophys Res Commun,* 211**,** 1023-30.

TANG, L. J., LI, Y., LIU, Y. L., WANG, J. M., LIU, D. W. & TIAN, Q. B. 2016. USP12 regulates cell cycle progression by involving c-Myc, cyclin D2 and BMI-1. *Gene,* 578**,** 92-9.

TANG, Z., SHU, H., ONCEL, D., CHEN, S. & YU, H. 2004. Phosphorylation of Cdc20 by Bub1 provides a catalytic mechanism for APC/C inhibition by the spindle checkpoint. *Mol Cell,* 16**,** 387-97.

TARASOVA, N. K., YTTERBERG, A. J., LUNDBERG, K., ZHANG, X. M., HARRIS, R. A. & ZUBAREV, R. A. 2015. Proteomics Reveals a Role for Attachment in Monocyte Differentiation into Efficient Proinflammatory Macrophages. *J Proteome Res,* 14**,** 3940-7.

TARCA, A. L., ROMERO, R. & DRAGHICI, S. 2006. Analysis of microarray experiments of gene expression profiling. *Am J Obstet Gynecol,* 195**,** 373-88.

TESTA, J. R., CHEUNG, M., PEI, J., BELOW, J. E., TAN, Y., SEMENTINO, E., COX, N. J., DOGAN, A. U., PASS, H. I., TRUSA, S., HESDORFFER, M., NASU, M., POWERS, A., RIVERA, Z., COMERTPAY, S., TANJI, M., GAUDINO, G., YANG, H. & CARBONE, M. 2011. Germline BAP1 mutations predispose to malignant mesothelioma. *Nat Genet,* 43**,** 1022-5.

THERMOFISHER 2017. Transcriptome Analysis Console (TAC) Software.

THIEME, S., GYARFAS, T., RICHTER, C., OZHAN, G., FU, J., ALEXOPOULOU, D., MUDERS, M. H., MICHALK, I., JAKOB, C., DAHL, A., KLINK, B., BANDOLA, J., BACHMANN, M., SCHROCK, E., BUCHHOLZ, F., STEWART, A. F., WEIDINGER, G., ANASTASSIADIS, K. & BRENNER, S. 2013. The histone demethylase UTX regulates stem cell migration and hematopoiesis. *Blood,* 121**,** 2462-73.

THOMPSON, J. F. & STEINMANN, K. E. 2010. Single molecule sequencing with a HeliScope genetic analysis system. *Curr Protoc Mol Biol,* Chapter 7**,** Unit7 10.

THUONG, N. T., HAWN, T. R., CHAU, T. T., BANG, N. D., YEN, N. T., THWAITES, G. E., TEO, Y. Y., SEIELSTAD, M., HIBBERD, M., LAN, N. T., CAWS, M., FARRAR, J. J. & DUNSTAN, S. J. 2012. Epiregulin (EREG) variation is associated with susceptibility to tuberculosis. *Genes Immun,* 13**,** 275-81.

TRAORE, K., TRUSH, M. A., GEORGE, M., JR., SPANNHAKE, E. W., ANDERSON, W. & ASSEFFA, A. 2005. Signal transduction of phorbol 12-myristate 13-acetate

(PMA)-induced growth inhibition of human monocytic leukemia THP-1 cells is reactive oxygen dependent. *Leuk Res,* 29**,** 863-79.

TRINKLEIN, N. D., ALDRED, S. F., HARTMAN, S. J., SCHROEDER, D. I., OTILLAR, R. P. & MYERS, R. M. 2004. An abundance of bidirectional promoters in the human genome. *Genome Res,* 14**,** 62-6.

TROEGELER, A., LASTRUCCI, C., DUVAL, C., TANNE, A., COUGOULE, C., MARIDONNEAU-PARINI, I., NEYROLLES, O. & LUGO-VILLARINO, G. 2014. An efficient siRNA-mediated gene silencing in primary human monocytes, dendritic cells and macrophages. *Immunol Cell Biol,* 92**,** 699-708.

TROLLE, C., NIELSEN, M. M., SKAKKEBAEK, A., LAMY, P., VANG, S., HEDEGAARD, J., NORDENTOFT, I., ORNTOFT, T. F., PEDERSEN, J. S. & GRAVHOLT, C. H. 2016. Widespread DNA hypomethylation and differential gene expression in Turner syndrome. *Sci Rep,* 6**,** 34220.

TSUCHIYA, S., YAMABE, M., YAMAGUCHI, Y., KOBAYASHI, Y., KONNO, T. & TADA, K. 1980. Establishment and characterization of a human acute monocytic leukemia cell line (THP-1). *Int J Cancer,* 26**,** 171-6.

URBE, S., LIU, H., HAYES, S. D., HERIDE, C., RIGDEN, D. J. & CLAGUE, M. J. 2012. Systematic survey of deubiquitinase localization identifies USP21 as a regulator of centrosome- and microtubule-associated functions. *Mol Biol Cell,* 23**,** 1095-103.

VALLELEY, E. M., CORDERY, S. F. & BONTHRON, D. T. 2007. Tissue-specific imprinting of the ZAC/PLAGL1 tumour suppressor gene results from variable utilization of monoallelic and biallelic promoters. *Hum Mol Genet,* 16**,** 972-81.

VAN DER MEULEN, J., SPELEMAN, F. & VAN VLIERBERGHE, P. 2014. The H3K27me3 demethylase UTX in normal development and disease. *Epigenetics,* 9**,** 658-68.

VAN HAAFTEN, G., DALGLIESH, G. L., DAVIES, H., CHEN, L., BIGNELL, G., GREENMAN, C., EDKINS, S., HARDY, C., O'MEARA, S., TEAGUE, J., BUTLER, A., HINTON, J., LATIMER, C., ANDREWS, J., BARTHORPE, S., BEARE, D., BUCK, G., CAMPBELL, P. J., COLE, J., FORBES, S., JIA, M., JONES, D., KOK, C. Y., LEROY, C., LIN, M. L., MCBRIDE, D. J., MADDISON, M., MAQUIRE, S., MCLAY, K., MENZIES, A., MIRONENKO, T., MULDERRIG, L., MUDIE, L., PLEASANCE, E., SHEPHERD, R., SMITH, R., STEBBINGS, L., STEPHENS, P., TANG, G., TARPEY, P. S., TURNER, R., TURRELL, K., VARIAN, J., WEST, S., WIDAA, S., WRAY, P., COLLINS, V. P., ICHIMURA, K., LAW, S., WONG, J., YUEN, S. T., LEUNG, S. Y., TONON, G., DEPINHO, R. A., TAI, Y. T., ANDERSON, K. C., KAHNOSKI, R. J., MASSIE, A., KHOO, S. K., TEH, B. T., STRATTON, M. R. & FUTREAL, P. A. 2009. Somatic mutations of the histone H3K27 demethylase gene UTX in human cancer. *Nat Genet,* 41**,** 521-3.

VARRAULT, A., CIANI, E., APIOU, F., BILANGES, B., HOFFMANN, A., PANTALONI, C., BOCKAERT, J., SPENGLER, D. & JOURNOT, L. 1998. hZAC encodes a zinc finger protein with antiproliferative properties and maps to a chromosomal region frequently lost in cancer. *Proc Natl Acad Sci U S A,* 95**,** 8835-40.

VEYRUNES, F., WATERS, P. D., MIETHKE, P., RENS, W., MCMILLAN, D., ALSOP, A. E., GRUTZNER, F., DEAKIN, J. E., WHITTINGTON, C. M., SCHATZKAMER, K., KREMITZKI, C. L., GRAVES, T., FERGUSON-SMITH, M. A., WARREN, W. & MARSHALL GRAVES, J. A. 2008. Bird-like sex chromosomes of platypus imply recent origin of mammal sex chromosomes. *Genome Res,* 18**,** 965-73.

WADDELL, N., PAJIC, M., PATCH, A. M., CHANG, D. K., KASSAHN, K. S., BAILEY, P., JOHNS, A. L., MILLER, D., NONES, K., QUEK, K., QUINN, M. C., ROBERTSON, A. J., FADLULLAH, M. Z., BRUXNER, T. J., CHRIST, A. N., HARLIWONG, I., IDRISOGLU, S., MANNING, S., NOURSE, C., NOURBAKHSH, E., WANI, S., WILSON, P. J., MARKHAM, E., CLOONAN, N., ANDERSON, M. J., FINK, J. L., HOLMES, O., KAZAKOFF, S. H., LEONARD, C.,

NEWELL, F., POUDEL, B., SONG, S., TAYLOR, D., WADDELL, N., WOOD, S., XU, Q., WU, J., PINESE, M., COWLEY, M. J., LEE, H. C., JONES, M. D., NAGRIAL, A. M., HUMPHRIS, J., CHANTRILL, L. A., CHIN, V., STEINMANN, A. M., MAWSON, A., HUMPHREY, E. S., COLVIN, E. K., CHOU, A., SCARLETT, C. J., PINHO, A. V., GIRY-LATERRIERE, M., ROOMAN, I., SAMRA, J. S., KENCH, J. G., PETTITT, J. A., MERRETT, N. D., TOON, C., EPARI, K., NGUYEN, N. Q., BARBOUR, A., ZEPS, N., JAMIESON, N. B., GRAHAM, J. S., NICLOU, S. P., BJERKVIG, R., GRUTZMANN, R., AUST, D., HRUBAN, R. H., MAITRA, A., IACOBUZIO-DONAHUE, C. A., WOLFGANG, C. L., MORGAN, R. A., LAWLOR, R. T., CORBO, V., BASSI, C., FALCONI, M., ZAMBONI, G., TORTORA, G., TEMPERO, M. A., AUSTRALIAN PANCREATIC CANCER GENOME, I., GILL, A. J., ESHLEMAN, J. R., PILARSKY, C., SCARPA, A., MUSGROVE, E. A., PEARSON, J. V., BIANKIN, A. V. & GRIMMOND, S. M. 2015. Whole genomes redefine the mutational landscape of pancreatic cancer. *Nature,* 518**,** 495-501.

WALLNER, S., SCHRODER, C., LEITAO, E., BERULAVA, T., HAAK, C., BEISSER, D., RAHMANN, S., RICHTER, A. S., MANKE, T., BONISCH, U., ARRIGONI, L., FROHLER, S., KLIRONOMOS, F., CHEN, W., RAJEWSKY, N., MULLER, F., EBERT, P., LENGAUER, T., BARANN, M., ROSENSTIEL, P., GASPARONI, G., NORDSTROM, K., WALTER, J., BRORS, B., ZIPPRICH, G., FELDER, B., KLEIN-HITPASS, L., ATTENBERGER, C., SCHMITZ, G. & HORSTHEMKE, B. 2016. Epigenetic dynamics of monocyte-to-macrophage differentiation. *Epigenetics Chromatin,* 9**,** 33.

WALPORT, L. J., HOPKINSON, R. J., VOLLMAR, M., MADDEN, S. K., GILEADI, C., OPPERMANN, U., SCHOFIELD, C. J. & JOHANSSON, C. 2014. Human UTY(KDM6C) is a male-specific N-methyl lysyl demethylase. *J Biol Chem,* 289**,** 18302-13.

WANG, J., QIAN, J., HU, Y., KONG, X., CHEN, H., SHI, Q., JIANG, L., WU, C., ZOU, W., CHEN, Y., XU, J. & FANG, J. Y. 2014. ArhGAP30 promotes p53 acetylation and function in colorectal cancer. *Nat Commun,* 5**,** 4735.

WATKINS, N. A., GUSNANTO, A., DE BONO, B., DE, S., MIRANDA-SAAVEDRA, D., HARDIE, D. L., ANGENENT, W. G., ATTWOOD, A. P., ELLIS, P. D., ERBER, W., FOAD, N. S., GARNER, S. F., ISACKE, C. M., JOLLEY, J., KOCH, K., MACAULAY, I. C., MORLEY, S. L., RENDON, A., RICE, K. M., TAYLOR, N., THIJSSEN-TIMMER, D. C., TIJSSEN, M. R., VAN DER SCHOOT, C. E., WERNISCH, L., WINZER, T., DUDBRIDGE, F., BUCKLEY, C. D., LANGFORD, C. F., TEICHMANN, S., GOTTGENS, B., OUWEHAND, W. H. & BLOODOMICS, C. 2009. A HaemAtlas: characterizing gene expression in differentiated human blood cells. *Blood,* 113**,** e1-9.

WATSON, J. V., CHAMBERS, S. H. & SMITH, P. J. 1987. A pragmatic approach to the analysis of DNA histograms with a definable G1 peak. *Cytometry,* 8**,** 1-8.

WELSTEAD, G. G., CREYGHTON, M. P., BILODEAU, S., CHENG, A. W., MARKOULAKI, S., YOUNG, R. A. & JAENISCH, R. 2012. X-linked H3K27me3 demethylase Utx is required for embryonic development in a sex-specific manner. *Proc Natl Acad Sci U S A,* 109**,** 13004-9.

WHYTE, W. A., ORLANDO, D. A., HNISZ, D., ABRAHAM, B. J., LIN, C. Y., KAGEY, M. H., RAHL, P. B., LEE, T. I. & YOUNG, R. A. 2013. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell,* 153**,** 307-19.

WIESNER, T., OBENAUF, A. C., MURALI, R., FRIED, I., GRIEWANK, K. G., ULZ, P., WINDPASSINGER, C., WACKERNAGEL, W., LOY, S., WOLF, I., VIALE, A., LASH, A. E., PIRUN, M., SOCCI, N. D., RUTTEN, A., PALMEDO, G., ABRAMSON, D., OFFIT, K., OTT, A., BECKER, J. C., CERRONI, L., KUTZNER,

H., BASTIAN, B. C. & SPEICHER, M. R. 2011. Germline mutations in BAP1 predispose to melanocytic tumors. *Nat Genet,* 43**,** 1018-21.
WILLAERT, A., KHATRI, S., CALLEWAERT, B. L., COUCKE, P. J., CROSBY, S. D., LEE, J. G., DAVIS, E. C., SHIVA, S., TSANG, M., DE PAEPE, A. & URBAN, Z. 2012. GLUT10 is required for the development of the cardiovascular system and the notochord and connects mitochondrial function to TGFbeta signaling. *Hum Mol Genet,* 21**,** 1248-59.
WILLIAMS, K., CHRISTENSEN, J., RAPPSILBER, J., NIELSEN, A. L., JOHANSEN, J. V. & HELIN, K. 2014. The histone lysine demethylase JMJD3/KDM6B is recruited to p53 bound promoters and enhancer elements in a p53 dependent manner. *PLoS One,* 9**,** e96545.
WRIGHT, S. D., RAMOS, R. A., TOBIAS, P. S., ULEVITCH, R. J. & MATHISON, J. C. 1990. CD14, a receptor for complexes of lipopolysaccharide (LPS) and LPS binding protein. *Science,* 249**,** 1431-3.
WU, C., JIN, X., TSUENG, G., AFRASIABI, C. & SU, A. I. 2016. BioGPS: building your own mash-up of gene annotations and expression profiles. *Nucleic Acids Res,* 44**,** D313-6.
XU, J., DENG, X., WATKINS, R. & DISTECHE, C. M. 2008. Sex-specific differences in expression of histone demethylases Utx and Uty in mouse brain and neurons. *J Neurosci,* 28**,** 4521-7.
XU, Y., YANG, H., JOO, H. Y., YU, J. H., SMITH, A. D. T., SCHNEIDER, D., CHOW, L. T., RENFROW, M. & WANG, H. 2013. Ubp-M serine 552 phosphorylation by cyclin-dependent kinase 1 regulates cell cycle progression. *Cell Cycle,* 12**,** 3219-27.
XUN, J., WANG, D., SHEN, L., GONG, J., GAO, R., DU, L., CHANG, A., SONG, X., XIANG, R. & TAN, X. 2017. JMJD3 suppresses stem cell-like characteristics in breast cancer cells by downregulation of Oct4 independently of its demethylase activity. *Oncotarget,* 8**,** 21918-21929.
YAN, Q., SUN, L., ZHU, Z., WANG, L., LI, S. & YE, R. D. 2014. Jmjd3-mediated epigenetic regulation of inflammatory cytokine gene expression in serum amyloid A-stimulated macrophages. *Cell Signal,* 26**,** 1783-91.
YANG, L., DAI, F., TANG, L., LE, Y. & YAO, W. 2017. Macrophage differentiation induced by PMA is mediated by activation of RhoA/ROCK signaling. *J Toxicol Sci,* 42**,** 763-771.
YANG, W., LEE, Y. H., JONES, A. E., WOOLNOUGH, J. L., ZHOU, D., DAI, Q., WU, Q., GILES, K. E., TOWNES, T. M. & WANG, H. 2014. The histone H2A deubiquitinase Usp16 regulates embryonic stem cell gene expression and lineage commitment. *Nat Commun,* 5**,** 3818.
YATES, A., AKANNI, W., AMODE, M. R., BARRELL, D., BILLIS, K., CARVALHO-SILVA, D., CUMMINS, C., CLAPHAM, P., FITZGERALD, S., GIL, L., GIRON, C. G., GORDON, L., HOURLIER, T., HUNT, S. E., JANACEK, S. H., JOHNSON, N., JUETTEMANN, T., KEENAN, S., LAVIDAS, I., MARTIN, F. J., MAUREL, T., MCLAREN, W., MURPHY, D. N., NAG, R., NUHN, M., PARKER, A., PATRICIO, M., PIGNATELLI, M., RAHTZ, M., RIAT, H. S., SHEPPARD, D., TAYLOR, K., THORMANN, A., VULLO, A., WILDER, S. P., ZADISSA, A., BIRNEY, E., HARROW, J., MUFFATO, M., PERRY, E., RUFFIER, M., SPUDICH, G., TREVANION, S. J., CUNNINGHAM, F., AKEN, B. L., ZERBINO, D. R. & FLICEK, P. 2016. Ensembl 2016. *Nucleic Acids Res,* 44**,** D710-6.
YOSHIMI, A. & KUROKAWA, M. 2011. Key roles of histone methyltransferase and demethylase in leukemogenesis. *J Cell Biochem,* 112**,** 415-24.
YU, H., PAK, H., HAMMOND-MARTEL, I., GHRAM, M., RODRIGUE, A., DAOU, S., BARBOUR, H., CORBEIL, L., HEBERT, J., DROBETSKY, E., MASSON, J. Y., DI NOIA, J. M. & AFFAR EL, B. 2014. Tumor suppressor and deubiquitinase BAP1 promotes DNA double-strand break repair. *Proc Natl Acad Sci U S A,* 111**,** 285-90.

ZHANG, D. E., HETHERINGTON, C. J., CHEN, H. M. & TENEN, D. G. 1994. The macrophage transcription factor PU.1 directs tissue-specific expression of the macrophage colony-stimulating factor receptor. *Mol Cell Biol,* 14**,** 373-81.

ZHANG, J. D., KOERNER, C., BECHTEL, S., BENDER, C., KEKLIKOGLOU, I., SCHMIDT, C., IRSIGLER, A., ERNST, U., SAHIN, O., WIEMANN, S. & TSCHULENA, U. 2011. Time-resolved human kinome RNAi screen identifies a network regulating mitotic-events as early regulators of cell proliferation. *PLoS One,* 6**,** e22176.

ZHANG, X., CHOI, P. S., FRANCIS, J. M., IMIELINSKI, M., WATANABE, H., CHERNIACK, A. D. & MEYERSON, M. 2016. Identification of focally amplified lineage-specific super-enhancers in human epithelial cancers. *Nat Genet,* 48**,** 176-82.

ZHANG, Z., YANG, H. & WANG, H. 2014. The histone H2A deubiquitinase USP16 interacts with HERC2 and fine-tunes cellular response to DNA damage. *J Biol Chem,* 289**,** 32883-94.

ZHAO, X., VALEN, E., PARKER, B. J. & SANDELIN, A. 2011. Systematic clustering of transcription start site landscapes. *PLoS One,* 6**,** e23409.

ZHOU, Y. & NESS, S. A. 2011. Myb proteins: angels and demons in normal and transformed cells. *Front Biosci (Landmark Ed),* 16**,** 1109-31.

ZIEGLER-HEITBROCK, H. W., SCHRAUT, W., WENDELGASS, P., STROBEL, M., STERNSDORF, T., WEBER, C., AEPFELBACHER, M., EHLERS, M., SCHUTT, C. & HAAS, J. G. 1994. Distinct patterns of differentiation induced in the monocytic cell line Mono Mac 6. *J Leukoc Biol,* 55**,** 73-80.

ZIEGLER-HEITBROCK, H. W., THIEL, E., FUTTERER, A., HERZOG, V., WIRTZ, A. & RIETHMULLER, G. 1988. Establishment of a human cell line (Mono Mac 6) with characteristics of mature monocytes. *Int J Cancer,* 41**,** 456-61.