The effect of hydroxyurea treatment on DNA methylation and gene expression in essential thrombocythaemia and polycythaemia vera: a cross-species study

Contreras Castillo, Stephania

Awarding institution:
King's College London

The effect of hydroxyurea treatment on DNA methylation and gene expression in essential thrombocythaemia and polycythaemia vera: a cross-species study

Submitted by

Stephania N. Contreras Castillo

September 2019

To King's College London for the degree of

Doctor of Philosophy

Department of Medical and Molecular Genetics

School of Basic & Medical Biosciences

Abstract

Myeloproliferative neoplasms (MPNs), which includes polycythaemia vera (PV) and essential thrombocythaemia (ET), are bone marrow disorders that give rise to a high production of red blood cells and platelets leading to thrombosis and haemorrhage. Hydroxyurea (HU), is the first line treatment for high-risk PV and ET patients since it effectively reduces the haematocrit, platelets and white blood cells counts. Mechanistically, HU acts by inhibiting the ribonucleotide reductase enzyme, blocking the cell cycle which can lead to cell death. However, additional effects of HU have also been observed which are unlikely related to this mechanism. Therefore, I hypothesize that HU has an influence on the epigenome, causing changes in gene expression which contribute to its therapeutic effects. For this, I assayed and analysed DNA methylation and gene expression from two differentially developed and clinically relevant cells types from MPN patients and a MPN mouse model, comparing samples prior to and following HU treatment. I observed that HU mainly changes gene expression and specifically affects DNA methylation at the stem cell level. Interestingly, several genes encoding transcription factors involved in haematopoiesis were also identified as potential mediators of HU effects in both species. Moreover, *SPI1* was found to be upregulated and differentially methylated following HU treatment. In addition, several differentially expressed genes and differentially methylated sites were enriched for SPI1 binding sites. Thus, I propose that SPI1 is involved in the pathogenesis of the disease and in the therapeutic effect of HU. Finally, I also provide a list of candidate genes to be further investigated for their role in the therapeutic effect of HU.

Acknowledgements

I would like to first thank to my supervisor Professor Rebecca Oakey for the support and trust to carry out this project. Thanks to my second supervisor Dr Reiner Schulz for the patience to teach me bioinformatics and for the critical input towards the project. Thank you to Matt, Sam and Pro for helping me in the lab.

Many thanks to Professor Claire Harrison for being always helpful and supportive in the clinical aspects of this project. Thank you to the clinical team especially Samah, Claire, Yvonne and Natalia, for helping me with the patients. Thank you to all the patients that participated and made this project meaningful.

I am also grateful to the funding received from Becas Chile (CONICYT) to undertake my PhD. Thank you to the funding bodies – King's Heatlh Partners and NIHR BRC at Guy's and St Thomas' NHS Foundation Trust – that supported the experimental work of this study. Thanks to the Genetic Society and King's College London for their support when presenting my work at international conferences.

I greatly appreciate the support received through the collaborative work undertaken with Dr Ann Mullally at Harvard Medical School, Boston (USA). Thank you to Ann for your advice and constructive comments towards the project. Thank you to Ann's research group that helped me during my studies in Boston, especially Azu and Will that taught me everything that I needed to perform my experiments. I gratefully acknowledge the funding received from the European Hematology Association that made this study in Boston possible.

Big thanks to Bertille, Lena, Brooke and Federico for your support and, more

importantly, for your friendship that made every day more enjoyable in the lab. A special double thank you to Bertille for the invaluable scientific input and encouragement through the last year that helped me to think critically about my project. Thank you to Elizabeth for kindly proofreading my thesis. Thank you to Ismael for helping me to set up the flow cytometry analysis.

I will always be grateful for meeting my great friend Marta. Thank you for listening and supporting me during all these four years. Just meeting her made it worth coming to the UK.

Eternal thanks to Matías. I definitely couldn't have done this without his support, understanding and company (and food!). And last but not least, thank you to my family. Gracias a mi mamá y papá por apoyarme y entenderme en todo lo que siempre quise hacer y ser personas a las que siempre admiraré y recordaré con orgullo. Gracias a mi hermano por ser un ejemplo que me ha inspirado desde pequeña. Gracias a mi abuelita Albertina y a mi tía Elena por ser el soporte de nuestra familia. Sin ustedes no podría haber logrado nada de esto.

Abbreviations

AML	Acute myeloid leukaemia
CBC	Cell blood counts
CD34+	Haematopoietic stem and progenitor cells
CGI	CpG island
DEGs	Differentially methylated genes
DMR	Differentially methylated region
ET	Essential thrombocythaemia
HbF	Foetal haemoglobin
HSCs	Haematopoietic stem cells
HU	Hydroxyurea
LSK	Lineage-Sca+cKit+; haematopoietic stem cell
MPNs	Myeloproliferative neoplasms
PMF	Primary myelofibrosis
PV	Polycythaemia vera
TF	Transcription factor
TN	Triple negative

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Haematopoiesis

Haematopoiesis is a dynamic process that is defined by the differentiation and self-renewal of multipotent cells that give rise to different cell lineages with various functions. In mammals, haematopoiesis occurs in two waves, the primitive wave and the definitive wave, which differ in their cell types produced and anatomical location (Galloway and Zon, 2003). The primitive wave occurs in the yolk sac and is transitory (Moore and Owen, 1965, 1967). It involves limited erythroid progenitor cells which lack renewal capacity and can only differentiate into erythrocytes and macrophages (Palis, 2001). The definitive wave occurs later in development in what is known as aorta–gonad–mesonephros, which give rises to haematopoietic stem cells (HSCs) (Medvinsky *et al.*, 1993; Ivanovs *et al.*, 2011). These cells will be responsible for the continuous production of all cell lineages during the life of the organism (Dzierzak and Speck, 2008; Medvinsky *et al.*, 2011). Haematopoiesis then progresses to the liver, spleen and finally to the bone marrow. After birth, bone marrow, and to some extent the spleen, are the main sites of haematopoiesis (Ciau-Uitz *et al.*, 2016). Although much of the information has been obtained from murine studies, it appears that the human haematopoietic system is closely related to the

mouse in terms of the sequential progression and tissues involved in haematopoiesis (Galloway and Zon, 2003).

Hematopoiesis is a hierarchical system where all cells are derived from one multipotent cell (Fig. 1.1). The HSCs are at the apex, these can self-renew and differentiate into several progenitor cells that give rise to mature blood cells (Fig. 1.1). The identification of surface markers has facilitated the characterization of the multiple cell types that make up the haematopoietic system. In mice, immature stem cells have low expression of lineage markers, expressing the stem cell antigen-1 (Sca-1) and CD117 (c-Kit), which are known as LSK (Lineage$^-$Sca$^+$cKit$^+$) (Spangrude *et al.*, 1988; Uchida, N, and Weissman, 1992). In humans, the CD34 antigen, expressed in less than 5% of all blood cells, has been widely used to identify haematopoietic stem and progenitor cells (Civin *et al.*, 1984). However, as this is still a heterogenous population, efforts have been made to sub-classify it further. This led to the identification of the stem cell marker CD90 antigen (Baum *et al.*, 1992) which, in combination with CD38 and CD45RA, that are expressed in differentiated progenitors (Sciences *et al.*, 1997), made it possible to call HSCs the CD34$^+$CD90$^+$CD38$^-$CD45RA$^-$ population. These HSCs give rise to intermediate progenitors, that progressively lose multilineage potential as more committed stages are reached. HSC differentiation eventually gives rise to common lymphoid progenitors (CLP) and common myeloid progenitors (CMP) (Kondo *et al.*, 1997; Akashi *et al.*, 2000). CLP differentiate into more specialized progenitors to give rise to T cells, B cells and natural killer cells. While, CMP differentiate to ultimately give rise to platelets, erythrocytes, granulocytes, macrophages, osteoclasts and dendritic cells (Fig. 1.1). Collectively, these mature blood cells constitute the adaptive and innate immune system.

The hierarchy tree concept has been challenged by recent studies which propose that lineage commitment occurs at the multipotent progenitor stage (Cabezas-Wallscheid *et al.*, 2014; Pietras *et al.*, 2015) and that the HSCs pool is more heterogenous in terms of functionality (Yamamoto *et al.*, 2013; Sanjuan-Pla *et al.*, 2013;

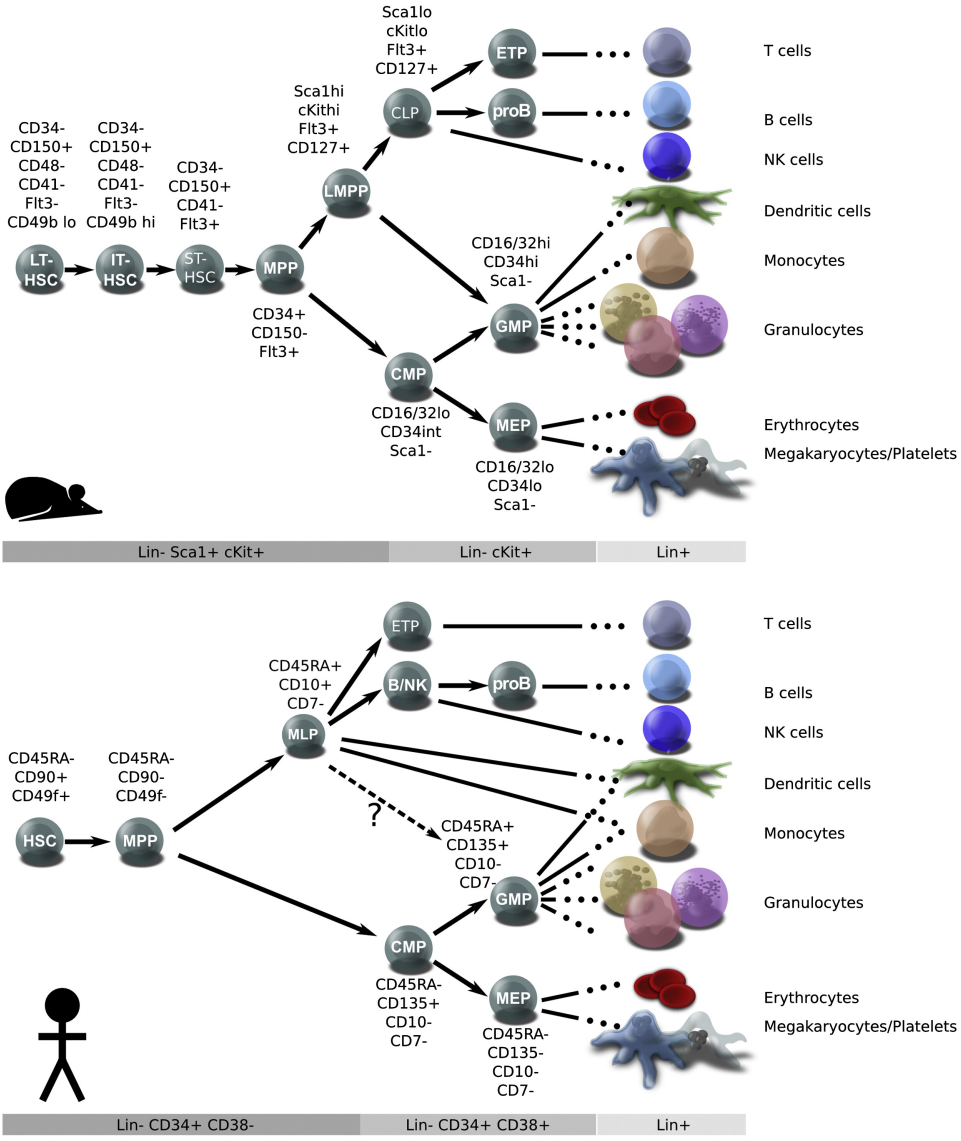

Figure. 1.1. Classical model of mouse and human haematopoiesis. HSC are at the apex and then progressively differentiate to multipotent progenitors (MPP) that give rise to common lymphoid progenitors (CLP) and common myeloid progenitors (CMP), which are committed cells with limited potency. These eventually differentiate into fully functional mature blood cells. Markers that identified the populations are indicated in grey and on each cell type. Reprinted with permission from Elsevier. Hematopoiesis: A human perspective, *Cell Stem Cell* Doulatov *et al.* (2012).

Pellin *et al.*, 2019). These investigations have led to new models being proposed where haematopoiesis is not a compartmentalized process, but instead, is flexible and occurs with gradual cell decisions that ultimately allow the cells to respond to changes in physiological needs (Laurenti and Göttgens, 2018). The hierarchy tree presented here will be used as a guide to discuss further concepts, since much of the current knowledge has been based on this model (Fig. 1.1).

1.2 Regulation of haematopoiesis

The haematopoietic system generates approximately one trillion cells every day in the adult human bone marrow, making blood the most regenerative tissue in mammals (Doulatov *et al.*, 2012). Therefore, a delicate balance between self-renewal and differentiation is crucial to prevent the development of haematopoietic disorders. Hence, the study of cell fate decisions has become an important topic in understanding how this balance is maintained.

HSC self-renewal is controlled by the stem cell microenvironment and several signalling pathways such as Wnt, Notch, integrin and cytokine receptors (Rieger and Schroeder, 2012; Orkin and Zon, 2008). In addition, there are genetic regulatory programs executed by transcriptional and epigenetic regulators, which are also involved in HSC differentiation (Orkin, 2000). Here, focus will be given to these genetic programs in definitive haematopoiesis.

1.2.1 Transcription factors

Transcription factors (TFs) regulate gene expression by binding to specific DNA sequences located in gene promoters, enhancer and silencer regions (Vaquerizas *et al.*, 2009). TFs, with their ability to reprogram the transcriptional landscape of a cell, play a key role in controlling the cell stage during cell development. Some TFs act broadly, controlling multiple genes in different cell types, such as MYC (Nie

et al., 2012). While other TFs are tissue specific and considered master regulators of a particular differentiation pathway. Studies involving the disruption of TFs in mice has helped to understand their relevance during development and how their deregulation can contribute to a disease stage.

In haematopoiesis, the Stem Cell Leukaemia (SCL), also known as T-cell Acute Lymphoblastic Leukaemia 1 (TAL1), TF is required for HSC formation that is necessary for primitive and definitive haematopoiesis. Deletion of *Scl* in the mouse embryo has shown that no development of blood occurs in its absence (Robb *et al.*, 1995). Similarly, in adult mice, its deletion blocks haematopoiesis because there are no HSCs (Porcher *et al.*, 1996). SCL is also involved at the quiescent and self-renewal stages of HSCs. High levels of SCL mRNA correlates with cells in G_0 and loss of *Scl* expression with cells in G_1 (Lacombe *et al.*, 2010). However, in HSCs from human cord-blood, SCL is correlated with self-renewal (Reynaud *et al.*, 2005). These investigations demonstrate that SCL has different functions in embryo (proliferative) and adult (quiescent) HSCs. Runt-related transcription factor 1 (RUNX1) is equally important for HSC formation. *Runx1* deletion blocks the definitive wave during embryonic development (Okuda *et al.*, 1996). During adult haematopoiesis, *Runx1* deletion results in lineage cell impairment with decreased platelet formation and a block of T cells and B cells maturation (Ichikawa *et al.*, 2004). Therefore, RUNX1 is not needed for HSC maintenance in adult haematopoiesis, but rather, for the development of lineage cells. Another TF involved in self-renewal is HOXB4, a member of the Hox homeo box genes, which increases self-renewal when it is over-expressed in HSCs *in vivo* and *in vitro* without losing any of its differentiation capacity (Sauvageau *et al.*, 1995; Thorsteinsdottir *et al.*, 1999; Antonchuk *et al.*, 2002). The TF pre-B cell leukemia transcription factor 1 (PBX1) impairs self-renewal when it is deleted using a conditional knock-out mouse (Ficara *et al.*, 2008). Similarly, deficiency of the early growth response 1 (EGR1) TF in mice, reduces HSC proliferation (Min *et al.*, 2008) and its overexpression increases differentiation

towards the macrophage fate (Krishnaraju *et al.*, 2001).

Other TFs work by restricting HSC proliferation. A knock-out mouse of the Growth factor independent 1 (*Gfi1*), exhibits more proliferative HSCs than wild-type mice. Moreover, *Gfi1*$^{-/-}$ HSCs are able to repopulate the bone marrow of irradiated mice upon transplantation. However, these transplanted mice eventually develop pancytopenia because of a lack of mature blood cells (Hock *et al.*, 2004).

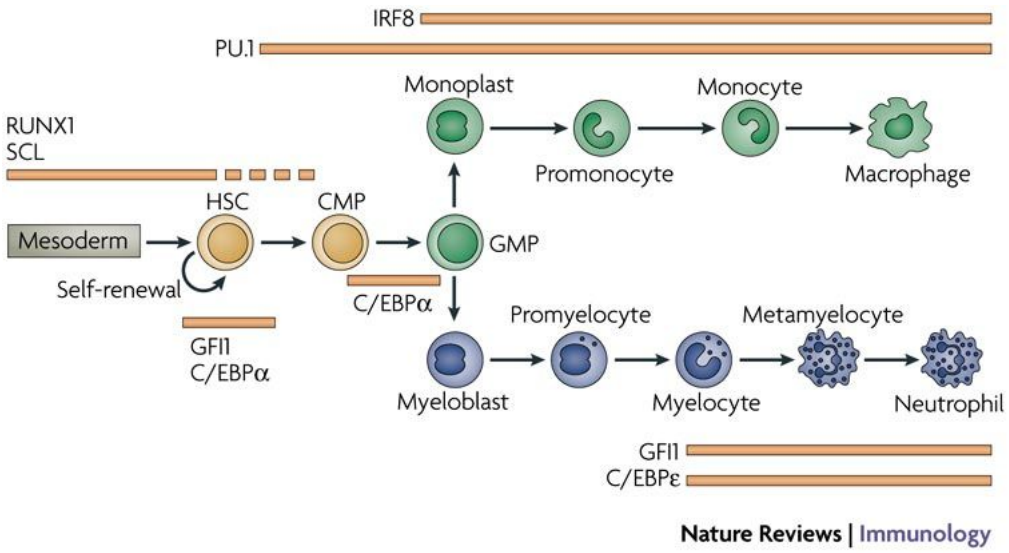

Figure. 1.2. Transcription factors of the myeloid differentiation process. The stem cell leukemia (SCL) and runt-related transcription factor 1 (RUNX1) TFs give rise to HSCs during embryonic development. The growth factor independent 1 (GFI1) is involved in HSC self-renewal and the CCAAT/enhancer binding protein-α (C/EBPα) TF is needed for myeloid commitment of the HSC to common myeloid progenitor (CMP) and granulocyte/macrophage progenitors (GMP). Macrophage differentiation is regulated by the PU.1 and interferon-γ (INF-γ)-responsive (IRF8) TF. Meanwhile, granulocyte differentiation is regulated by the GFI1 and CCAAT/enhancer binding protein-ε (C/EBPε) TFs. Reprinted with permission from Springer Nature. Transcription factors in myeloid development: balancing differentiation with transformation, *Nature Reviews Immunology,* Rosenbauer and Tenen (2007).

These studies demonstrate that self-renewal and differentiation needs to be balanced. In order for HSCs to progress to differentiation, self-renewal genes need to turn off. Investigations of TFs in cell lineage choice have mostly followed the classical model (Fig. 1.1), wherein a TF can either promote the lymphoid or myeloid lineages. However, the same TFs can act in both lineages at distinct differentiation stages and can also reprogram cells from one lineage to another (Orkin and Zon,

2008). The most relevant TFs will be discussed below and focus will be given to the myeloid lineage.

The myeloid transcriptional network starts with PU.1 (encoded by the oncogene *SPI1*), which is only expressed in blood cells. PU.1 is able to induce multipotent progenitor (MPP) cell commitment to the myeloid lineage prior to downregulation of GATA-1 (Nerlov and Graf, 1998), another necessary TF for erythroid maturation (Pevny *et al.*, 1991). PU.1 is not restricted to the myeloid lineage, as it is an important regulator during early T cell development (Rothenberg *et al.*, 2019). Its expression levels can also influence the development of progenitors towards B cells or macrophages (DeKoter and Singh, 2000). This is also the case in the myeloid lineage, where low or high levels of PU.1 lead to the production of either neutrophils or macrophages, respectively (Dahl *et al.*, 2003). In order to progress from CMP to the subsequent (GMP) and (MEP) stages, other TFs need to activate. The CCAAT/enhancer binding protein-α (C/EBPα) is important for the MPP to GMP transition. C/EBPα is expressed during myeloid commitment and in granulocytes but not in MEP (Radomska *et al.*, 1998; Akashi *et al.*, 2000). Deletion of C/EBPα impairs GMP production and granulocyte differentiation (Zhang *et al.*, 1997). Interestingly, if C/EBPα is deleted after GMP, it does not affect granulopoiesis. Therefore, C/EBPα must play a key role during the first stages of myeloid commitment (Zhang *et al.*, 2004). After this point, PU.1 and interferon-γ-responsive transcription factor (IRF8) are involved in the macrophage versus granulocyte differentiation (Laslo *et al.*, 2006; Kurotaki *et al.*, 2014). For further specification to neutrophils, both C/EBPε and GFI1 are necessary. C/EBPε is only expressed in the myeloid lineage (Williamson *et al.*, 1998) and its deletion in mice impairs granulopoiesis, resulting in a lack of mature neutrophils (Yamanaka *et al.*, 1997). Likewise, GFI1 deletion in mice has a similar phenotype. These GFI1 knock out mice also show accumulation of neutrophil precursors with abnormal expression of monocyte-specific genes, suggesting that GFI1 is needed to repress those genes (Hock *et al.*, 2003).

In MEP specification, the most relevant TF is GATA1. Deletion of GATA1 during embryonic development cause severe anaemia leading to death (Pevny *et al.*, 1991). Similarly, during erythroid development low expression of GATA1 blocks primitive erythropoiesis with substantial accumulation of undifferentiated erythroid cells (Takahashi *et al.*, 1997). During definitive haematopoiesis, GATA1 is necessary for erythroid and megakaryocytic development (Takahashi *et al.*, 1998). On the other hand, overexpression of GATA2 increases megakaryocytic differentiation (Ikonomi *et al.*, 2000), and it is downregulated during erythroid differentiation (Welch *et al.*, 2004).

These studies have led to a better understanding of the balance between self-renewal and differentiation which are crucial to not develop haematological disorders. This balance has to exist at every stage of differentiation, since every cell population have a distinct transcriptomic landscape. However, TFs are not the only actors in gene regulation as they are immersed within a complex network of transcriptomic regulation that includes epigenetic modifications.

1.2.2 DNA methylation in haematopoiesis and disease

Epigenetic modifications are heritable through mitosis and modulate gene expression without affecting the DNA sequence (Berger *et al.*, 2009). The two most studied epigenetic modifications are DNA methylation and histone modifications. DNA methylation is the addition of a methyl group to the 5'-position of a cytosine nucleotide in the context of the CpG dinucleotide. Histones are the proteins that package the DNA into nucleosomes. These are post-transcriptionally modified mainly at their tails, affecting the chromatin organization, and consequently transcription (Kouzarides, 2007). Because of its particular relevance to this thesis, the role of DNA methylation in haematopoiesis will be discussed further.

In mammals, DNA methylation mostly occurs at CpG dinucleotides and these are not evenly distributed in the genome. Regions rich in CpGs are called CpG

islands (CGIs). CGIs are defined as regions where the ratio between observed CpGs and expected CpGs is greater than 0.6 using a 200 bp window (Gardiner-Garden and Frommer, 1987). CGIs are usually located at gene promoters and these are mostly not methylated. CGIs are also present at intergenic and intragenic locations and these are typically tissue-specifically and developmentally methylated (Illingworth *et al.*, 2008; Deaton *et al.*, 2011).

DNA methylation is crucial in biological processes including gene silencing, genomic imprinting and X chromosome inactivation (Reik, 2007). DNA methylation is carried out by DNA methyltransferases (DNMTs), a group of proteins which include DNMT1, DNMT3A and DNMT3B. DNMT1 is involved in the maintenance of DNA methylation during replication whereas DNMT3A and DNMT3B are involved in *de novo* methylation (Okano *et al.*, 1998; Lei *et al.*, 1996). The DNMT family establish and maintain methylation patterns across the genome, which is necessary for development and cell differentiation (Huang and Fan, 2010). DNA methylation controls few developmental genes and tissue-specific genes, producing distinct patterns of DNA methylation along the differentiation of some cell lineages (Meissner *et al.*, 2008; Bock *et al.*, 2012). Therefore, DNA methylation is also important for haematopoiesis. During haematopoiesis, a gain in expression and loss of DNA methylation on some genes that are cell type specific has been observed (i.e *Scl* gene) (Attema *et al.*, 2007; Bock *et al.*, 2012). Moreover, comparative analyses between myeloid and lymphoid lineages reveals that DNA methylation in lymphoid development is gained at regions that are important for myeloid development (Ji *et al.*, 2010; Bock *et al.*, 2012; Farlik *et al.*, 2016). In this context, analysis of differentially methylated regions between lineages (myeloid and lymphoid) in human and mouse reveals enrichment of TF binding sites that are known to be lineage specific, such as GATA1, GATA2 and RUNX1 (Bock *et al.*, 2012; Farlik *et al.*, 2016). These findings demonstrate that DNA methylation may help to protect lineage-committed cells from accidental activation of TFs that could occurs during oncogenesis (Bock

et al., 2012; Langstein *et al.*, 2018).

Mutations in *DNMTs* are common in hematological diseases and their role on the fate of HSCs has been demonstrated. HSCs from *Dnmt3a* conditional knock-out mice exhibit increased self-renewal and decreased expression of TFs that are crucial for cell differentiation (Challen *et al.*, 2011). These effects are enhanced in *Dnmt3a* and *Dnmt3b* null mice demonstrating a synergistic effect of the two enzymes (Challen *et al.*, 2014). In acute myeloid leukaemia (AML) patients, the first mutation described in epigenetic modifiers was in the DNMT3A enzyme (Ley *et al.*, 2010). This finding prompted the investigation of the *DNMT3A* mutation in other hematological diseases, including myeloproliferative neoplasms (MPNs), where it was identified in 10% of the patients (Table 1.2)(Roller *et al.*, 2013). Other mutations involved in epigenetic regulators found in hematological diseases are in TET2, IDH1 and IDH2 enzymes (Table 1.2). The TET enzymes are a family of proteins that includes TET1, TET2 and TET3, and their function includes the ability to remove DNA methylation by oxidation of the 5-methylcytosine (5-mC) to 5-hydroxymethylcytosine (5-hmC), which can be further oxidized (Tahiliani *et al.*, 2009). The 5-hmC is not recognized properly by the DNA maintenance machinery and DNA methylation is lost through cell division (Valinluck and Sowers, 2007). TET2 mutations mostly affect the functional domains compromising the enzymatic activity to oxide 5-mC to 5-hmC (Ko *et al.*, 2010). Mutations in TET2 have been identified in MPN patients and other myeloid malignancies, where bone marrow samples show low levels of 5-hmC and hypermethylation in specific regions compared to healthy controls or patients with no mutations in TET2 (Ko *et al.*, 2010). Moreover, mice with conditional *Tet2* knock-out in the haematopoietic compartment exhibit increased self-renewal and proliferation of HSCs, with enhanced myeloproliferation and myeloid differentiation (Moran-Crusio *et al.*, 2011). Analysis of TET2 depleted mouse embryonic stem cells revealed that enhancer regions are mostly hypermethylated compared to wild-type cells which are enriched in 5-hmC. These findings suggest that TET2 prevents

aberrant hypermethylation at enhancer regions that could contribute to changes in gene expression (Hon *et al.*, 2014). IDH1/IDH2 enzymes convert isocitrate to α-ketoglutarate producing NADPH and CO_2 in the cytoplasm/peroxisome and mitochondrial matrix, respectively. Mutations in these enzymes have been described in AML patients but rarely in MPN patients (Mardis *et al.*, 2009; Gross *et al.*, 2010; Pardanani *et al.*, 2010; Brecqueville *et al.*, 2012). Mutations in IDH1/2 switches its catalytic activity to the conversion of α-ketoglutarate to 2-hydroxyglutarate (2-HG) by NADPH consumption. The 2-HG molecule is highly similar to α-ketoglutarate. Therefore, it works as a competitive inhibitor of α-ketoglutarate-dependent dioxygenases, including the TET enzymes (Gross *et al.*, 2010; Xu *et al.*, 2011). Conditional knock-in mice for *Idh1* in haematopoietic cells develop splenomegaly, anaemia and enhanced proliferation of HSCs. Moreover, hypermethylation is observed at promoters and intragenic regions, which correlates with the DNA methylation pattern of AML patients with IDH1/2 mutations (Akalin *et al.*, 2012; Sasaki *et al.*, 2012).

These somatic mutations can also be found in cells from individuals with no haematopoietic malignancies. Most of these mutations have been identified in around 10 to 15% of people aged 70 and older. The most common mutations are in the *DNMT3A* and *TET2* genes. Mutations in these genes confer a high risk for developing haematopoietic malignancies, for which this condition has been termed clonal haematopoiesis of indeterminate potential (CHIP) (Steensma *et al.*, 2015; Gibson *et al.*, 2017). Although 0.5 to 1% of individuals with CHIP develop haematopoietic neoplasms, studying the cause of these mutations may help to prevent the disease (Heuser *et al.*, 2016).

These studies have demonstrated the importance of conducting whole genome sequencing in patients to detect point mutations that could explain the causes of haematological diseases. More importantly, these investigations have revealed that DNA methylation seems to be an important factor to develop haematological disorders. Therefore, it is also useful to test whether these mutations contribute to

changes in the epigenomic landscape that can be translated to deregulation of gene expression. Accordingly, different methods have been developed to assess DNA methylation. These methods can be locus specific such as direct bisulfite sequencing in a small region, which is helpful to investigate specific questions related to one gene or region (Parrish *et al.*, 2012). However, assaying DNA methylation in whole genome provides valueble information to characterize and to model the effect of specific factors, such as mutations. Currently, bisulfite sequencing is considered to be the "gold standard" method to assay DNA methylation. However, this method is relatively expensive due to the high sequencing depth required. Other cost-effective alternatives do exist though, such as the Illumina MethylationEPIC BeadChip, a microarray that is able to quantify DNA methylation in specific genomic regions in the whole genome (Pidsley *et al.*, 2016). Assaying DNA methylation in haematological diseases can provide another layer of information to understand the deregulated gene expression usually observed in cancer. DNA methylation assays can also constitute a source of information to understand how a drug could be regulating gene expression in disease and ultimately link this information to its therapeutic effect.

1.3 Myeloproliferative neoplasms

Myeloproliferative neoplasms (MPNs) are a heterogenous group of haematological disorders characterized by uncontrolled proliferation of haematopoietic stem and progenitor cells in the bone marrow and peripheral blood. The term MPN collectively refers to chronic myeloid leukaemia (CML), polycythaemia vera (PV), essential thrombocythaemia (ET) and primary myeloid fibrosis (PMF) conditions (Arber *et al.*, 2016). CML is characterized by the presence of the fusion gene BCR-ABL1, which is a constitutively active form of the ABL1 tyrosine kinase involved in cell proliferation and differentiation (Nowell and Hungerford, 1985). The so-called 'Philadelphia Chromosome negative' MPN which include PV, ET and PMF lack this

mutation. MPN patients have a heterogeneous phenotype and may develop disease related symptoms, such as splenomegaly and have an increased risk of thrombotic events and hemorrhage. The criteria to diagnose each of these disorders have been dictated by the World Health Organization (Table 1.1 diagnosis criteria for PV and ET) (Arber *et al.*, 2016). PV is characterized by erythrocytosis, as defined by a sustained increase in haematocrit and frequently may present with concomitant leucocytosis or thrombocytosis. Diagnostic confirmation of ET may require a bone marrow morphology assessment to distinguish it from PV with co-existent thrombocytosis or pre-fibrotic MF (Tefferi *et al.*, 2018b). PMF is a more severe form of MPN where patients progressively develop fibrous "scar-like" tissue deposition within the bone marrow that can lead to progressive bone marrow failure, anaemia and extramedullary haematopoiesis. PV and ET patients may also develop MF – so called post-PV and post-ET MF – and later progress to acute myeloid leukaemia (AML), so called post-MPN AML (Iurlo *et al.*, 2019).

The incidence of ET, PV and PMF are 1, 0.7 and 0.5 cases per 100,000 per year in the European Union, respectively (Frederiksen *et al.*, 2016). The main three driver mutations contributing to these disorders can be found in the Janus Kinase 2 (*JAK2*), Myeloproliferative Leukemia Virus (*MPL*) and Calreticulin (*CALR*) genes. These mutations are normally mutually exclusive and ultimately result in constitutive activation of the JAK-STAT pathway which is involved in cell proliferation, cell differentiation, cell migration and apoptosis. These alterations occur at the stem cell level, where normal stem cells cohabitate with mutant lineage cells (Jamieson *et al.*, 2006). Some patients do not present with any detectable mutations in these genes, and are thus classified as triple negative (TN). More rarely, other mutations can affect the *ASXL1, EZH2, TET2, IDH1/2* genes, which are involved in epigenetic regulation (Table 1.2) (Brecqueville *et al.*, 2012; Stegelmann *et al.*, 2011; Tefferi *et al.*, 2009).

Characterization of these somatic mutations has helped considerably in under-

Table. 1.1. World Health Organization diagnosis criteria for PV and ET

Diagnosis	Major criteria	Minor criteria
PV (All 3 major criteria, or the first 2 major criteria and the minor criterion)	1. Hb >16.5 g/dL in men and >16.0 in women or Hematocrit >49% in men and >48% in women or increased red cell mass (>25% mean predicted 2. Bone marrow biopsy • hypercellularity for age • trilineage hyperproliferation (panmyelosis) • megakaryocytic proliferation with pleomorphic, mature megakaryocytes 3. Presence of JAK2V617F or JAK2 exon 12 mutation	Subnormal serum erythropoietin level
ET (All 4 major criteria or the first 3 major criteria and the minor criterion)	1. Platelet count $\geq$ 450 x 10^9/L 2. Bone marrow biopsy • Megakaryocyte proliferation with hyperlobulated nuclei. • No significant increase or left shift in granulopoiesis or erythropoiesis • very rarely minor (grade 1) increase in reticulin fibers 3. Not meeting WHO criteria for BCR-ABL1 CML, PV, PMF, myelodysplastic syndromes, or other myeloid neoplasms 4. Presence of JAK2, CALR, or MPL mutation	Presence of a clonal marker or absence of evidence for reactive thrombocytosis

Table. 1.2. Genes involved in epigenetic mechanisms, type of mutation, function, frequency and consequences in MPN patients.

Gene	Type of mutation	Protein function	Frequency	Disease consequence	Reference
		DNA methylation			
TET2	Missense, nonsense deletion	Remove DNA methylation by oxidation of 5mC into 5hmC **Mutation:** loss of function	10%-20% MPN (ET, PV, and PMF)	Initiation, Mutations on 2 alleles associated with progression	Delhommeau *et al.* (2009)
DNMT3A	Missense, hotspot	DNA de novo methylation during replication **Mutation:** loss of function	5%-10% MPN (ET, PV, and PMF)	Initiation	Stegelmann *et al.* (2011); Roller *et al.* (2013)
IDH1	Missense, hotspot	Generation of α-ketoglutarate **Mutation:** generation of 2-hydroxyglutarate blocking α-ketoglutarate enzymes	1%-3% PMF	Initiation, Disease progression	Tefferi *et al.* (2010)
IDH2	Missense, hotspot	Generation of α-ketoglutarate **Mutation:** generation of 2-hydroxyglutarate blocking α-ketoglutarate enzymes	1%-3% PMF	Initiation, Disease progression	Tefferi *et al.* (2010)
		Histone modification			
ASXL1	Nonsense/ indel	Chromatin-binding protein associated with PRC1 and PRC2 **Mutation:** loss of function	25% PMF 1%-3% ET/PV	Initiation, Rapid progression	Carbuccia *et al.* (2009)
EZH2	Missense, indel	H3K27 methyltransferase. **Mutation:** Loss of function.	5%-10% PMF	Initiation, Disease progression	Kamminga (2006); Ernst *et al.* (2010); Puda *et al.* (2012)

MPN: myeloproliferative neoplasms; PV: polycythaemia vera; ET: essential thrombocythaemia; PMF: primary myelofibrosis

standing the basis of the disease and account for common phenotypes. MPN patients have elevated circulating levels of pro-inflammatory cytokines and deregulation of anti-inflammatory cytokines have also been observed (Pourcelot *et al.*, 2014; Vaidya *et al.*, 2012; Panteli *et al.*, 2005). The JAK-STAT signalling pathway is involved in the release of these molecules and thus inflammation is thought to play an important role in the development of MPN (Kleppe *et al.*, 2015).

1.3.1 The JAK-STAT pathways and mutations in the *JAK2* gene

The JAK family includes the JAK1, JAK2, JAK3 and TYK2 proteins. These are protein tyrosine kinases that associate with cytokine receptor subunits such as thrombopoietin receptor (MPL), erythropoietin receptor (EPOR) and granulocyte colony-stimulating factor receptor (G-CSF-R). These receptors are activated upon binding of cytokines that induce conformational changes resulting in activation of the JAK proteins by trans-phosphorylation (Fig. 1.3). These activated JAKs are able to phosphorylate several targets including a tyrosine residue of the STAT transcription factor family that resides in the cytoplasm. Phosphorylation of STATs allows dimerization through interaction of their phosphotyrosine and Src homology 2 (SH2) domain. Once phosphorylated, STATs are able to enter the nucleus and bind to specific regulatory sequences of target genes (Rawlings *et al.*, 2004). Besides STAT activation, the JAK signalling pathway can also activate the mitogen-activated protein kinase (MAPK) and phosphatidylinositol-3'-kinase (PI3K) pathways.

The myeloid JAK-STAT signalling pathway is regulated by the homodimeric receptors which employ mostly JAK2. EPOR mediates the erythroid proliferation and differentiation mainly through activation of STAT5 (Klingmüller *et al.*, 1997; Grebien *et al.*, 2008). MPL regulates megakaryocyte and platelet production by activation of STAT3 and STAT5 (Bacon *et al.*, 1995; Miyakawa *et al.*, 1996). G-CSF-R modulates granulopoiesis by strong activation of STAT3 (Dwivedi and Greis, 2017).

Lymphoid proliferation is also regulated by the JAK-STAT signalling pathway. In this case, the cytokine receptors such as IL-2, IL-4, IL-7, IL-9, IL-15 and IL-21 activate the STAT proteins by phosphorylation of JAK1 and JAK3. This signalling regulates the formation and function of T cells, natural killer cells and B cells (Liao *et al.*, 2011).

JAK-STAT signalling is transient and three major classes of negative regulators have been described: PTPs (protein tyrosine phosphatases), SOCS (suppressors of cytokine signalling), and PIAS (protein inhibitors of activated STAT) (Fig. 1.3) (Yasukawa *et al.*, 2000). The most characterized PTP is SHP-1, which is mostly expressed in haematopoietic cells. This protein recognizes phosphorylated tyrosine residues from receptors such as EPOR or JAK2 proteins and dephosphorylates them (Fig. 1.3) (Klingmüller *et al.*, 1995). The SOCS proteins work as a negative feedback loop after JAK-STAT activation which stimulates SOCS transcription. SOCS can bind to either the phosphotyrosine residues of cytokine receptors (SOCS-2, SOCS-3 and CIS) or to JAKs (SOCS-1) inhibiting their kinase activity (Fig. 1.3) (Endo *et al.*, 1997; Yoshimura *et al.*, 1995; Schmitz *et al.*, 2000; Greenhalgh *et al.*, 2002). Another mechanism of inhibition involves the ubiquitination of JAKs which are targeted for proteosomal degradation (Kile *et al.*, 2002). The PIAS proteins interact with STAT proteins and block their transcriptional activity by preventing them from binding to DNA or recruiting co-repressor molecules (Fig. 1.3) (Liu *et al.*, 1998; Chung *et al.*, 1997; Arora *et al.*, 2003).

In 2005, several published studies have described a mutation in the *JAK2* gene which is present in 90- 95% of PV patients and approximately 50% of ET and MF patients (Kralovics *et al.*, 2005; Scott *et al.*, 2005; Ugo *et al.*, 2005). This is a point mutation, a G to T transversion in exon 14 of the *JAK2* gene, that causes a substitution of valine to phenylalanine at amino acid 617 (V617F) in the JAK2 molecule (Kralovics *et al.*, 2005). This mutation activates the JAK2 protein independently of cytokine stimulation – so called constitutive activation – which

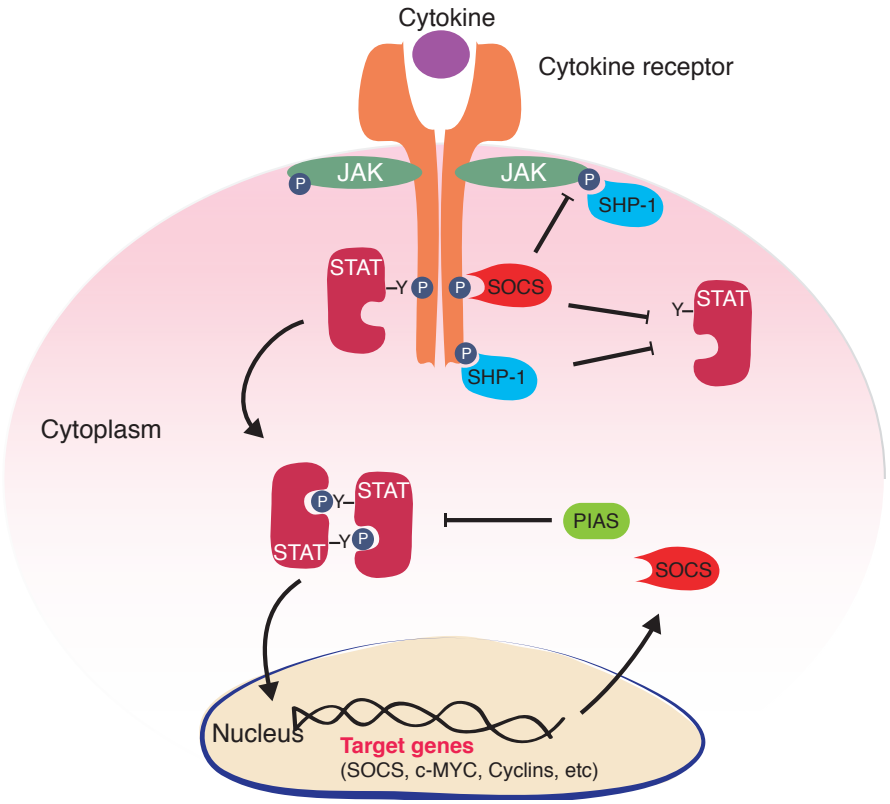

Figure. 1.3. JAK-STAT pathway. JAK proteins bind to cytokine receptors that are activated upon cytokine ligation. The receptor once activated, produces conformational changes that brings two JAK proteins close together so that they can trans-phosphorylate each other. These active phosphorylated JAKs are then able to phosphorylate other molecules such as STAT transcription factors. STAT proteins are phosphorylated at one tyrosine residue allowing them to homodimerize with another phosphorylated STAT. This allows them to enter to the nucleus and bind to their target genes. The JAK-STAT pathways can also regulate themselves by allowing the transcription of SOCS proteins through STAT. SOCS proteins act as negative regulators, which bind to JAKs or receptors inhibiting their kinase activity. Another two negative regulators are SHP-1 and PIAS. SHP-1 dephosphorylates JAK, receptors and PIAS prevents STAT from binding to the DNA. (Y: tyrosine residue, P: phosphorylated tyrosine)

is associated with homodimeric cytokine receptors, leading to the expansion of all myeloid lineages (Lu *et al.*, 2008).

Another four somatic mutations in the *JAK2* gene have also been described. These are gain-of-function mutations located in the exon 12. These mutations are less common, presenting in only approximately 10% of PV patients (Scott *et al.*, 2010).

Interestingly, the ratio between STAT3 and STAT5 phosphorylation has been shown to differ between MPN categories. PV patients have increased STAT3 and STAT5 phosphorylation whereas ET patients have increased STAT3 and reduced STAT5 phosphorylation. PMF patients have reduced phosphorylation of both STAT3 and STAT5. These findings are independent of the mutational state of JAK2, suggesting that other factors contribute to the state of disease (Teofili *et al.*, 2007).

1.3.2 Clinical management of PV and ET

In the clinical management of PV and ET, patients need to be stratified according to their risk of thrombotic events. A high-risk patient typically has a history of thrombosis, or is over 60 years of age, and harbours mutations in the *JAK2* gene. By contrast, a low-risk patient is typically below 60 years, has not had any thrombotic events, and has no mutations in the *JAK2* gene (McMullin *et al.*, 2019).

The goal of the treatment is to relieve the symptoms and prevent cardiovascular events, such as thrombosis and bleeding, and disease-related symptoms where present.

Low risk ET patients are generally treated with aspirin and a healthy lifestyle is encouraged. In addition to aspirin, high risk patients may also require a cytoreductive drug treatment such as hydroxycarbamide (hydroxyurea,HU), interferon-α, anagrelide and rarely busulfan. These treatments aim to lower the levels of platelets (frequently $<400 \times 10^9$/L) (Harrison *et al.*, 2010). HU is typically the first line of treatment, and has been demonstrated to reduce the occurrence of thrombotic

events significantly (from 24 to 3.6%) (Harrison *et al.*, 2010). However, HU may have several side effects such as drug related neutropenia, mouth ulcers, fever and skin lesions (Randi *et al.*, 2005). In younger patients (<60 years of age), interferon-α is frequently recommended particularly for patients planning to get pregnant (Harrison *et al.*, 2010).

In the case of PV patients, the risk of thrombosis is reduced by controlling erythrocytosis, leukocytosis and thrombocytosis with cytoreductive therapy. Aspirin also helps to reduce thrombotic events (Landolfi *et al.*, 2004). The aim of the treatment is to maintain hematocrit levels under 45%, for which phlebotomy is recommended (McMullin *et al.*, 2019). For the reduction of leukocytosis and thrombocytosis, HU or interferon-α are the first line treatments. Some patients develop resistance to HU requiring them to switch therapy. Resistance is associated with a more aggressive stage of the disease and these patients usually present with a poorer prognosis (Tefferi *et al.*, 2018a).

HU therapy has been the first-line treatment for MPN patients for many years. However, limited studies have been conducted to understand better its mechanism of action in MPN patients. Below, information on what is known about HU in the context of the disease is reviewed.

1.4 Hydroxyurea

1.4.1 Mechanism of action

The primary target of HU is the inhibition of the ribonucleotide reductase (RNR) enzyme which catalyses the reduction of ribonucleotide diphosphates to their corresponding deoxyribunocleotide triphosphates (dNTPs), used in DNA synthesis and repair (Young and Hodas, 1964; Krakoff *et al.*, 1968). HU inhibits the RNR enzyme and depletes dNTPs during the synthesis (S) phase of the cell cycle. In consequence, the DNA polymerase movement is slowed down at the replication forks during repli-

cation (Sogo *et al.*, 2002). This process activates the S phase checkpoint, which is a highly conserved intracellular signalling pathway that maintains genome stability and is activated in response to replication stress. The S checkpoint normally prevents the cell from moving to the next stage (G_2) to allow recovery from the stalled replication forks (Boddy and Russell, 2001). Cell lines treated with HU have demonstrated that the S arrest induced by HU is reversible since DNA synthesis can be restarted when the drug is removed (Petermann *et al.*, 2010). However, when cells are treated with high concentrations or with prolonged exposure, HU can be cytotoxic as demonstrated experimentally in primary cells and in cell lines (Johnson *et al.*, 1992; Warren K. Sinclair, 1967). The cytotoxic effect is produced by fork collapse, where DNA synthesis can no longer be resumed at the site of stalling, inducing double strand breaks (DSB) that are accumulated with longer exposure to the drug (Saintigny *et al.*, 2001). During S phase, DSB are repaired by homologous recombination using the sister chromatid as a template. Failure to repair can result in chromosome loss, chromosomal rearrangements or apoptosis.

1.4.2 Clinical outcomes

HU was first synthesized in 1869 in Germany and its effects on lymphocyte formation in animals were first reported in 1928 (Rosenthal *et al.*, 1928). But it was not until the early 60's when HUs potential use in leukemia was demonstrated with the use of mouse models and later with clinical trials (Stearns *et al.*, 1963). HU was first used in CML patients where it caused myelosuppression, decreased spleen size and increased haemoglobin levels (Thurman *et al.*, 1963; Fishbein *et al.*, 1964). The need to treat high risk MPN patients with more than just aspirin and phlebotomy, led clinicians to start using HU in MPN patients in 1981 (Löfvenberg and Wahlin, 1988). The outcomes were beneficial for MPN patients, with resultant cytoreduction, potential relief of symptoms and a low side effect burden. It was also demonstrated that HU could reduce the risk of thrombotic events in ET patients

(Cortelazzo *et al.*, 1995), hence HU has been the first-line drug for high risk PV and ET patients over the age of 60 years (Mcmullin *et al.*, 2015). However, resistance to the drug can occur, whereby the reduction of blood counts is not achieved within three months of dose optimised therapy. Some patients may also present with HU intolerance due to adverse side effects including neutropenia, anaemia and skin ulcers. In these patients, a switch to an alternative myelosuppressive drug may be required, such as anagrelide, interferon-α or busulfan, which are normally used in younger patients (Alvarez-Larrán *et al.*, 2014). HU has also been suggested to be associated with a low incidence of leukemic transformation (3-4%), however this is controversial, albeit this is a risk that increases when used in combinations with other older myelosuppressive agents (Sterkers *et al.*, 1998; Harrison *et al.*, 2005). Deletion of chromosome arm 17p has been described among individuals treated with HU that have progressed to AML or myelodysplastic syndrome (MDS) (Sterkers *et al.*, 1998; Merlat *et al.*, 1999; Bernasconi *et al.*, 2002). Other chromosome abnormalities might vary depending on the type of therapy used before progressing to AML or MDS (Swolin *et al.*, 2008). HU has also been associated with the induction of macrocytosis. Macrocytosis is characterized by an increase in the mean corpuscular volume (MCV) levels, due to large erythrocytes that are over-hydrated and contain more quantities of haemoglobin than normal (Burns *et al.*, 1986; Spier, 1971). However, according to what has been observed in patients treated with the drug, this does not seem to affect the microcirculation since the geometry and deformability of erythrocytes are not affected (Engström and Löfvenberg, 1998). Another effect of HU is a reduced expression of endothelial adhesion molecules which are essential for the inflammatory process, specifically in leading leucocytes, such as neutrophils, to the site of inflammation (Gambero *et al.*, 2007). HU has also been associated with nitric oxide (NO) production by two mechanisms: by acting as a NO donor (a product of its metabolism) (Gladwin and Schechter, 2001), and by stimulating the endothelial nitric oxide synthase (eNOS) (Cokic *et al.*, 2006). NO, that is also

secreted by phagocytes and endothelial cells, acts as a signalling molecule causing vasodilatation, increased local blood flow and reduced platelet activation. NO activates the soluble guanylate cyclase, which catalyses the formation of cyclic guanosine monophosphate (cGMP) involved in the vascular regulation tone and platelet function (Arnold *et al.*, 1977; Brüne and Ullrich, 1987). The NO/cGMP pathway has also been proposed as a mechanism of foetal haemoglobin (HbF) induction by HU that has been observed in erythroid progenitors cells (Fig. 1.4) (Platt *et al.*, 1984; Cokic *et al.*, 2006; Lou *et al.*, 2009). Due to these effects, HU is also used to treat sickle cell anaemia (SCA), which is a blood disorder caused by a point mutation in the β-globin gene producing a sickle haemoglobin that decreases erythrocyte deformability leading to vascular obstruction and ischemia (Platt, 2008). HbF induction may relieves patients' symptoms and several studies have been conducted to understand further the mechanism of action of HU. Some of these studies observed hypomethylation at the γ-globin promoter, which encodes HbF, following HU treatment in SCA patients (Walker *et al.*, 2011). Other drugs that induce HbF include 5-azacytidine and 5-aza-2'-deoxycytidine (decitabine) which inhibit DNMT1 (Saunthararajah *et al.*, 2003; Ley *et al.*, 1983; Christman, 2002).

This evidence indicates that HU has multiple effects that are not always attributed to its main mechanism of action. Efforts have been made to understand HU clinical effects, however most of the studies have been focussed in SCA rather than in MPN. One study attempted to assess gene expression in MPN patients comparing samples from before HU treatment and after one week of treatment, however this was a very short period of time where some of the patients did not show signs of therapeutic effects (Bruchova *et al.*, 2002). Therefore, there is motivation to understand how HU works in MPN patients. Since effects on DNA methylation have been attributed to HU clinical outcomes, it is an interesting drug to investigate in the context of a myeloproliferative disease, where abnormal gene expression and DNA methylation patterns can be expected.

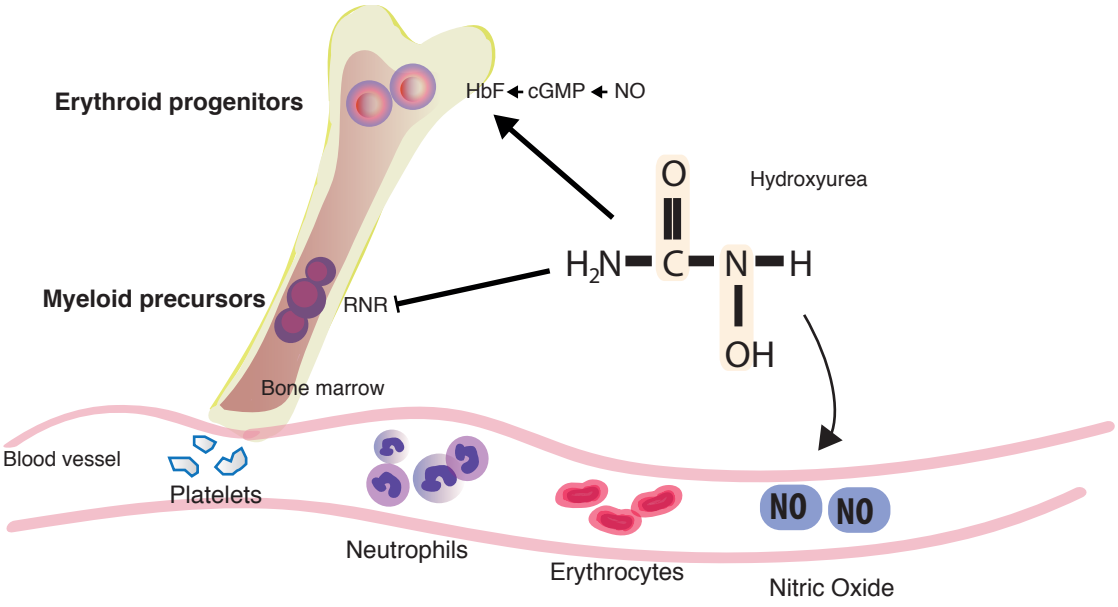

Figure. 1.4. The effects of hydroxyurea. Hydroxyurea (HU) mechanism of action is the inhibition of the ribonucleotide reductade (RNR) that is necessary for DNA synthesis. The result of the RNR inhibition is cell cycle arrest in synthesis phase, which can lead to apoptosis. This is the main mechanism of action to explain the reduction of blood cells seen in patients treated with the drug. Another effect is the release of the vasodilator nitric oxide (NO) caused from the metabolism of HU. NO also induces cyclic guanosine monophosphate (cGMP) synthesis which, in turn,induces the foetal haemoglobin (HbF) in erythroid progenitor cells.

1.5 Mouse models of the disease

After the description of the JAK2V617F mutation in MPN patients, several mouse models were generated with the aim of understanding the exact role this single point mutation holds in the pathogenesis of MPN. The first models were generated by transplantation of bone marrow cells that were retrovirically transduced with JAK2V617F which resulted in the development of PV and sometimes, full progression to a fulminant myelofibrosis-like disease (Bumm *et al.*, 2006; Lacout *et al.*, 2008; Wernig, 2006; Zaleskas *et al.*, 2006). To generate an inducible/tissue specific model the Cre/loxP system was used. Transgenic mice were developed to express JAK2V617F when crossed with mice expressing Cre recombinase (Tiedt *et al.*, 2008). The level of JAK2V617F expression was determined by using either *Vav*-Cre mice, which produce haematopoietic expression only (Georgiades *et al.*, 2002), or *Mx1*-Cre mice, where the expression in the haematopoietic compartment was transient and dose-dependent with respect to polyinosine-polycytosine (pIpC) injection (Kühn *et al.*, 1995). The *Vav*-Cre mice produced low JAK2V617F expression in the haematopoietic compartment and resembled an ET phenotype with elevated platelet counts and normal haematocrit. However, using *Mx1*-Cre mice, higher *JAK2*V617F expression was observed and a PV phenotype was obtained with increased erythrocytosis and thrombocytosis (Tiedt *et al.*, 2008). It was demonstrated that the level of JAK2V617F expression is important in determining the phenotypic subtypes of MPN. This was also observed in patient samples whereby homozygous mutations are usually present in PV patients rather than those with ET (Godfrey *et al.*, 2012).

The next generation of MPN models were used to in investigate the effects of JAK2V617F on a physiological level by expressing it from the endogenous promoter of *Jak2*. The first Jak2V617F knock-in model was designed to express mutated Jak2 after Cre-mediated recombination (Akada *et al.*, 2010). Homozygous and heterozygous Jak2V617F mutants were generated (Akada *et al.*, 2010). The

expression activity of Jak2V617F in haematopoietic cells was controlled by breeding the mice with Mx1-Cre mice, where Cre expression is conditionally induced by pIpC. Homozygous mice (Mx1-Cre;V617F/V617F) resembled a PV phenotype with more appreciable levels of reticulocytosis, leukocytosis, neutrophilia, thrombocytosis and splenomegaly than heterozygous mice (Mx1-Cre;V617F/+). These homozygous mice also had a higher platelets count, which is something expected from the ET phenotype. Moreover, immunoblotting revealed STAT5 constitutive phosphorylation in homozygous and heterozygous models (Akada *et al.*, 2010). Another two independent knock-in Jak2V617F mouse models were engineered to have a point mutation in the exon 13 of the mouse *Jak2* gene. One was constitutively active (Marty *et al.*, 2010) and with STAT5 constitutively phosphorylated. The other mouse model was conditionally activated by breeding the Jak2V617F floxed mouse with a E2ACre mouse inducing germ-line expression of Jak2V617F during mouse embryogenesis (Mullally *et al.*, 2010). Only heterozygous mice that resembled a PV phenotype were obtained from both research groups. These mice presented erythrocytosis, leukocytosis, enlarged spleen size and reduced survival. Another knock-in model expressed human JAK2V617F under the control of endogenous *Jak2* promoter. The expression was controlled by using Mx1-Cre mice and pIpC injections. After six weeks the mice developed human ET and after 26 weeks, developed PV-like disease (Li *et al.*, 2010).

These mouse models demonstrate that one point mutation is enough to produce the MPN phenotype and according to the levels of JAK2V617F expression different type of disease are generated. These models are a valuable source of information to investigate MPN as well as investigate drug efficacy.

Accordingly, for this study, the previously generated Jak2V617F floxed mice (Mullally *et al.*, 2010) were bred with *vav*Cre mice to induce Jak2V617F expression in the haematopoietic compartment (Chen *et al.*, 2015). These mice resemble the MPN phenotype with extramedullary haematopoiesis, reflected by splenomegaly

(Chen *et al.*, 2015). These mice were chosen as they have been routinely used in Dr Mullally's laboratory where she investigates the molecular mechanisms of MPN. I was fortunate enough to have the opportunity to visit her laboratory and learn their techniques first hand.

In summary, the first-line treatment for MPN patients, used for over 30 years, is HU. The mechanism of action of HU is the inhibition of the RNR enzyme, that is necessary for DNA synthesis during replication. However, HU also have several effects that cannot be explained by this mechanism, such as the induction of foetal haemoglobin. Some investigations suggest that HU is able to affect DNA methylation and control gene expression but this mechanism is poorly understood and has not been investigated in MPN before. Therefore, I have hypothesized that HU treatment influences the epigenome and changes gene expression which contributes to its therapeutic effect in patients. To test this hypothesis, I have assayed the effect of HU treatment on gene expression and DNA methylation in MPN patients. Since transcription is tightly regulated during haematopoiesis, two clinically relevant and differentially developed cell types where isolated: $CD34^{+}$ cells and neutrophils. Moreover, to strengthen the findings identified in patients, the effect of HU was also assayed in a well-characterized mouse model of the disease. By investigating the effect of HU in human and mouse, genes that are affected in both species can be proposed as candidates that explain HU mechanism and clinical effects.

Chapter 2

Materials and methods

2.1 Human sample collection and processing

This study was developed in collaboration with the Department of Haematology, Guy's and St Thomas' NHS Foundation Trust. Ethical approval was obtained from the NRES Committee London - City Road & Hampstead, REC reference 15/LO/0265.

Men and women over 18 years old diagnosed with MPN and prompted to start HU treatment were included in this study after providing informed consent according to the Helsinki Declaration. Peripheral whole blood was collected before HU treatment and then at three, six and nine months of treatment. Demographic information and cell blood counts (CBC) were obtained for each patient and time point (Appendix A.1).

2.1.1 Neutrophils isolation

Neutrophils were purified from 10 mL of peripheral blood as described previously (Heit *et al.*, 2002). Briefly, red blood cells were removed using dextran sedimentation (6% Dextran, 0.9% NaCl) and two rounds of hypotonic lysis with ddH_2O. Neutrophils were purified by density-gradient centrifugation with Histopaque-1077

(cat#10771, Sigma-Aldrich) at 700 x g for 20 mins. The mononuclear layer of cells was removed with a Pasteur pipette and the supernatant decanted. Neutrophils were resuspended in 5 mL of Phosphate-buffered saline (PBS) and kept on ice until use.

2.1.2 Neutrophils purity

An 80 µl aliquot of cells was incubated with 20 µl of Human Fc Receptor Binding Inhibitor (eBioscience, cat#14-9161-73) for 20 minutes at RT. Cells were immediately incubated with 5 µl of CD16b-PE (Clone CLB-gran11.5, BD Biosciences), CD14 (Clone M5E2, Biolegend), CD3 (Clone UCHT1, Biolegend) and CD19 (Clone HIB19, Biolegend) antibodies at RT for 30 minutes in the dark. The cells were washed to remove unbound antibody with 2 mL of PBS, centrifuged at 300 x g for 5 minutes and resuspended with 300 µL of PBS. The cells were analyzed on a BD FACSCanto^{TM} II Flow Cytometer (NIHR BRC Flow Cytometry Core Facility). The data were obtained as .fcs files which were analyzed using FlowJo (Version X) (Fig. 2.1).

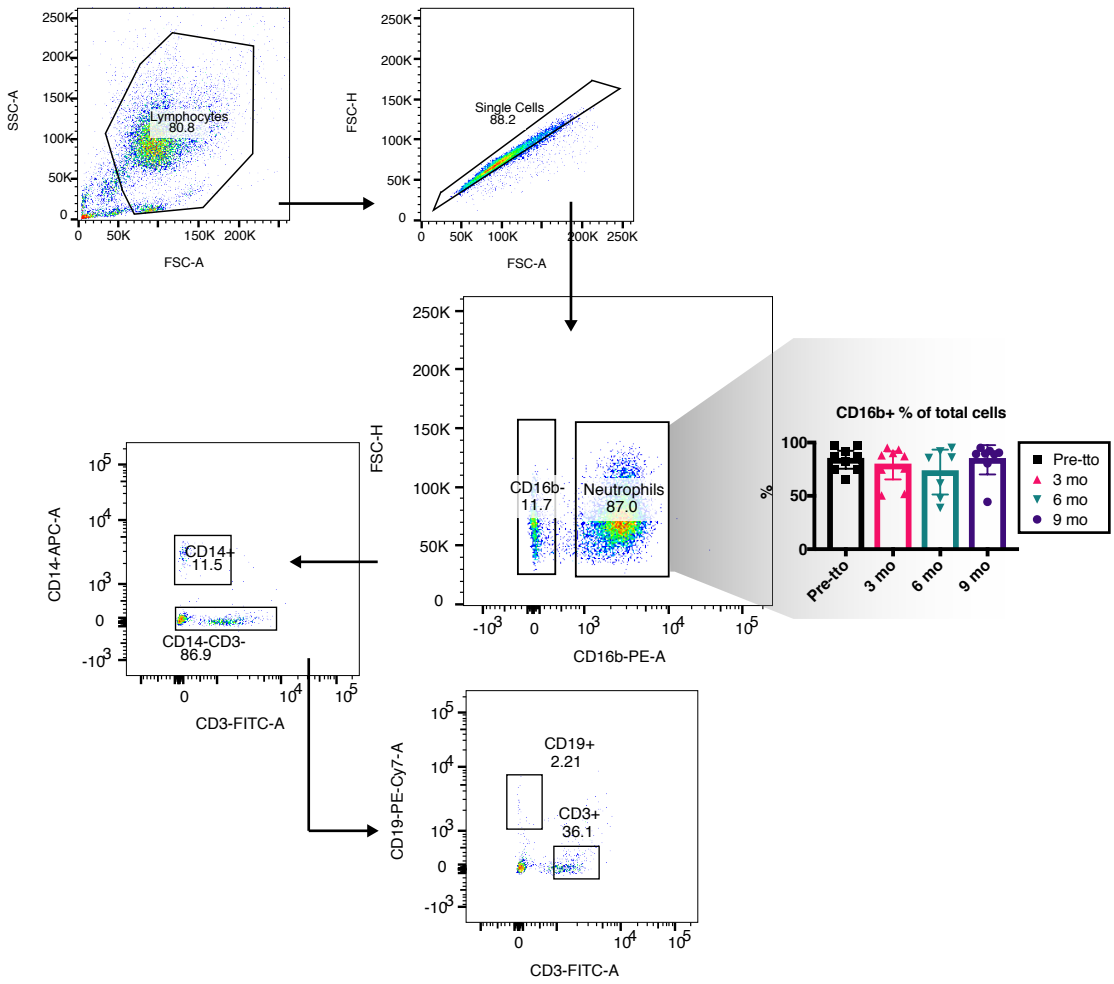

Figure. 2.1. Gate strategy for purity assessment of neutrophils ($CD16b^+$) . Flow cytometry was used to assess purity of neutrophils after granulocyte enrichment. For this, an aliquot of cells was stained with four antibodies: CD16b (neutrophils), CD14 (monocytes), CD19 (B cells), CD3 (T cells). The purity of CD16b purity was analyzed following the gating strategy presented here. Live and single cells were gated and then separated between $CD16b^+$ cells and $CD16b^-$cells. To further identify the type of cells in the $CD16b^-$ population, these were separated according to the expression of other markers: CD14 and CD3. The $CD3^+CD14^-$ fraction was further gated in CD3 and CD19.

2.1.3 $CD34^+$ cells isolation

Peripheral blood was collected in EDTA tubes and diluted with PBS at a 1:4 ratio (blood:PBS) and centrifuged for 10 minutes at 3000 rpm. The buffy layer was collected in a single tube and the same volume of PBS was added. Gradient centrifugation with Histopaque-1077 (cat#10771, Sigma-Aldrich) was used to separate the

peripheral blood mononuclear cells (PBMC) from red blood cells and granulocytes following the manufacturer's instructions. The PBMC were collected in a separate tube and washed with PBS by centrifugation at 250 x g for 10 minutes. Pelleted cells were resuspended at about $2x10^7$ cells/mL with RPM1-1640 media completed with 10% FBS, and 5% Penicillin/Streptomycin. Freezing medium, FBS+20%DMSO, was added to double the cell suspension and 1 ml was dispensed into cryovials, stored in Mr Frosty at -70^{o}C and then moved to liquid nitrogen.

Previously frozen PBMC aliquots, were rapidly thawed in a water-bath at 37^{o}C, resuspended with 1 mL of complete RPM1-1640 media and collected in 50 mL falcon tubes containing 3 mL of media. Cells were centrifuged at 300 x g for 5 minutes and the supernatant aspirated. The pellet was resuspended in 300 µL of PBS + 0.5% BSA + 2mM EDTA and passed through a 40 µm mesh to obtain a single cell suspension. $CD34^+$ cells were isolated using Miltenyi CD34 Ultrapure Microbeads (Miltenyi Biotec, Cat#130-100-453) following the manufacturer's instructions. To increase purity, double column separation was performed.

2.2 Jak2V617F knock-in and control mice

Heterozygous Jak*2*V617F knock-in (Jak2VF) mice were generated as previously described (Mullally *et al.*, 2010). Floxed Jak*2*V617F animals were crossed with *vav*-Cre transgenic mice to induce expression of Jak2V617F in haematopoietic lineages only (Georgiades *et al.*, 2002). The generated Jak2VF knock-in mice were backcrossed and maintained with a C57Bl/6 background. The genotype was confirmed by PCR as previously described (Chen *et al.*, 2015). Wild-type C57Bl/6 and *vav*-Cre mice were used as a genotype control and are referred to as wild-type (WT) indistinctly in the text. All mice were maintained in pathogen-free facilities at the Brigham and Women's Hospital, Boston, MA, USA. All mouse experiments were approved by the institutional ethic committee of Brigham and Women's Hospital,

Boston (Protocol#: 2017N000025).

2.2.1 HU mouse treatment

HU was purchased from Sigma-Aldrich (H8627-10G) and diluted with injectable 0.9% NaCl solution (Baxter, Cat# 0338-0051-44). HU solution was prepared freshly upon administration under sterile conditions. A 0.2 μm syringe filter (Acrodisc 25mm Syringe Filter, Life Science, Cat#PN4192) was used to sterilize the solution and dispensed into a 5 mL eppendorf tube.

Intraperitoneal injections were administered to Jak2VF and WT mice groups 5 days a week for 6 weeks. The first two weeks, the mice received 50 mg/kg from a 5 mg/mL solution of HU and the following four weeks 100 mg/kg from a 10 mg/mL solution of HU. Similarly, control groups per genotype received vehicle only. The treatment was performed in sterile conditions under a Class 2 A1 Biological Safety Cabinet.

2.2.2 Clinical parameters

CBC were measured weekly. Anesthesia was applied using a Table Top Research Anesthesia Machine w/O_2 Flush. Mice were placed in an induction chamber with O_2 flowmeter adjusted to 0.6 L/min and isoflurane vaporizer to 3.5%. Peripheral whole-blood was obtained from the retro-orbital sinus using Micro-Hematocrit Capillary Tube, Sodium Heparinized. Blood was collected in K_2EDTA coated tubes (BD Microtainer, Cat#365974) and 75 μL was diluted with 300 μL of PBS to measure CBC on Advia 2120i Hematology System (Brigham and Women's Hospital), results were multiplied by the dilution factor. Mice were weighed every week for dose adjustment and toxicity monitoring.

2.2.3 Sample collection

After six weeks of treatment, the mice were divided into groups of 4-5 mice to be sacrificed daily. Each group received a final dose of HU or vehicle the day before sacrifice. The weight of each mouse was registered prior to anesthesia. The mice were placed within an induction chamber with O_2 flowmeter adjusted to 0.3 L/min. Subsequently, isoflurane vaporizer (at 3.5%) was used to anesthetize them. The peripheral blood was collected by bleeding the mouse from the retro-orbital cavity. Cervical dislocation was undertaken on all mice to proceed with the dissection. Femur, tibia and spine were collected in PBS + 3% FBS and kept on ice until use.

2.2.3.1 Isolation of Lineage - Sca-1$^+$c-Kit$^+$(LSK) cells

Whole bone marrow was obtained by crushing the femur, tibia and spinal column in presence of 5 mL of PBS + 3% FBS. The cells were washed with PBS + 3% FBS, passed through a 100 µm filter and collected in 50 mL falcon tubes, this was repeated up to five times. The cells were pelleted by centrifugation at 300 x g for 5 minutes. To remove erythrocytes, the pellet was resuspended with 10 mL of 1x erythrocyte lysis buffer (10x BD-lysis-buffer, BD Bioscience, Cat#555899) diluted in ddH_2O and centrifuged again. To remove muscle material, the cells were resuspended in 10 ml of PBS and passed through a 70 µm cell strainer, then washed with 10 mL of PBS + 3%FBS. This step was repeated using a 40 µm cell strainer and centrifuged at 300 x g for 5 minutes. The cells were resuspended with 300 µL PBS followed by magnetic enrichment using 60 uL of Miltenyi CD117 microbeads (Miltenyi Biotec, Cat#130-091-224). Cells were left to incubate at 4ºC for 20 min. Remaining antibody was removed by washing with 10 ml of PBS + 3% FBS and centrifuged at 300 x g for 5 mins. The cells were resuspended with 4 mL AutoMACS running buffer (MACS Miltenyi Bioctec, Cat#130-091-221), passed through 40 µm cell strainer and washed with 3 mL of PBS + 3% FBS. The cells were then separated using autoMACS®

Pro Separator.

The positive cell fraction was collected and counted. Aliquots of $5x10^5$cells were used for fluorochromes compensation. The remaining cells were centrifuged and resuspended in 200 µL of PBS and incubated on ice for 20 minutes with 1 µL of each antibody: CD5 (Clone 53-7.3, Bioscience), Gr1 (Clone RB6-8C5, Biolegend), B220 (Clone RA3-6B2, Biolegend), CD3e (Clone 17A2, Biolegend), CD11b (Clone M1/70, Biolegend), Ter119 (Clone Ter-119, Biolegend), all Pacific Blue fluorochrome. After a washing step, cells were resuspended in 200 µL of PBS + 3% FBS and stained with 2 µL of c-Kit (APC, Clone 2B8, Bioscience) and Sca-1 (FITC, Clone D7, Bioscience) on ice for 20 minutes. Finally, cells were sorted on a FACSAria instrument and LSK cells collected in RLT-plus buffer (Qiagen) with 1% of β-mercaptoethanol to further isolate DNA and RNA. The gating strategy followed is indicated in Figure 2.2.

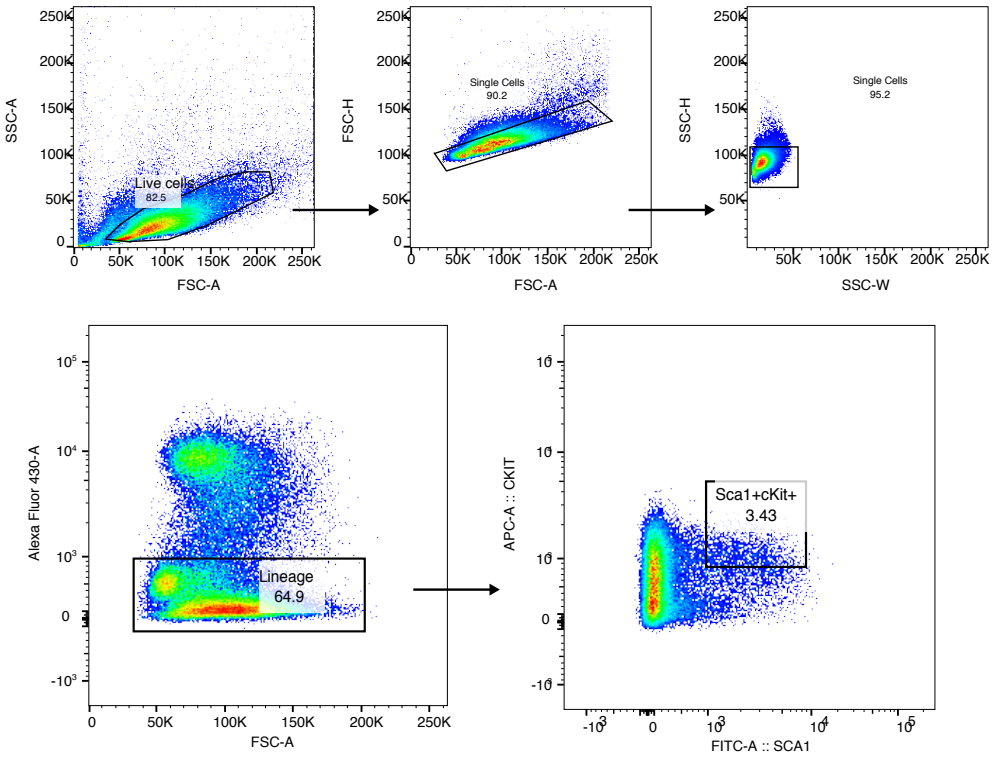

Figure. 2.2. Fluorescence-activated cell sorting of Lin-Sca+cKit+ (LSK) cells. Cells were isolated from bone marrow and enriched for CD117 marker. These were subsequently stained using lineages markers, Sca-1 and c-Kit. Live cells were selected in the first gate and then single cells were double gated. The negative portion of the lineage markers was selected and then the double positive cells for c-Kit and Sca-1 antibodies were sorted and collected in an eppendorf tube containing LRT buffer.

2.2.3.2 Neutrophil isolation

Approximately 1 mL of blood was extracted from retro-orbital bleeding and collected in 2 mL tubes with 100 µL of 0.5 M EDTA. The plasma was removed by centrifugation at 600 x g for 5 minutes. The pelleted cells were transfer to 15 mL falcon tubes and red blood cells were removed using 10 mL of 1x erythrocyte lysis buffer (10x BD-lysis-buffer, BD Bioscience, Cat#555899) diluted in ddH_2O and centrifuged at 300 x g for 5 minutes. Cells were resuspended in PBS and passed through a 40 µm strainer and centrifuged at 300 x g for 10 minutes. Neutrophils were isolated using Neutrophil Isolation Kit, mouse (MACS Miltenyi Biotec, Cat#130-097-658) following the manufacturer's instruction using the autoMACS® Pro Separator. The

fraction containing the enriched neutrophils was collected and centrifuged for 5 mins at 300 x g. Supernatant was removed and cells resuspended in 75 µL of RLT plus buffer (Qiagen) with 1% β-mercaptoethanol. The samples were stored at -70ºC until DNA and RNA purification.

2.3 DNA and RNA purification

DNA and RNA were simultaneously purified using AllPrep DNA/RNA Micro Kit (Qiagen) for mouse samples and AllPrep DNA/RNA Mini Kit (Qiagen) for human samples, following the manufacturer's instructions. DNA was resuspended in 60 µL of EB buffer using two rounds of 30 µL to improve recovery from the columns. RNA was eluted with 14 µL of RNase and DNase Free water, which yields an effective volume of 12 µL. Due to the limited number of $CD34^+$ cells, RNA extraction was performed with the PicoPure RNA isolation kit (Thermo Fisher Scientific). DNA and RNA were quantified with Quibit using dsDNA HS Assay Kit (Thermo Scientific) and RNA HS Assay Kit (Thermo Fisher Scientific), respectively. RNA quality was assessed with the Agilent 4200 Tapestation™ System (Agilent Technologies, BRC Genomics Facility) using Tapestation™ High sensitivity RNA ScreenTape (Agilent, Cat#5067-5579) following the manufacturer's instructions. The Tapestation™ System calculates the RNA integrity number (RIN) which represents the quality of the RNA according to an algorithm based on the electrophoretic trace of the RNA sample. This value can range between 1 and 10, 1 being poor quality RNA and 10 good quality RNA.

2.4 RNA sequencing

The workflow of the RNA-sequencing was followed as indicated in figure 2.3.

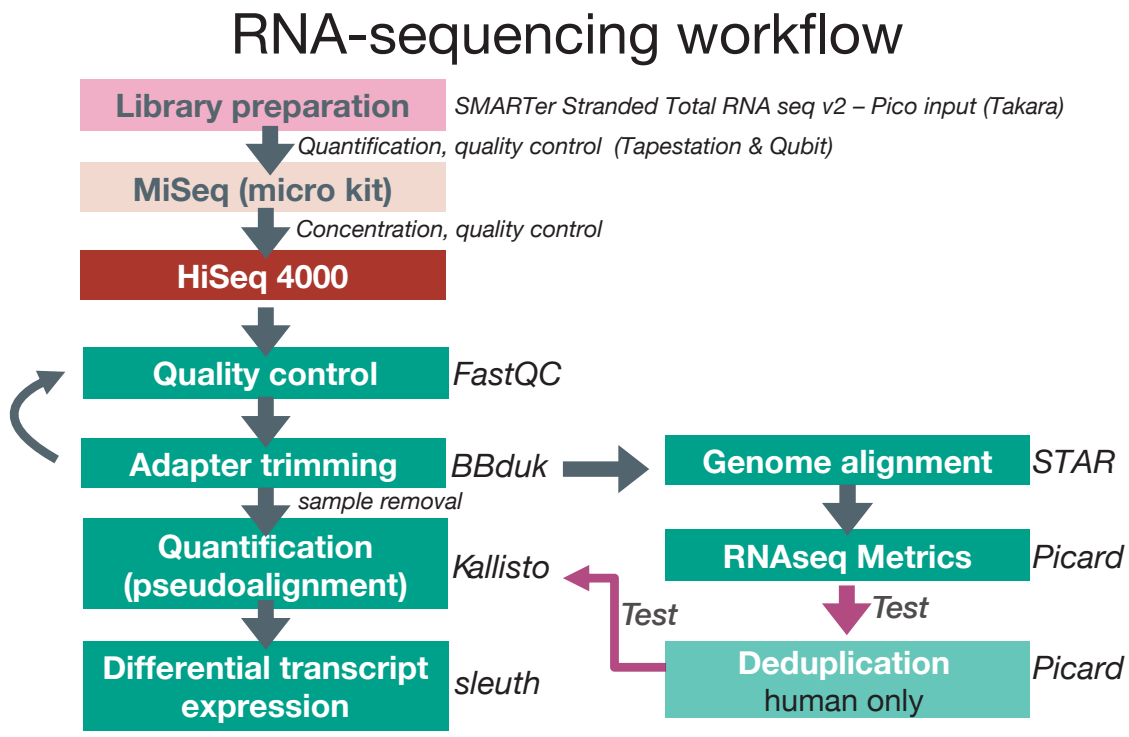

Figure. 2.3. RNA-sequencing workflow. Library preparation was performed using the SMARTer® Stranded Total RNA - Seqv2 - Pico Input (Clontech), MiSeq sequencing was performed to assess quality and concentration. All libraries were finally sequenced in HiSeq 4000. Raw reads were processed with FastQC. Adapters were trimmed using BBduk and reads were aligned to the reference genome using STAR. RNAseqMetrics were obtained using Picard. A deduplication step was performed in the human samples which was not included in the final pipeline (light green). After trimming, transcripts were quantified with Kallisto package and differential transcript expression was performed with sleuth package.

2.4.1 Human samples

Gene expression was assayed in neutrophils and $CD34^+$ cells from pre-treatment and following nine months of treatment. RNA concentration and quality were assessed as described previously (Section 2.3). For library preparation, equal amounts of RNA were used for each sample pair (pre- and post-treatment).

A total of 34 samples, corresponding to nine patients, were selected for sequencing (Table 2.1). Quality control (QC) of the samples using Tapestation™ revealed that some samples presented a low RNA integrity number (RIN), indicating degraded RNA. RNA degradation mostly affected the $CD34^+$ samples (Table 2.1). However,

these samples were not discarded as the library preparation kit used (SMARTer® Stranded Total RNA - Seqv2 - Pico Input Mammalian) was designed to work with low concentration (250 pg-10 ng) and degraded RNA by adjusting two steps from the standard protocol, according to the manufacturer's instructions (see Section 2.4.3.1).

Table. 2.1. RNA integrity number and concentration of $CD34^+$ cells and neutrophil samples

Patient	Sample	$CD34^+$			Neutrophils		
		ng/µL	RIN	Total ng	ng/µL	RIN	Total ng
P002	S01	0.14	4.6	2.8	22.8	6.6	456
	S04	0.276	2	5.52	1.36	8.4	27.2
P003	S01	0.532	8.4	10.64	0.246	7.7	4.92
	S04	0.214	2	4.28	0.289	7.6	5.78
P004	S01	0.239	3.1	4.78	0.251	7	5.02
	S04	0.121	2.3	2.42	0.452	8	9.04
P006	S01	0.336	8.1	6.72	6.48	8.5	129.6
	S04	0.922	3.2	18.44	2.32	9	46.4
P007	S01	0.399	8.6	7.98	12.1	8.4	242
	S04	0.228	7.1	4.56	1.3	8.8	26
P008	S01	4.86	8.7	97.2	1.45	8.8	29
	S04	0.435	8.5	8.7	0.227	7.3	4.54
P009	S01	7.5	8.7	150	12.8	7.7	256
	S04	0.741	8.8	14.82	0.604	7.1	12.08
P010	S01	0.355	8.1	7.1			
	S04	0.287	7.3	5.74			
P011	S01	0.174	7.7	3.48	3.16	8.2	63.2
	S04	0.476	9	9.52	19.4	8.8	388

RIN: RNA integrity number; red indicates low RIN

2.4.2 Mouse samples

Gene expression was assayed in neutrophils and LSK cells from each mouse treated with HU or vehicle. RNA concentration and quality were assessed as described previously (Section 2.3). For library preparation, 5 ng of RNA was used from LSK cells and 1 ng of RNA from neutrophils (Appendix D.1).

A total of 34 samples were selected for sequencing. The overall concentration of the neutrophils samples was low, and some were of poor quality as characterised by

a low RIN (Table 2.2). However, these samples were not removed.

Table. 2.2. RNA concentration (ng/µl) and quality (RIN) from LSK cells and neutrophils

Sample	Treatment	LSK			Neutrophils		
		ng/µl	RIN	Total ng	ng/µl	RIN	Total ng
MM_4964	VEH	2.37	8.7	28.4	0.197	7.8	2.4
MM_5276	HU	1.78	6.2	21.4	0.18	7.1	2.2
MM_5277	HU	3.25	9.1	39	0.15	6.3	1.8
MM_5322	HU	2.83	9.1	34	0.167	5.6	2
MM_5327	VEH	5.29	8.4	63.5	0.655	8.9	7.9
RB_427	VEH	2.06	8.5	24.7	0.235	6	2.8
RB_429	HU	1.38	8.2	16.6	0.116	3.7	1.4
RB_430	VEH	5.05	9	60.6	0.905	9.1	10.9
RB_431	HU	1.75	9.4	21	0.121	4.1	1.5
RB_438	VEH	3.44	8.4	41.3	0.133	2.4	1.6
RB_441	VEH	1.53	8.8	18.4	0.352	8.6	4.2
RB_444	HU	6.05	8	72.6	0.119	6.2	1.4
RB_447	HU	1.13	6.9	13.6	0.329	8.7	3.9
RB_458	HU	1.76	8.1	21.1	0.228	6.5	2.7
RB_460	VEH	2.44	8.2	29.3	0.215	6.7	2.6
RB_461	HU	1.89	5.5	22.7	0.398	6.9	4.8
RB_462	VEH	2.85	9.1	34.2	0.702	6.5	8.4
RB_470	VEH	3.1	9.2	37.2	0.0921	2.4	1.1
RB_481	VEH	1.28	9.3	15.4	0.648	8.2	7.8
RB_489	HU	1.95	8.8	23.4	0.183	5.7	2.2
RB_490	HU	1.96	9.2	23.5	0.566	8.3	6.8
RB_491	VEH	3.17	8.3	38	0.277	3.2	3.3
SX_581	HU	2.18	9.3	26.2	0.393	8.1	4.7

Low RIN indicated in red. VEH: vehicle; HU: hydroxyurea; RIN: RNA integrity number.

2.4.3 Library preparation

Libraries were prepared using the SMARTer® Stranded Total RNA - Seqv2 - Pico Input Mammalian (Clontech, cat#634411), following the manufacturer's instructions. Briefly, total RNA (RIN >4) was fragmented and random primers were used for the synthesis of the first-strand cDNA. Illumina adapters and barcodes were added in a first PCR using 3'- and 5'- primers for paired-end indexing (Table 2.3). The PCR products were purified using AMPure XP magnetic beads (cat#BECLA63881). Ribosomal cDNA, originated from rRNA (18S and 28S) and mitochondrial rRNA

(m12S and m16S), were cleaved by Zap-R in presence of mammalian-specific R-Probes. The libraries were further amplified by a second PCR using primers that are universal to all libraries (PCR2 primers v2). The number of cycles used were calculated according to the initial RNA concentration. A final purification step using AMPure XP magnetic beads was performed to finally elute the libraries with 12 µL of tris buffer.

Table. 2.3. Indexes used for library preparation. Libraries were paired-end indexed. i5 and i7 indexes were used in different combinations to obtain libraries uniquely indexed.

i5 Illumina Index Name (3')	i5 Bases for Sample Sheet MiSeq®, HiSeq® 2000/2500	i7 Illumina Index Name (5')	i7 Bases for Sample Sheet
D501	TATAGCCT	D701	ATTACTCG
D502	ATAGAGGC	D702	TCCGGAGA
D503	CCTATCCT	D703	CGCTCATT
D504	GGCTCTGA	D704	GAGATTCC
		D705	ATTCAGAA
		D706	GAATTCGT
		D707	CTGAAGCT
		D708	TAATGCGC
		D709	CGGCTATG
		D710	TCCGCGAA
		D711	TCTCGCGC
		D712	AGCGATAG

2.4.3.1 Steps modified for degraded RNA samples

The library preparation protocol was modified according to the concentration and RIN of each individual sample, following the manufacturer's instruction. One step was the first-strand cDNA synthesis, where samples with a RIN <4 were not fragmented enzymatically. In those samples with a RIN >4 the enzymatic fragmentation time was increased at higher RIN. The second step modified was the number of cycles in the last PCR amplification, which was adjusted according to the amount of input RNA. Details about the conditions used for each sample are listed in Appendix

B.1 for human samples and in Appendix D.1 for mouse samples.

2.4.4 Library quantification and QC

Libraries were quantified with Qubit dsDNA HS Kit (Thermo Fisher Scientific, Cat#Q32851) following the manufacturer's instructions. According to the concentration, the libraries were diluted to 1.5 ng/µl to evaluate the size distribution. Size distribution was assessed using the High Sensitivity D1000 ScreenTape (Agilent, Cat#5067-5584) on Agilent 2200 TapeStation (Agilent Technologies, BRC Genomics Facility) following the manufacturer's instructions. The libraries concentration was obtained by converting ng/µL to nM using the following formula:

$$Concentration\,in\,nM = \frac{Concentration(ng/\mu l)}{660g/mol\ \times\ average\,library\,size\,in\,bp} \times 10^6$$

According to the concentration in nM the libraries were diluted to 4 nM with Tris buffer and then each library pooled in a 0.25 mL eppendorf tube.

2.4.4.1 Human libraries

Following library preparation of all 34 human samples, concentration and the fragment distribution were assessed using Qubit™ and Tapestation™, respectively, as described previously (Section 2.4.4). According to the manufacturers of the library preparation kit, the expected size distribution of the libraries generated should range between 200 and 1000 bp, with a peak between 300 and 400 bp. According to the Tapestaion™ results, all libraries fell within this distribution range (Appendix B.1). Libraries with a shorter amount of product (between 150 and 200 bp) were cleaned again following the manufacturer's instructions. Three $CD34^+$ samples: P04S01, P02S04 and P06S04 and one neutrophil sample P04S01 fell into this category (Fig. 2.4A). The clean-up step was then validated by assaying the samples using the Tapestation™ (Fig. 2.4B): the P006S04 $CD34^+$sample was the only one where

cleaning was not successful (Fig. 2.4C). A further clean-up step was not attempted, in order to avoid losing extra sample material. Finally, the concentration for these libraries was re-calculated, and all libraries were then pooled in one aliquot. The Tapestation™ was used again to confirm library size, and Qubit™ was used to calculate final concentration.

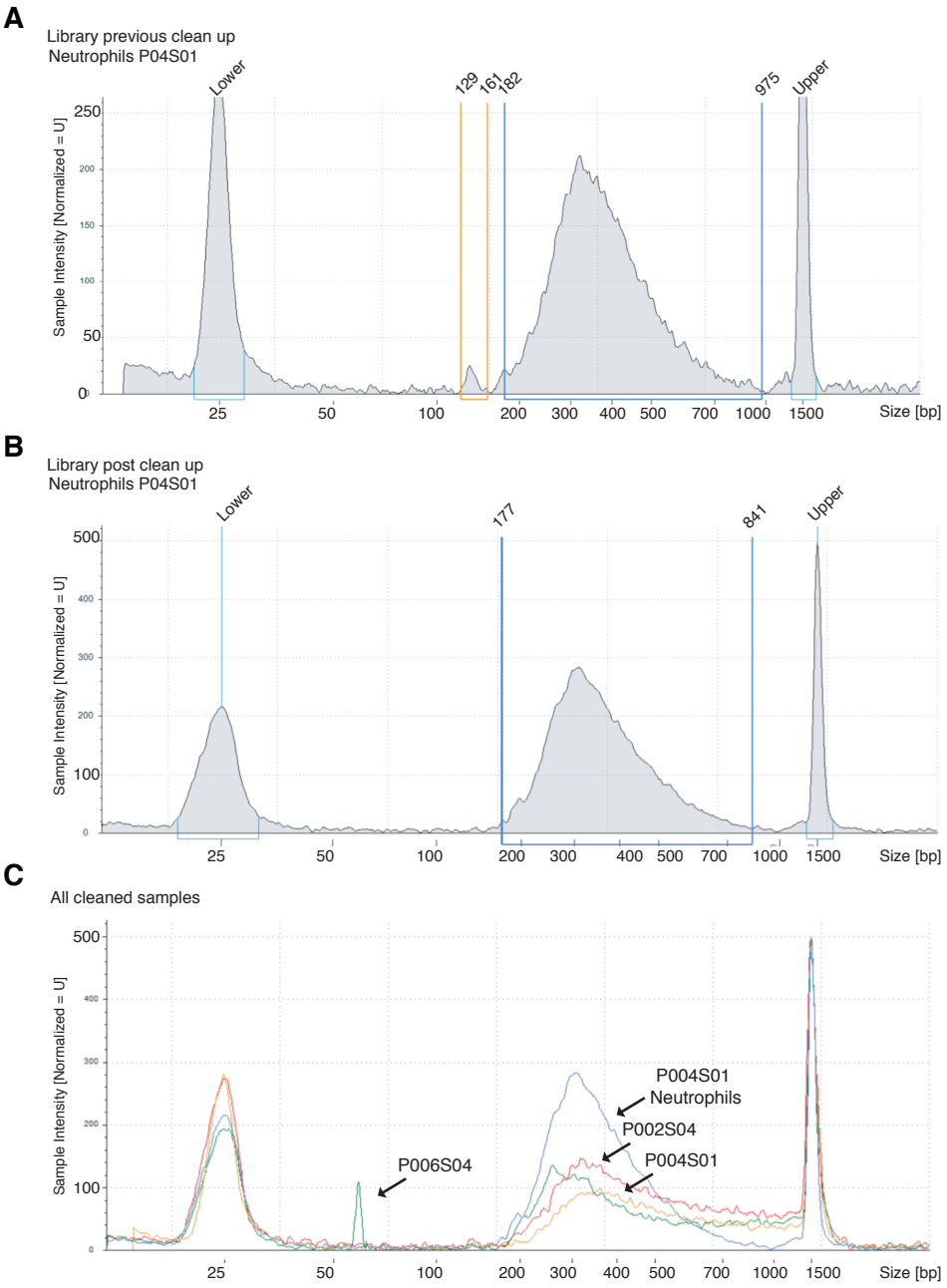

Figure. 2.4. Fragment size distribution of samples with adapter dimers. A) Fragment size distribution of the sample P04S01 (neutrophils) which shows a small peak between 100 and 200 bp. According to the manufacturer's instruction, these fragments can be removed by repeating the last clean-up step using magnetic beads. This procedure was performed on four samples. B) Fragment size distribution after adapter removal of the sample P04S01 (neutrophils). C) Fragment size distribution of all four samples after clean-up step; each sample name is indicated. Only sample P06S04 was not completely clean.

2.4.4.2 Mouse libraries

Following library preparation of all 34 mouse samples, the concentration and the fragment distribution was assessed using Qubit™ and Tapestation™ as described previously (Section 2.4.4)

Tapestation™ revealed that the distribution of fragment sizes was between the expected ranges according to the manufacturer's instructions (Fig. 2.5). Moreover, no adapter dimers were observed, therefore no additional cleaning steps were carried out.

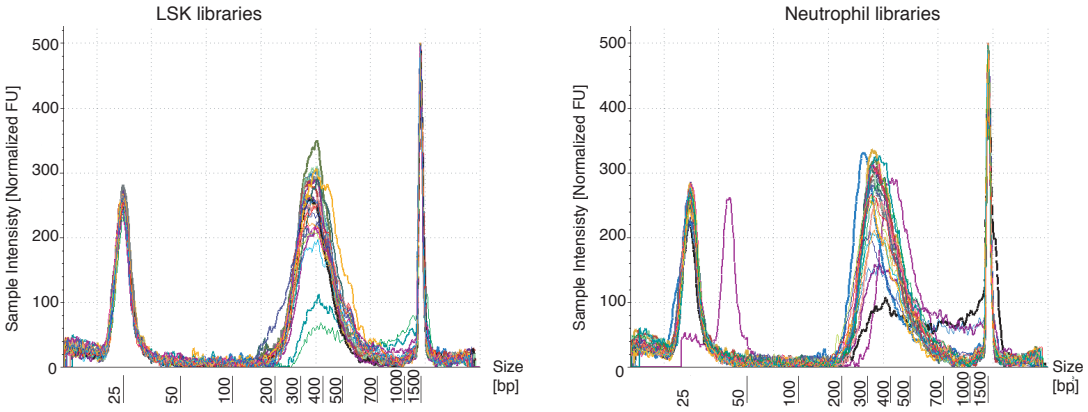

Figure. 2.5. RNA library fragment distribution from LSK cells and neutrophils samples . After library preparation, the libraries were analyzed with Tapestation™ to confirm their fragment size distribution and to evaluate if adapters dimers were remaining in the libraries. The fragment size distribution observed was as expected and no adapter dimers were identified.

2.4.5 Library normalization using low depth sequencing

The libraries were sequenced on an Illumina MiSeq (BRC Genomics Core Facility) using the Reagent v2 Micro Kit (Illumina), which yields *circa* four million reads in total. Raw reads were analyzed using FastQC (ver. 0.11.8) and the results were summarized using MultiQC (Ewels *et al.*, 2016). The number of sequences per library was considered as concentration of the library and this value was used to re-pool the libraries at the same concentration.

2.4.5.1 Low depth sequencing QC

FastQC uses the Phred score which represents the probability of obtaining an incorrect base call. This value should be maintained above 30 (one error in every 1000 base calls). Another metric for quality assessment is the percentage of GC content (%GC), which reveals the presence of high levels of ribosomal RNA (rRNA), or of microorganism contamination (Delhomme *et al.*, 2015). In the case of RNA libraries, the theoretical %GC distribution range is between 20- 60% when the libraries are PolyA enriched. Libraries with high levels of rRNA have a theoretical %GC distribution of between 40-70% (Delhomme *et al.*, 2015). In this case, rRNA depletion was used, since PolyA enrichment is not recommended for degraded samples. Sequence duplication was also evaluated. Duplication can occur during the library preparation, specifically at the PCR amplification step that over amplifies highly expressed transcripts. In the case of RNA libraries, it is common to sequence at high depth to get information from lowly expressed transcripts, creating more duplication (Delhomme *et al.*, 2015).

Human libraries: QC results All libraries had a Phred score above 30 (Fig. 2.6A). The GC content results revealed a distribution between 20 and 70% which is close to what is normally observed in RNA libraries (Fig. 2.6B). However, some libraries demonstrated quality issues as the distribution was skewed to the right and this was critical in the P003S04 sample from $CD34^+$ cells. With the MiSeq, sequencing depth was low, and one particular sample – once again, the P003S04 sample from $CD34^+$ cells had more duplication levels than the other samples (Fig. 2.6C).

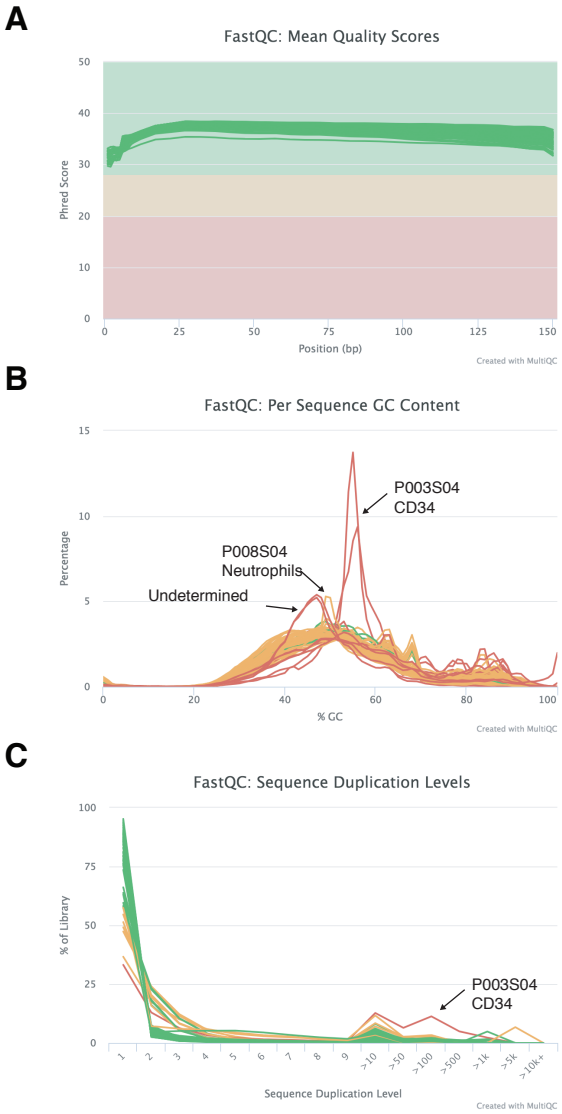

Figure. 2.6. FastQC results from CD34$^+$ and neutrophil libraries summarized with MultiQC. A) FastQC mean quality scores represents the Phred score which should be kept >30%. B) FastQC per sequence GC content. C) Duplication levels

Although some quality issues were observed in some samples, these were not removed from the high depth sequencing. The rationale for this was that all 34 libraries would be in the same lane for sequencing, and removing one or two samples would not greatly affect the overall results for the rest of the libraries. Accordingly, all libraries were normalized using the number of sequences obtained in the MiSeq sequencing, as described in the Methods Section (Section 2.4.5) (Fig. 3.2A), and re-pooled in one aliquot. All samples were sequenced and processed as described in the next section 2.4.6)

Mouse libraries: QC results All libraries had a Phred score above 30 (Fig. 2.7A). The GC content results were under the expected ranges (Fig. 2.7B). Similarly, no problems of duplication levels were observed (Fig. 2.7C). All samples were sequenced and processed as described in the next section 2.4.6)

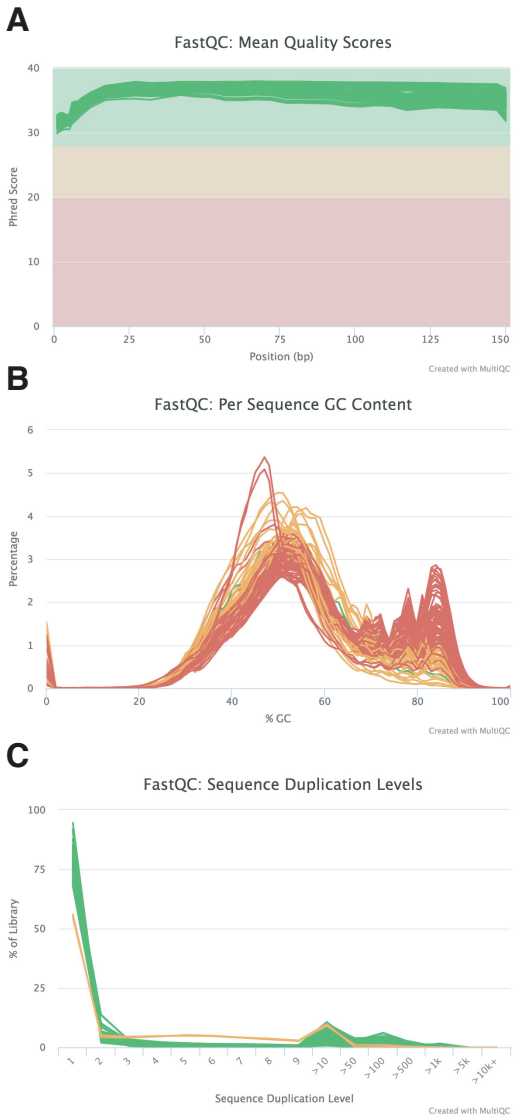

Figure. 2.7. FastQC results from LSK cells and neutrophils libraries summarized with MultiQC. A) FastQC mean quality scores represents the Phred score which should be kept >30%. B) FastQC per sequence GC content. C) Duplication levels

2.4.6 Hight depth sequencing and data processing

Libraries were sequenced in HiSeq4000 sequencing system (Genewiz). Read adapters were trimmed using BSTools (ver. 38.22 Wang *et al.*, 2015) and QC performed with FastQC. Reads were aligned to the corresponding reference genome (Hg38 and mm10, human and mouse respectively) using STAR (ver. 2.6.1b Dobin *et al.*, 2013). QC metrics were obtained using "collectRNAmetrics" from Picard (ver. 2.18.14). Deduplication was performed with "MarkDuplicates" from Picard and it was only applied to human samples as an alternative approach. Transcript abundance was quantified using Kallisto (ver. 0.44.0, Bray *et al.*, 2016) and Sleuth (ver. 0.30.0 Pimentel *et al.*, 2017) to identify differentially expressed transcripts.

2.4.6.1 Assessment of the normalization step using MiSeq: human

After normalization of the libraries using the approach described previously (Section 2.4.5) the libraries were sent to Genewiz for sequencing in a HiSeq4000 sequencing system. A total of four lanes were used to yield around 40 million reads per library. The number of sequences from HiSeq4000 were compared to the sequences from MiSeq (2.8A) to assess the effectiveness of the library normalization method employed (Fig. 2.8B). Homogeneity was observed among the sequences from HiSeq4000.

2.4.6.2 Assessment of the normalization step using MiSeq: mouse

The normalization of the libraries using MiSeq sequences was not completely homogeneous, as some samples have more sequences than others (Fig. 2.9A and B)

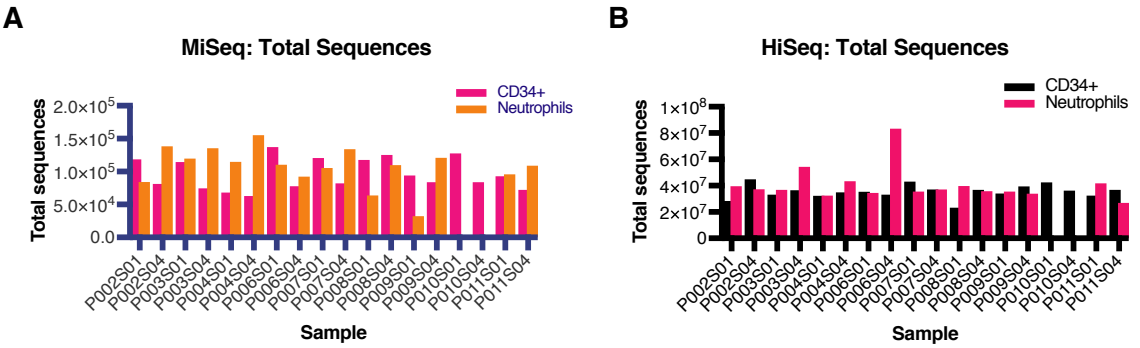

Figure. 2.8. Assessment of the normalization step using MiSeq . A) Total sequences obtained from Illumina MiSeq sequencing system in $CD34^+$ cells and neutrophils used to normalize concentration of each library. B) Total sequences obtained from Illumina HiSeq 4000 sequencing runs on $CD34^+$ and neutrophil libraries after using MiSeq sequencing system to normalize concentration. In general, homogeneity was observed among the libraries.

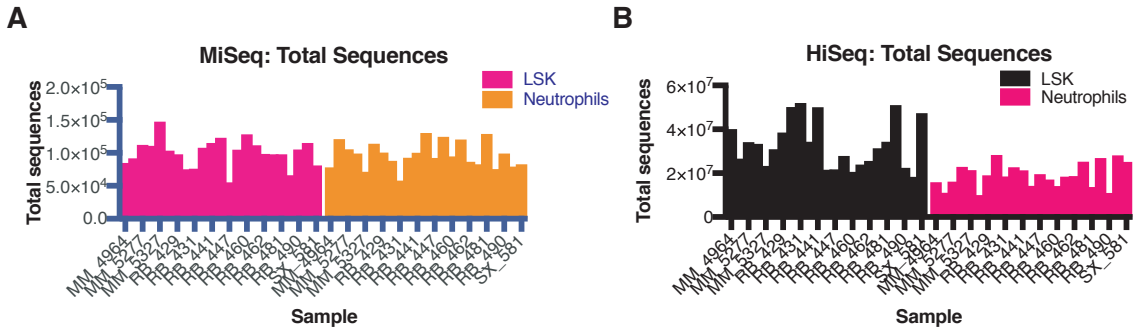

Figure. 2.9. Assessment of the normalization step using MiSeq: mouse libraries. A) Total sequences obtained from Illumina MiSeq sequencing system in LSK cells and neutrophils used to normalize concentration of each library. B) Total sequences obtained from Illumina HiSeq 4000 sequencing runs on LSK and neutrophil libraries after using MiSeq sequencing system to normalize concentration. Some libraries have greater sequencing depth than others, revealing a non-homogeneous concentration.

2.5 DNA methylation

2.5.1 Infinium MethylationEPIC BeadChip

2.5.1.1 Bisulfite conversion

Three hundred nanograms of DNA from neutrophils or up to 50 ng of DNA from $CD34^{+}$ cells were processed using the EZ DNA Methylation^{TM} Kit (Zymo Research, cat#D5001) following the manufacturer's instructions for Infinium assays. Samples were stored at -20^{o}C for no longer than one week before use.

The bisulfite converted DNA was sent to the BRC Genomic Core where DNA methylation was measured using the Infinium MethylationEPIC^{TM} BeadChip Kit (Illumina, cat#WG-317-1001). The samples were distributed across four EPIC BeadChips (eight samples per BeadChip).

2.5.1.2 Data processing

Raw files were processed in R (ver. 3.5.1). Workflow of the analysis was completed as indicated in figure 2.10.

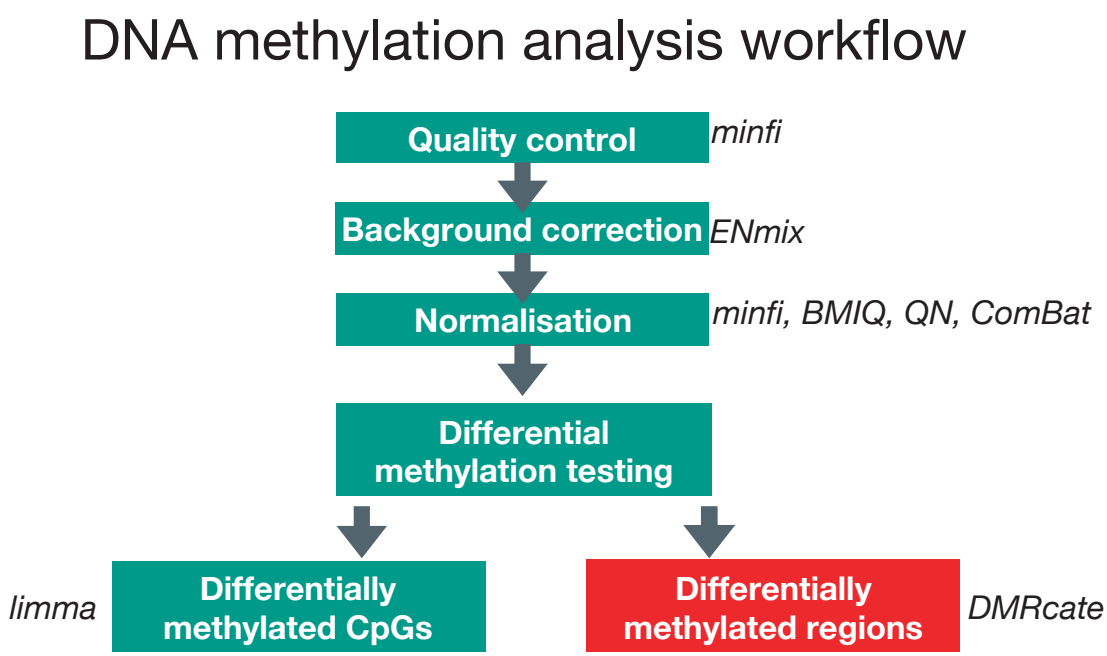

Figure. 2.10. DNA methylation workflow. Samples were analysed in R using minfi, ENmix and RnBeads packages for QC steps. Background correction was then applied using ENmix package. Then, four normalisation steps were carried out with minfi, BMIQ, quantile normalization (QN) and batch correction (ComBat). Finally, differentially methylated CpGs were obtained with limma package and differentially methylated regions with DMRcate package.

QC QC was performed using Minfi (ver 1.28.4, Aryee *et al.*, 2014), and RnBeads's Greedy cut algorithm (ver. 2.0.1, Assenov *et al.*, 2014). Samples, and probes with poor quality signals were removed.

Normalization The normalization steps are detailed in section 2.5.1.2. Briefly, background correction was performed using the out-of-band model from Enmix (ver. 1.18.2, Xu *et al.*, 2015). Subsequently, the probes were separated by type and color to perform quantile normalization and reduce probe-design type bias between Type I and Type II probes. After, the beta mixture quantile normalization (BMIQ) from wateRmelon package (ver. 1.26 Pidsley *et al.*, 2013) was used for correction of probe-design type bias affecting Infinium II probes. Finally, to remove the batch effect between samples, ComBat from sva package (ver. 3.30.1 Leek *et al.*, 2012) was used.

Differential methylation testing Differential methylation testing was carried out at two levels: I) differential methylation of individual CpG sites using limma (ver. 3.38.3 Ritchie *et al.*, 2015) and, II) differential methylation between regions, using DMRcate (ver. 1.18, Peters *et al.*, 2015) which agglomerates significant CpGs within a 1 kb region and computes a q-value. The testing was performed by using pair time point comparison (pre- versus three-, six- and nine- months of treatment) for each patient. Technical replicates were accounted in the differential methylation design with dupcorr from limma (Ritchie *et al.*, 2015).

Normalization steps and results The EPIC array has two types of probe which differ in the way that they measure DNA methylation intensity (Fig. 2.11). Type II probes utilise one probe per CpG, which is able to measure "methylated" (M) in the green color channel and "unmethylated" (U) in the red color channel. Type I probes, on the other hand, are essentially two probes per CpG (pairs), one measuring M and the other U. Type I probes can measure this either in the red or the green channel. Consequently, scanning the array generates six sets of measurements and each set with its own characteristic signal distribution: type II green, type II red, type I green M, type I green U, type I red M and type I red U.

Methylation of each CpG is indicated using the β value, which is calculated as $\beta = M/(M+U+\alpha)$ where $\alpha = 100$. Another means of quantifying methylation is through use of the M-value, $M = \log2(\beta/(1-\beta))$, which is the logit-transformed form of the β-value (Wang *et al.*, 2017).

Normalization steps are performed in order to reduce technical variance between samples, which can lead to confounding results when performing differential methylation testing. In order to be able to compare different samples, these need to have the same statistical properties. Accordingly, background correction was performed: this normalizes the signal of the tested probes using the signal from control probes which are integrated in the array. Multiple methods of background correction meth-

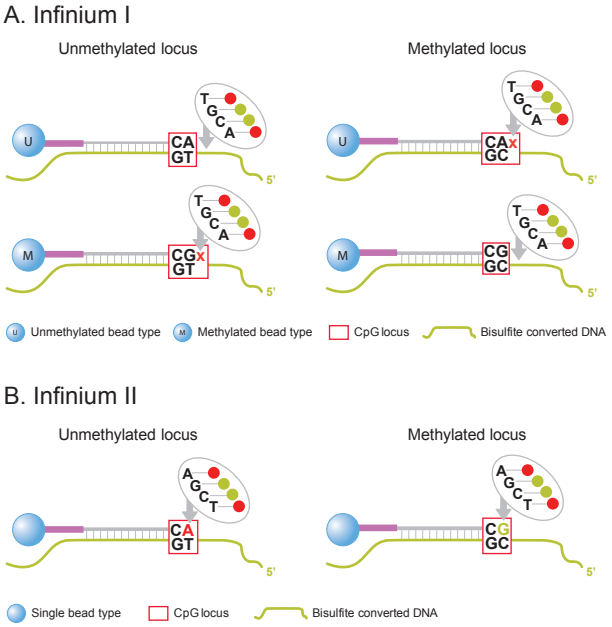

Figure. 2.11. Design of the EPIC array probes. A) Type I probes (or Infinium I) utilise two probes per locus, which will hybridize with the DNA according to the methylation status of the CpG assayed. B) Type II probe (or Infinium II) utilises one probe per locus, which will emit a different signal according to the methylation status of the CpG assayed. Image extracted from the online documentation of the Illumina MethylationEPIC BeadChip.

ods have been described. In this case, the ENmix package, which claims to perform better than previous methods, was used (Xu *et al.*, 2015) (Fig. 2.12A and Appendix B.1A).

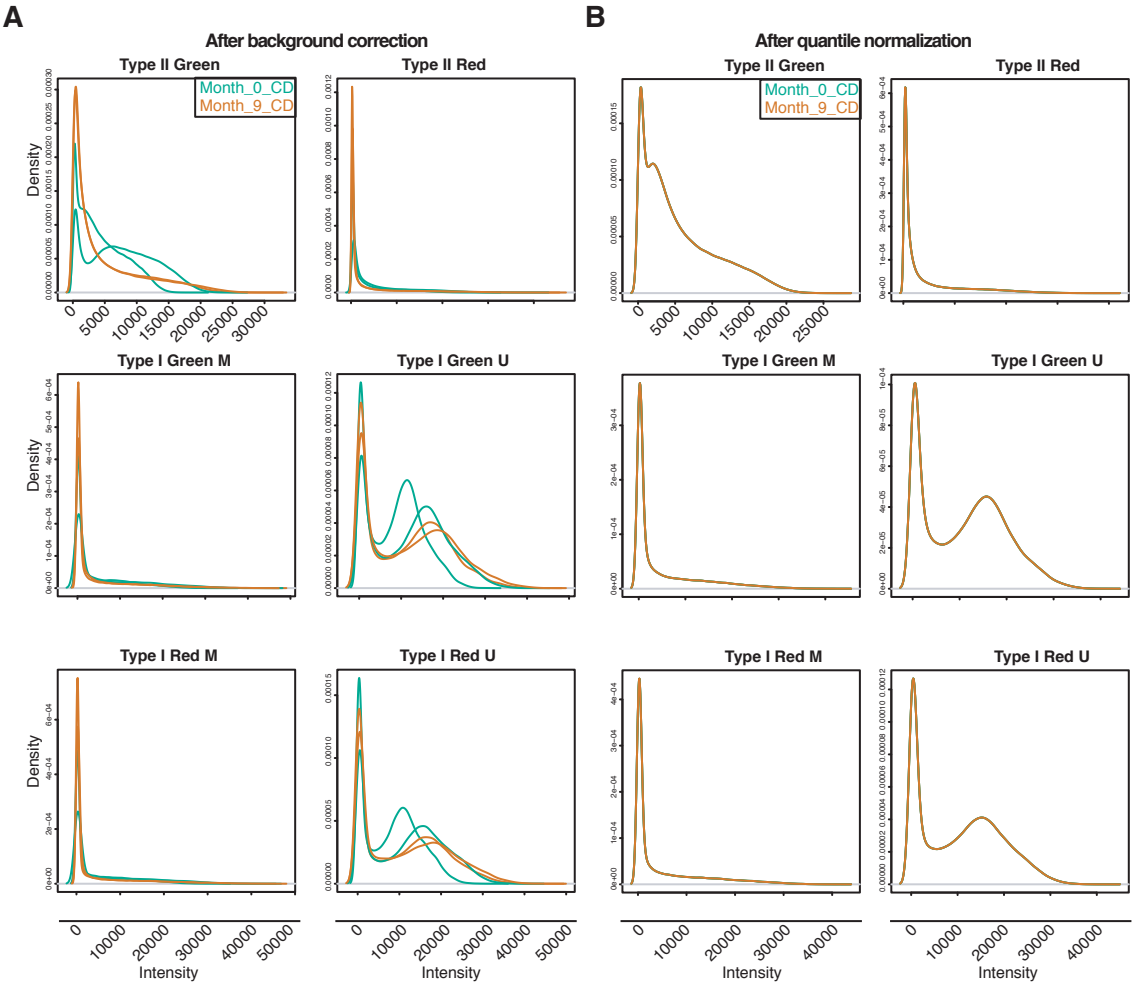

Figure. 2.12. Intensity plots of CD34$^+$ dataset before and after quantile normalization. Each graph represent a type of probe (Type I or Type II) and its channel (Red or Green). Type II probe utilises one probe able to emit signal in Red and Green channels. Type I probe utilises two probes that measure methylation (M) and unmethylation (U), each emitting signal in two channels, Red or Green. A) Distribution of the signal intensity before quantile normalization and after background correction. B) Distribution of the signal intensity after quantile normalization.

The common packages used to normalize EPIC arrays do not discriminate between the types of probes, which can lead to bias, due to differences in the distribution of probe intensity between probes, as observed in Fig. 2.12A. Accordingly, a method provided by Dr Reiner Schulz, which separates each type of probe, and

applies the quantile normalization method on each set of probes to make the distributions statistically identical, was applied to the dataset. It was observed that the intensity of the post-normalization signal was smoother than before, as expected (Fig. 2.12B and Appendix B.1B).

Finally, a further normalization was performed to reduce the differences between the type I and type II probes: this can be better appreciated using the beta value distribution, as shown in figure 2.13A. Multiple methods have been described to normalize this type of bias. In this case, the Beta-Mixture Quantile (BMIQ) Normalization from the wateRmelon package was used (Teschendorff *et al.*, 2013). This method yielded better similarities between the two types of probes and also helped to reduce sample-variation, compared to other methods tested (data not shown) (Fig. 2.13B).

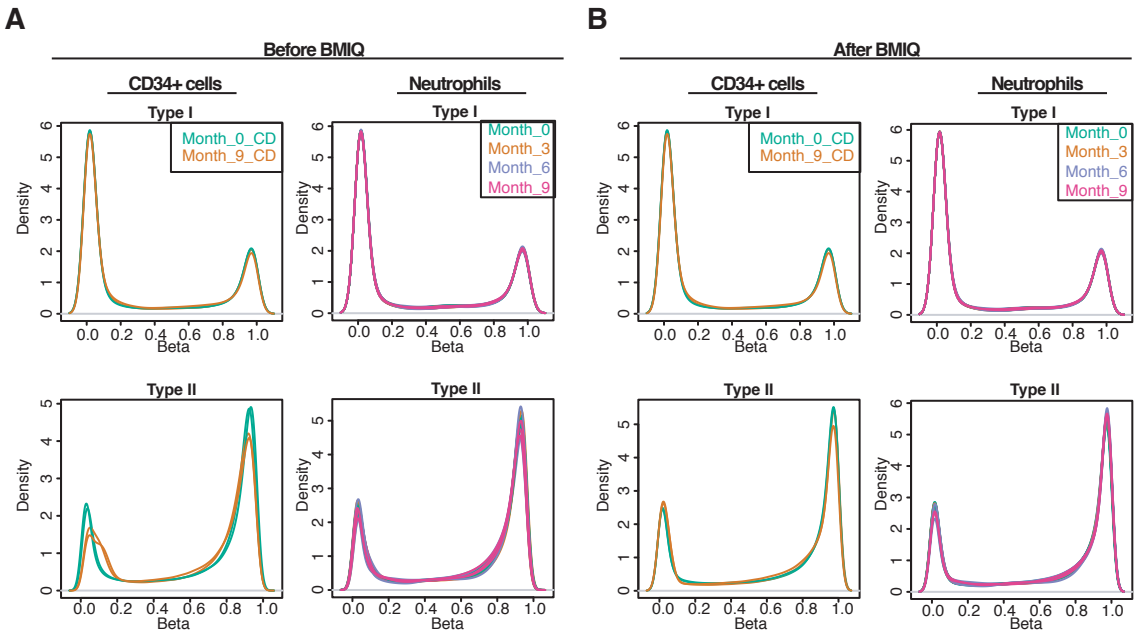

Figure. 2.13. Beta value distribution before and after BMIQ normalization. Type I and Type II probes measure DNA methylation in different ways, which leads to a different beta distribution (x-axis). BMIQ normalization is used to reduce the differences between these two types of probes. It also helps to reduce differences between patients. A) Before BMIQ normalization and B) after BMIQ normalization.

Combating the batch effect The EPIC array can assay eight samples per array. Every array is technically different from all others. Therefore, when analyzing

DNA methylation in more than eight samples, which requires additional arrays, the measurements from each array will not be fully comparable. This is known as a batch effect. In this study, differential methylation testing was performed in multiple samples from the same patient. Accordingly, if all samples from one patient were included in the same array, no batch effect correction would be needed. However, the neutrophil samples for some patients were distributed in different arrays, which led to undesirable technical differences.

First, to identify if batch and other known factors such as "patient" and "treatment" were the source of any variation, a function from ENmix - principal component regression (PCR) - was used. PCR performs a principal component regression analysis, using a data frame with factors (including "patient", "treatment", "batch", etc) per sample. In neutrophils, it was observed that the use of "patient" as a factor was driving the variation in the data (Fig. 2.14A). This information reinforced the importance of comparing patients with themselves. The "batch" factor was positioned in the second principal component, confirming that "batch" is also a factor that drives variation in the data. Accordingly, it was decided to remove the batch effect using the ComBat function from the sva package (Leek *et al.*, 2012). ComBat successfully removed the batch effect, and also increased the effect of "treatment" and "time" (Fig. 2.14B). In the case of $CD34^{+}$cells, "treatment" and "time" were the only factors that drove variation in the data (Fig. 2.14C).

After these steps, the data was tested for differentially methylated CpGs and regions as indicated previously 2.5.1.2.

2.5.1.3 Correlation with gene expression

The coordinates of the differentially methylated regions detected by DMRcate were subsetted by regions that mapped to CpG islands, shores, and shelfs, using GenomicRanges and the RnBeads annotation file as a reference. The genes associated with these regions were correlated with gene expression data.

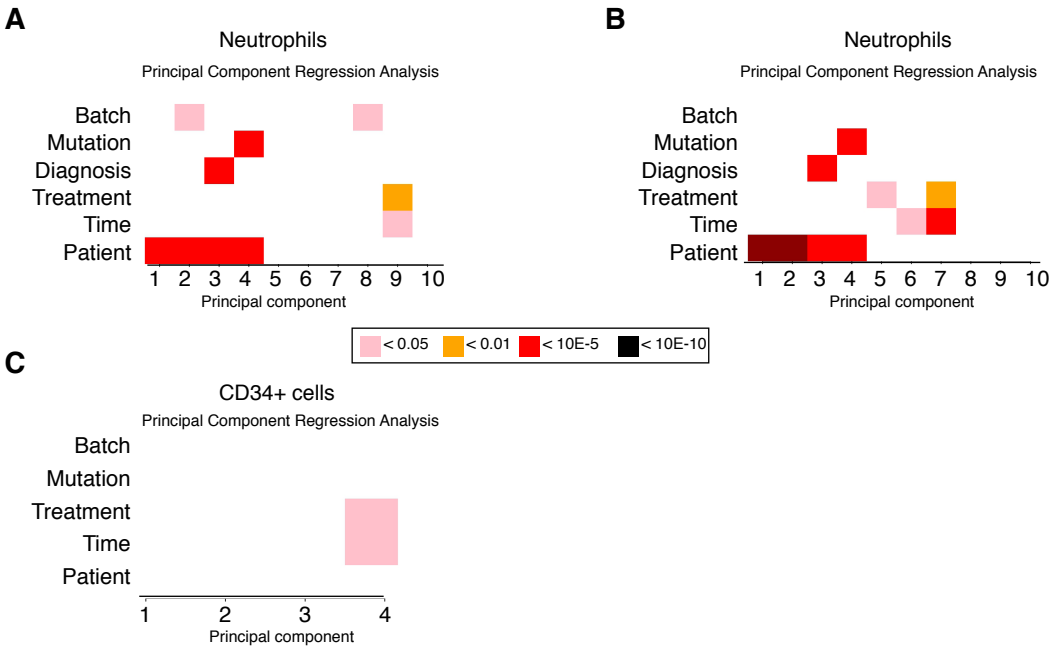

Figure. 2.14. Principal component regression analysis Principal component regression (PCR) analysis was performed to identify the source of data variation. A) PCR from neutrophils; a source of undesired variation was "batch", which was removed using the ComBat function from sva package in R. B) PCR from neutrophils after using ComBat, where "batch" is no longer driving variation. C) PCR from $CD34^{+}$cells, where "treatment" is a source of variation.

2.5.1.4 Enrichment analysis

The coordinates of the differentially methylated regions and differentially methylated CpGs were used to create a bed file and perform enrichment analysis. To identify enrichment for transcription binding sites, the online resource i-cisTarget (Imrichová *et al.*, 2015) was used. The regions were divided between hypo and hyper methylated prior submission. The configuration used was: minimum fraction of overlap, normalized enrichment score (NES) threshold, and AUC threshold were set to 0.4, 3.0, and 0.005, respectively.

2.5.1.5 Online databases

Processed DNA methylation data from a healthy donor (peripheral blood $CD34^{+}CD15^{-}$) was downloaded and formatted in a bed file to upload to the UCSC genome browser. The data was downloaded through the GSE106600 (Maupetit-Mehouas *et al.*, 2018).

These data was used to visually compare the levels of DNA methylation with the patients analyzed in this study.

Chapter 3

The effect of hydroxyurea treatment on DNA methylation and gene expression in polycythaemia vera and essential thrombocythaemia patients

3.1 Introduction

Myeloproliferative neoplasms (MPNs) are bone marrow disorders that give rise to an excess of blood cells. Classic MPNs include polycythaemia vera (PV), essential thrombocythaemia (ET) and myelofibrosis (MF). Patients with these disorders have a higher risk of suffering cardiovascular events, such as thrombosis or bleeding, or of developing more severe haematopoietic neoplasms, including acute myeloid leukaemia (AML). Hydroxyurea (HU) is the first-line treatment for high-risk PV and ET patients. The mechanism of action of HU is the inhibition of the ribonucleotide reductase enzyme (RNR) which blocks DNA synthesis, resulting in cell cycle arrest (Young and Hodas, 1964; Krakoff *et al.*, 1968). HU effectively reduces cell blood counts (CBC) and improves patients' symptoms (Harrison *et al.*, 2010). Another

effect of HU is the induction of foetal haemoglobin (HbF), explaning its use in sickle cell anaemia (SCA) patients. Several investigations have tried to understand the mechanism of action of HU in SCA, and some have suggested that HU is able to modulate DNA methylation (Walker *et al.*, 2011). However, these studies have been performed in a very site-specific manner and the mechanism is poorly understood.

DNA methylation is an important mechanism for gene regulation during development, and has been implicated in carcinogenesis (Esteller *et al.*, 2000). During haematopoiesis, correct DNA methylation appears to be essential for normal cell proliferation and differentiation (Challen *et al.*, 2011; Trowbridge *et al.*, 2009). In the case of MPN, somatic mutation in genes that are involved in DNA methylation have been identified. Moreover, the various therapeutic effects of HU have been linked to changes in DNA methylation, but no studies have been conducted in MPN patients to assess this effect.

Therefore, I investigated the effect of HU treatment on gene expression and DNA methylation in PV and ET patients. I collected blood samples from patients prior to HU treatment, and after three, six and nine months of treatment. I isolated two clinically relevant cell populations: $CD34^+$ cells, which are haematopoietic stem and progenitor cells, and neutrophils, which are fully differentiated cells from the myeloid lineage. The rationale underlying the choice of these two study populations is that $CD34^+$ cells in adults represent the most primitive cell type that will give rise to all cell lineages. By identifying the effects of HU at the stem cell level, may be predicted haematopoietic outcomes. In the case of neutrophils, these are significantly affected by HU treatment, are involved in the inflammatory response, and have been linked to thrombosis in MPN patients (Marin Oyarzún *et al.*, 2016; Stone *et al.*, 2018). Analysing of the effect of HU treatment on neutrophils would permit the identification of pathways related to the clinical outcomes of HU. Finally, by analyzing these two populations in parallel, it should be possible to recognize changes at the stem cell level that are translated to more differentiated cells, such as neutrophils.

Overall, through the work described in this thesis I sought to identify genes that are affected by HU treatment, with the aim of achieving a better understanding of its mechanism of action.

This chapter characterizes the patients included in the study, describes the generation of the data for gene expression and DNA methylation, and, lastly, reports the results of these analyses in terms of their biological significance.

3.2 Results

3.2.1 Patient sample collection and clinical response to HU treatment

For eighteen months, I took part in weekly clinical meetings held within the Haematology Department at Guy's Hospital. In collaboration with the clinicians, we identified patients suitable for inclusion in the study. These were patients who needed to commence HU treatment, who had not previously received any other cytoreductive treatment. Informed consent was obtained in accordance with the Declaration of Helsinki and local ethical guidelines (NRES Committee London - City Road & Hampstead, REC reference 15/LO/0265), and peripheral blood samples were then collected. During this time, I collected a total of 50 samples, distributed across four time points: prior to treatment, and after three-, six-, and nine-months of treatment. For some of these patients it was not possible to collect samples at the stipulated time point, and others switched to a different therapeutic pathway. A total of nine patients were deemed suitable for inclusion in the study (Table 3.1). Among these patients, three had been diagnosed with PV, and six with ET. Seven patients exhibited the JAK2V617F mutation. The remaining two patients (P004 and P010), did not have known mutations: these were termed "triple negative" (TN). All patients were regularly reviewed by the clinical team, and their clinical parameters were evaluated on each occasion.

The treatment response was evaluated by monitoring of their CBC at each time point (Fig. 3.1). Platelet, white blood cell and neutrophil counts decreased in all patients and most of them reached normal values at nine month of treatment (Fig. 3.1). In PV patients (dashed lines) the haematocrit was maintained below 45%. In

Table. 3.1. Clinical parameters of patients included in the study and the time points of sample collection

Patient ID	Age	Sex	Diagnosis	Mutation	Sample collection (time points)			
					Pre-treatment	3 months	6 months	9 months
P002	50	F	PV	JAK2	Yes	No	Yes	Yes
P003	55	F	ET	JAK2	Yes	Yes	Yes	Yes
P004	42	F	ET	TN	Yes	Yes	No	Yes
P006	67	F	ET	JAK2	Yes	Yes	Yes	Yes
P007	77	F	ET	JAK2	Yes	Yes	Yes	Yes
P008	88	F	ET	JAK2	Yes	Yes	Yes	Yes
P009	68	M	PV	JAK2	Yes	Yes	Yes	Yes
P010	77	F	ET	TN	Yes	Yes	Yes	Yes
P011	52	M	PV	JAK2	Yes	Yes	Yes	Yes

F: female, M: male; PV: polycythaemia vera; ET: essential thrombocythaemia
TN: triple negative; JAK2: *JAK2*V617F

ET patients (solid lines) platelet levels were kept below 400x10^9/L in most of the patients (Fig. 3.1). These observations are in accordance with those expected when PV and ET patients undergo HU treatment to reduce their risk of thrombotic events (Harrison *et al.*, 2010; McMullin *et al.*, 2019). Haemoglobin level is considered to be diagnostic of PV when it is higher than 160-165 g/L. In this cohort of patients, none had elevated haemoglobin levels prior to treatment, and some of them showed an increase in haemoglobin levels, within the normal range, following treatment (Fig. 3.1). White blood cell and neutrophil counts above 15 x10^9/L and 10 x10^9/L, respectively, increases the risk of thrombotic events (Harrison *et al.*, 2010). These values were present in one patient (P002), but HU treatment effectively reduced their cell counts. It has been documented that the mean corpuscular volume (MCV) normally increases when patients receive HU and such an increase was observed in these patients (Burns *et al.*, 1986; Spier, 1971) (Fig. 3.1). Although MCV is associated with the HU intake, this parameter was not used to assess treatment compliance. Overall, all patients were controlled as CBCs decreased throughout the study period. In a single patient (P011) haematocrits increased, but the remainder of the parameters (WBC, platelet and neutrophil counts) responded as expected.

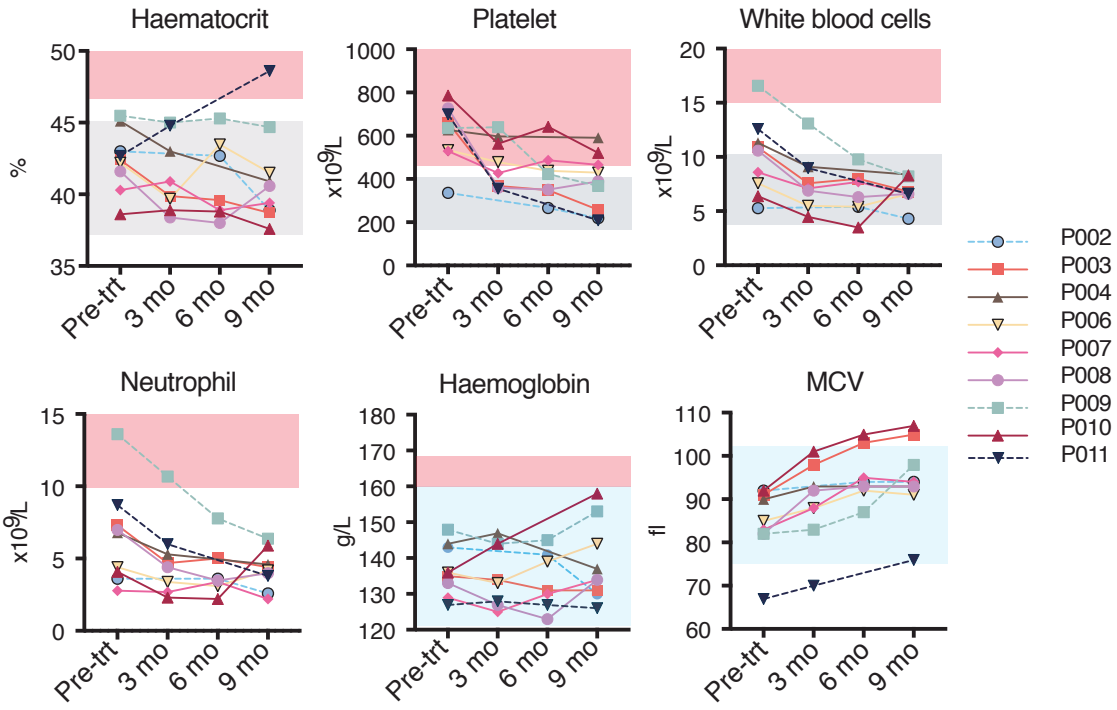

Figure. 3.1. CBC over the course of the study. CBC were monitored to assess the effectiveness of the HU treatment. Haematocrit, platelet, white blood cells, and neutrophil counts decreased in most patients. According to the aim of the treatment for PV (haematocrit <45%) and ET (platelets <400x10^9/L) patients, HU was effective in most patients. The haematocrit increased in one patient (P011) during HU treatment, however the other parameters were well controlled by HU. Haemoglobin levels were maintained within the normal ranges in most patients. The mean corpuscular volume (MCV) increased in all patients. Solid lines corresponds to ET patients and dashed lines to PV patients. Red shading corresponds with diagnostic criteria levels (Arber *et al.*, 2016). Grey shading represents the target level of the pharmacological treatment (McMullin *et al.*, 2019; Harrison *et al.*, 2010). Light blue shading correspond with reference values (Kambali and Taj, 2018).

3.2.2 Data generation

The data generation for this study did not follow the standard protocols and pipelines. These needed to be adapted to the design of the study and the quality of the samples. Some RNA samples were of poor quality and low yield. Therefore, a special library preparation kit designed for samples with these characteristics was used. The design of the study was also an important factor that needed to be considered when selecting the appropriate bioinformatic tools.

All the libraries and bioinformatic analyses were self-performed. The details and results of the library preparation steps are indicated in Methods Section 2.4. In this section, the details about the gene expression analysis and QC steps are presented.

Similarly for DNA methylation analysis, the bioinformatic pipeline used was designed especially for this study. Therefore, different tools were combined to ensure that the differentially methylated results were meaningful for the hypothesis of this project. In the following section (3.2.2.2), the microarray used for the DNA methylation assay is introduced and the QC steps conducted are detailed.

3.2.2.1 Gene expression analysis

Data processing Four sequencing lanes in the HiSeq4000 were used to yield around 40 million reads per library. The raw reads were processed as described in the Methods Section 2.4.6. The libraries were analyzed with FastQC to obtain QC metrics (Appendix B.2) and then aligned to the reference genome (Hg38) using the STAR package (Table 3.2). This alignment allowed further QC metrics to be obtained using the "CollectRnaSeqMetrics" tool from Picard (Appendix B.3). The QC results showed some libraries had high duplication levels (>85%), which were correlated with low RNA integrity number (RIN), a measure of DNA degradation and yield (Fig. 3.2A). These results showed that some samples with a high percentage of duplication also had high levels of ribosomal RNA (rRNA) (Fig. 3.2B). To

avoid losing samples from the analysis, removal of duplicates was attempted using the "MarkDuplicates" tool from Picard. This procedure was successfully carried out, but the number of reads left was too low to permit any statistical assessment of these samples (Appendix B.4). Therefore, samples with a duplication rate of >85% were removed; these included six $CD34^+$samples (P002S01, P002S04, P003S04, P004S01, P004S04 and P006S04) and one neutrophil sample (P004S01) (Fig. 3.2B).

After QC, the Kallisto package (Bray *et al.*, 2016) was used to quantify transcript abundance. This package uses an accurate pseudo-alignment method which requires less computational power, and the pipelines were well established in our research group. Finally, differentially expressed genes (DEGs) were identified using Sleuth package (Pimentel *et al.*, 2017).

Table. 3.2. Percentages of uniquely mapped reads of $CD34^+$ cells and neutrophils. Raw reads were processed and aligned to the reference genome (Hg38) using the STAR package.

ID	Uniquely mapped reads	ID	Uniquely mapped reads
P002S01CD34	54.24%	P002S01NEU	70.47%
P002S04CD34	66.75%	P002S04NEU	62.92%
P003S01CD34	58.69%	P003S01NEU	57.30%
P003S04CD34	28.76%	P003S04NEU	58.52%
P004S01CD34	49.56%	P004S01NEU	64.07%
P004S04CD34	70.84%	P004S04NEU	60.30%
P006S01CD34	60.38%	P006S01NEU	62.35%
P006S04CD34	54.56%	P006S04NEU	62.40%
P007S01CD34	59.91%	P007S01NEU	65.40%
P007S04CD34	63.42%	P007S04NEU	63.47%
P008S01CD34	61.49%	P008S01NEU	68.01%
P008S04CD34	63.67%	P008S04NEU	65.25%
P009S01CD34	66.52%	P009S01NEU	69.71%
P009S04CD34	58.58%	P009S04NEU	70.14%
P010S01CD34	55.19%		
P010S04CD34	56.82%		
P011S01CD34	58.27%	P011S01NEU	59.45%
P011S04CD34	53.42%	P011S04NEU	57.08%

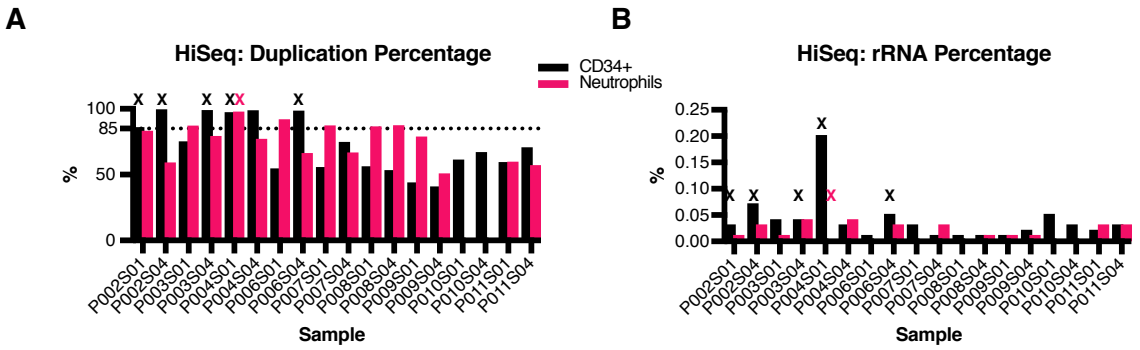

Figure. 3.2. QC results from sequencing. A) Duplication percentages and B) rRNA percentages from RNAseqMetrics from Picard. Samples removed from differential transcript analysis are noted with an X.

3.2.2.2 DNA methylation processing and QC results

DNA methylation was measured in $CD34^+$ cells and neutrophils using the Infinium MethylationEPIC^{TM} BeadChip (EPIC). The EPIC methylation array measures quantitatively the methylation of >850,000 cytosines genome-wide. DNA is chemically treated with bisulfite to convert all unmethylated cytosines into uracil. The bisulfite converted DNA is then hybridized to the EPIC array. The EPIC array has two types of probes, Type I and Type II, which measure DNA methylation (Fig. 2.11). Type I utilises two probes per locus, one probe detects a methylated CpG, and the other an unmethylated CpG. Depending on the methylation status of the locus and the CpG analyzed, it will hybridize with one of the two probes (Fig. 2.11A). By contrast, Type II utilises a single probe per locus, which is able to generate different signals according to the methylation status of the CpG assayed (Fig. 2.11B).

For the bisulfite conversion, the manufacturer of the EPIC array recommended using a minimum of 300 ng of DNA for this assay. In this study, due to the limited amount of DNA obtained from $CD34^+$ cells, only three patients were assayed prior to treatment (S01), and nine months of HU treatment (S04) using all the DNA available (<50 ng) (Table 3.3). In neutrophils, eight patients were assayed prior to treatment, and at three months (S02), six months (S03) and nine months of HU

treatment, using 300 ng of DNA (Table 3.3).

The bisulfite converted samples were submitted to the BRC Genomics Facility for the EPIC assay. The raw files were processed in R as described in Methods Section 2.5.1.2.

Table. 3.3. Samples included in DNA methylation analysis

Patient	Pre-treatment (S01)	3 months (S02)	6 months (S03)	9 months (S04)
P002	CD34 (removed) & NEU	NEU	NEU	CD34 & NEU
P003	CD34 & NEU	NEU	NEU	CD34 & NEU
P004	CD34 & NEU	NEU	NEU	CD34 & NEU
P006	NEU	NEU	NEU	NEU
P007	NEU	NEU (removed)	NEU	NEU
P008	NEU	NEU	NEU	NEU
P009	NEU	NEU	NEU	NEU
P011	NEU	NEU	NEU	NEU

NEU: neutrophils

QC of the DNA methylation assay: samples QC and probe removal The quality of the data was evaluated by using the detection p-values, which indicate the quality of the signal of each probe in the sample. The minfi package (Aryee *et al.*, 2014) was used to obtain these values, which compare the signal of each probe to the background signal level obtained from the negative control probes. Samples with a mean p-value >0.05 were considered to have a poor quality signal, and were excluded from the analysis. Accordingly, P002S01 from $CD34^+$ and P007S02 from neutrophils were removed from the analysis (Fig. 3.3A). Probes with quality issues (p-value $\geq$ 0.000001) were also removed (Lehne *et al.*, 2015). Additionally, two other packages, RnBeads (Assenov *et al.*, 2014) and ENmix, (Xu *et al.*, 2015), which use algorithms based on the p-value detection previously calculated, were used to identify bad probes. ENmix does not work properly on small numbers of samples, and was not used in the $CD34^+$ dataset. In neutrophils, ENmix overestimated the number of probes, compared with the other methods (Fig. 3.3B). Therefore, only

those probes identified as poor performers by RnBeads and minfi were removed (Fig. 3.3B). Cross-reactive probes, which are probes that map to multiple sites in the genome, and probes that are positioned in common SNPs are usually removed when comparing samples from different patients. However, in this study design it was intended that differential methylation testing should be performed in the same patient and, consequently these probes were not removed.

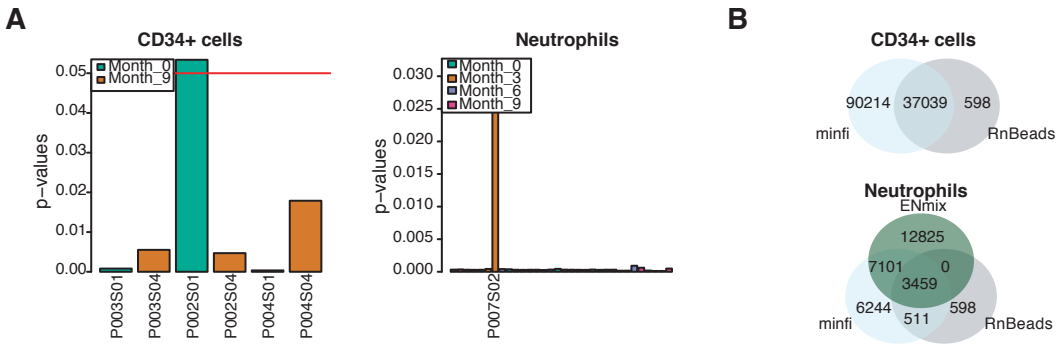

Figure. 3.3. QC results from DNA methylation analysis. The p-value detection values were calculated using the minfi package. A) Mean p-value per sample. Samples with p-value >0.05 were removed from further analysis. B) Probes with quality issues were also removed from the analysis. Three packages were used to identify poor quality probes in neutrophils and two packages in $CD34^{+}$ cells. Probes identified with RnBeads and minfi were removed.

After sample removal, normalization steps were conducted to reduce technical variance that are inherent to microarrays. These normalization steps are described in detail in Methods Section 2.5.1.2. Briefly, four normalization steps were applied: I) background correction which uses the signal from control probes included in the array to normalize the signal from the tested probes; II) quantile normalization (QN) method designed by Dr Reiner Schulz that normalize the signal of each type of probe per color channel; III) Beta-Mixture Quantile (BMIQ) Normalization, that corrects the differences between Type I and Type II probes; IV) Batch correction between arrays.

Finally, differential methylation testing was conducted with limma package (Ritchie *et al.*, 2015) to identify differentially methylated CpGs and DMRcate package (Peters *et al.*, 2015) to identify differentially methylated regions (DMRs) as described

in the Methods Section 2.5.1.2.

3.2.3 The effect of HU treatment on gene expression

Differentially expressed genes (DEGs) were identified using Sleuth to test whether the expression levels of any given gene across patient samples were significantly better explained by a linear model comprising patient and time point (pre-treatment versus nine months post-treatment) factors, as opposed to a model with only the patient factor. The expression difference of the significant DEGs (qval <0.05) between "before" and "after" treatment per patient were calculated. These values were hierarchically clustered (Euclidean correlation distance and ward.D2 agglomeration method), and gene set enrichment analysis was performed for each gene cluster using Reactome.org (Fabregat *et al.*, 2017) and enrichR (Kuleshov *et al.*, 2016) with gene sets defined by biological pathway or by the target of transcription factors (TFs).

A total of 18 samples from $CD34^+$ cells were sequenced and six were removed from the analysis for QC-related issues (Section 3.2.2.1). The removal of samples left five patients with paired samples at different time points. Enrichment analysis for pathways showed that downregulated genes were enriched for protein translation terms, due to a high number of ribosomal protein genes (Fig. 3.4A). Upregulated genes were enriched mainly for immune system related terms including cytokine signalling and neutrophil degranulation (Fig. 3.4A). TF enrichment analysis identified SPI1 (also known as PU.1) and RUNX1 as regulators of most of the gene clusters. The genes encoding for these TFs were also differentially expressed, *SPI1* was upregulated and *RUNX1* was downregulated (Fig. 3.4A). Interestingly, these TFs are known to be master regulators of haematopoiesis (Imperato *et al.*, 2015).

A similar analysis was performed on neutrophils. In this case, 16 samples were sequenced, and one was removed (Section 3.2.2.1). Clustering of samples revealed variability between PV and ET patients, where PV patients showed lower expression change compared to ET patients (Fig. 3.4B). As seen in $CD34^+$ cells, gene

clustering analysis identified that downregulated genes were enriched for protein translation terms and upregulated genes were enriched for immune system related terms, including interleukin signalling and neutrophil degranulation. Enrichment for TFs identified SPI1, ZBTB7A and CEBPD as regulator of upregulated genes and all three were also differentially expressed.

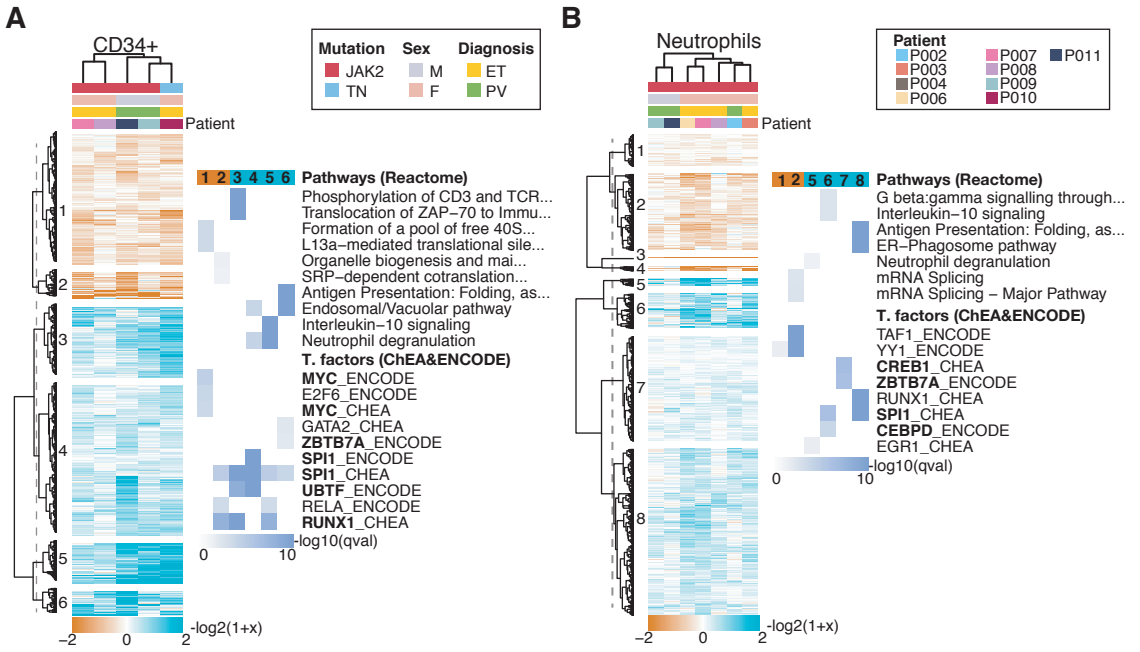

Figure. 3.4. Gene expression analysis of human CD34$^+$ cells and neutrophils. Significant differentially expressed genes (DEGs) (q-value <0.05) were clustered according to their log2-fold change expression (Euclidean correlation distance and ward.D2 agglomeration method). Clusters were divided at the same branch heights and numbered. Enrichment analyses for pathways and transcription factors were performed in each cluster. The top significantly enriched results from each cluster are represented as a heatmap plot where color scale indicates significance. A) DEGs from CD34$^+$ cells divided into 6 clusters; 1 and 2 represent downregulated genes and 3 to 6 represent upregulated genes. B) DEGs from neutrophils divided into 8 clusters; 1 and 2 represent downregulated genes and 3 to 8 represent upregulated genes. Differentially expressed genes are shown in bold type. JAK2: JAK2V617F mutation; TN: triple negative; M: male; F: female; ET: essential thrombocythaemia; PV: polycythaemia vera.

These enrichment results demonstrate that similar pathways are being modulated by HU treatment at two differentiation stages. In line with these similarities, common DEGs between CD34$^+$ cells and neutrophils were examined. Enrich-

ment analysis of upregulated genes confirmed the terms seen in CD34$^+$ cells and neutrophils. These terms were neutrophil degranulation and antigen presentation, among others (Fig. 3.5). Genes encoding important transcription factors involved in haematopoiesis such as *RARA*, *SPI1*, *CEBPB* and *ZBTB7A* (all upregulated) were also identified. Significant terms of downregulated genes included pathways related to protein translation, as seen in the previous analysis of CD34$^+$ cells and neutrophils.

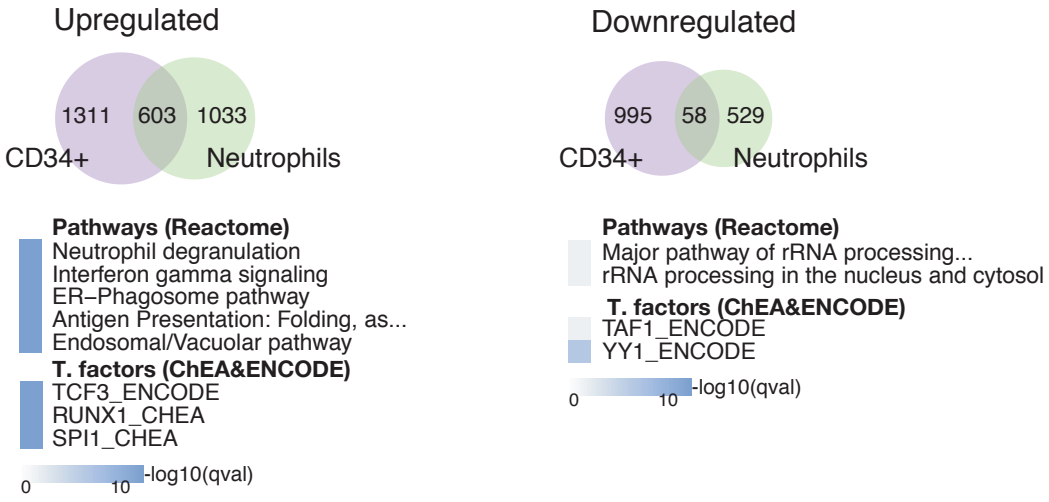

Figure. 3.5. Total number of common differentially expressed genes between CD34$^+$ cells and neutrophils and results of enrichment analysis. Differentially expressed genes from each cell type were compared and enrichment analysis was performed in separately for upregulated and downregulated genes.

3.2.3.1 The effect of HU in inflammation

Chronic inflammation in MPN has been directly linked to the somatic mutations that affect the JAK-STAT pathway. HU treatment helps to reduce inflammation, decrease expression of endothelial adhesion molecules, and induce nitric oxide (NO) production (Gambero *et al.*, 2007). Neutrophils detect the endothelial adhesion molecules allowing them to migrate to the site of injury where they release NO to kill microorganisms. However, the enrichment analysis previously carried out in neutrophils did not reveal any significant enrichment of pathways involved in NO

release or endothelial adhesion.

To further investigate the reduction of inflammation by HU, the analysis was focused towards canonical pathways of inflammation which included cytokine signalling e.g. by interferon and interleukin signallings. According to the previous enrichment results, most of the significantly enriched terms identified corresponded with immune system functions (Fig. 3.4). Therefore, the entire list of up- and downregulated was re-submitted to the pathway enrichment tool (Reactome.org) to identify significantly enriched pathways involved in inflammation, such as cytokine signalling.

In $CD34^+$ cells, there was no enrichment of downregulated genes involved in immune system responses that could lead to inflammatory pathways. Among the interferon pathways, the upregulated genes were significantly enriched for interferon gamma (INF-γ) signalling. Genes that are necessary for INF-γ responsiveness (*IFNGR1* and *IFNGR2*) were identified, together with genes involved in the phosphorylation and activation of INF-γ (*JAK1, JAK2* and *STAT1*). DEGs regulated by INF-γ stimulation included interferon regulatory factors (*IRF1*, *IRF7* and *IRF8* genes), promyelocytic leukemia (*PML*) and SP100 nuclear antigen (*SP100*), which together play roles in cell growth, differentiation and apoptosis. Differentially expressed negative regulators of the interferon pathways included *PTPN1, PTPN*2, and *PTPN11* genes which block INF-γ signalling. Although INF-γ is involved in inflammatory responses, evidence indicates that in hematopoietic stem cells (HSCs) it induces cell proliferation and differentiation, and can also stimulate apoptosis when HSCs are exposed to stress stimuli. Accordingly, these findings in $CD34^+$ cells might not be linked directly to the inflammatory response (Morales-Mantilla and King, 2018).

Interleukin signalling is another important pathway in inflammation. Studies in ET and PV patients have shown elevated levels of pro-inflammatory cytokines and deregulated anti-inflammatory cytokines (Pourcelot *et al.*, 2014; Vaidya *et al.*,

2012; Panteli *et al.*, 2005). In upregulated genes, significant enrichment of IL-10 signalling was observed, which is part of the nuclear factor kappa-light-chain enhancer of activated B cells (NF-κB) pathway, and is considered to be an anti-inflammatory interleukin (Iyer and Cheng, 2012). However, IL-10 signalling in HSCs has been proposed to increase self-renewal (Kang *et al.*, 2007).

In neutrophils, as in $CD34^+$ cells, downregulated genes were not enriched for genes involved in immune system responses. In upregulated genes, INF-γ and interleukin signallings were significantly enriched. Among the negative regulators of the INF-γ pathway, *SOCS3* and *PTPN6* genes were identified as genes which block INF-γ signalling. DEGs regulated by INF-γ stimulation included *IRF1*, *IRF7*, *IRF9*, *PML* and *SP100*, as seen in $CD34^+$ cells. Among interleukins, significant enrichment of IL-4 and IL-13 signalling was observed. *IL4R*, *IL13RA1*, *JAK3*, *STAT3*, *TYK2* were some of the genes shown to be upregulated. IL-4 and IL-13 have been reported to restrict neutrophil expansion and infiltration into tissues (Seki *et al.*, 2012; Woytschak *et al.*, 2016).

In conclusion, HU perturbed genes involved in INF-γ and interleukin signallings pathways in both $CD34^+$ cells and neutrophils. In $CD34^+$ cells, the upregulated DEGs identified were mostly involved in positive regulation of INF-γ. In neutrophils, conversely, two types of negative regulators of INF-γ were upregulated (the SOCS and PTP families). In the interleukin signaling pathway, upregulated genes from both cell types were enriched in anti-inflammatory interleukins signalling.

3.2.3.2 The effect of HU on haemoglobin synthesis

MPN patients treated with HU usually exhibit increased haemoglobin levels, hence its therapeutic application in sickle cell anaemia (SCA). In SCA, HU increases foetal haemoglobin (HbF), which helps to ameliorate patients' symptoms. Exhaustive research has been carried out in an attempt to elucidate the molecular mechanism of HbF induction. Some studies have shown that TFs such as BCL11A and ZBTB7A

repress the expression of the γ-globin gene (HBG1), which is responsible for HbF synthesis in erythroid cells (Zhou *et al.*, 2010; Norton *et al.*, 2017). These findings have led to proposals that the mechanism of action of HU involves the downregulation of BCL11A and ZBTB7A (Chondrou *et al.*, 2018). In this $CD34^+$ dataset, *ZBTB7A* gene was upregulated and *BCL11A* gene was downregulated following HU treatment. It is possible that $CD34^+$ cells do not constitute the optimal study population for attempts to elucidate the role of HU in HbF induction, since they are not fully committed erythroid progenitor cell differentiation. However, it is possible that *BCL11A* downregulation at the stem cell level is necessary, and that it is maintained throughout differentiation, as seen in *in vitro* assays (Chang *et al.*, 2015).

3.2.3.3 The effect of HU on the cell cycle

Inhibition of the ribonucleotide reductase enzyme (RNR), which leads to cell cycle arrest, constitutes the main mechanism of HU action. Accordingly, HU has been used to elucidate the mechanisms of checkpoint activation and DNA repair during cell cycle arrest *in vitro* (Young and Hodas, 1964; Krakoff *et al.*, 1968). Pathways involved in the cell cycle were therefore searched, as a means of observing the effect of HU on the cell cycle *in vivo*. In $CD34^+$ cells, *CDKN1A*, *ZNF385A*, *UBE2D1*, *HIST1H4A* and *BACH1* genes were shown to be upregulated. These genes participate in the transcriptional activation of p21 and are activated after DNA damage in G_1/S checkpoints. However, *TP53* which is responsible for DNA repair, was downregulated. *CHECK1* and *CDC25A* genes, which participate in the delay of cell cycle progression in response to double-stranded DNA breaks, were also downregulated. High doses of HU or prolonged exposure, leads to apoptosis *in vitro* (Johnson *et al.*, 1992; Gui *et al.*, 1997). In this case, the caspase activation pathway was found to be enriched in upregulated genes and including *TNFSF10* and *CFLAR*.

In the case of neutrophils, genes involved in cell cycle pathways were downregulated, specifically those involved in the mitotic phase (*SMC1A*, *SMC3 STAG2*,

PDS5B, *NIPBL*). Genes involved in the apoptotic pathway, such as *CASP8*, *BIRC2* and *TNFSF10,* among others were also downregulated.

The effect of HU treatment on the cell cycle was shown to be stronger in $CD34^+$ cells, affecting genes that participate in DNA damage and check points. In the case of neutrophils, it is difficult to assess the effect on the cell cycle since these cells do not replicate.

These results clearly indicated that the expression of key transcription factors for haematopoiesis were greatly affected by HU in $CD34^+$ cells and neutrophils. These TFs may explain the global effect of HU on the immune system, specifically those that were upregulated. In addition to regulation via TFs, DNA methylation was also investigated as a potential mediator of HU effect.

3.2.4 The effect of HU treatment on DNA methylation

DNA methylation was measured in $CD34^+$ cells and neutrophils using the Illumina MethylationEPIC BeadChip (EPIC) (Methods Section 2.5 and Chapter 3 Section 3.2.2.2). Differential methylation analysis was performed at individual CpGs and regions (1 kb) containing CpGs with consistent methylation changes. Since CpGs methylation states, typically, are highly correlated within short genomic distances, the methylation data for individual CpGs were averaged, and differentially methylated regions (DMRs) were identified using the DMRcate package (Peters *et al.*, 2015). DNA methylation level is expressed as β-values, which are computed for each CpG and range between 0 and 1, with 0 being not methylated and 1 fully methylated.

The DMR analysis was mainly focussed on CpG-rich regions - the so-called CpG islands (CGIs) (Fig. 3.6A). CGIs are frequently located at gene promoters, which are commonly unmethylated even in genes that are not expressed (Ioshikhes and Zhang, 2000; Bird, 2002). However, DNA methylation plays an important role during development, by silencing tissue-specific genes (Bird, 2002). This mechanism

also seems to be dependent on the density of CpGs in CGIs, and their locations in the genome (Weber *et al.*, 2007; Illingworth *et al.*, 2008, 2010). DMRs were therefore classified according to their genomic location (promoter, intragenic or intergenic) and association with CGIs (3.6A). These DMRs and associated genes were correlated with the gene expression results obtained in the previous Section 3.2.3.

3.2.4.1 DNA methylation in neutrophils is not significantly affected by HU treatment

In neutrophils, DNA methylation was measured at four time points for eight patients (Table 3.3). Differential methylation testing was performed using paired samples between pre-treatment, and following three-, six-, and nine- months of treatment. Most of the significant changes at CpG level (q-value <0.05) occurred at nine months of treatment (37 CpGs) (Appendix C.1), but no DMRs where identified by any of the time point comparisons. When testing all treated versus all non-treated samples, ten regions were found to be differentially methylated (Table 3.4). However, the average DNA methylation change was lower than 7%, and was not correlated to gene expression, thus these regions were not subjected to further analysis. In conclusion, DNA methylation in neutrophils was highly stable and nine months of HU treatment was not enough time to detect any significant changes.

Table. 3.4. Differentially methylated regions identified in neutrophils comparing pre-treatment vs all treated.

Position of the DMR			Width	Nº CpGs	Mean methylation change	Overlapping promoters
Chr	Start	End				
chr7	84001627	84002356	730	4	-0.07	SEMA3A
chr3	153942348	153942432	85	3	-0.08	ARHGEF26
chr17	79813410	79813507	98	3	0.038	P4HB
chr5	33362853	33363045	193	2	-0.07	-
chr10	134362126	134362170	45	3	0.04	-
chr17	40714979	40715281	303	6	0.03	COASY
chr5	177870177	177870407	231	6	0.03	CTB-26E19.1
chr17	79816504	79817271	768	7	0.04	P4HB
chr14	103415458	103416268	811	5	0.03	-
chr6	31762353	31762723	371	13	0.01	VARS

DMR: differentially methylated region

3.2.4.2 DNA methylation in CD34$^+$ cells is greatly affected by HU after nine months of treatment

Due to the limited amount of DNA obtained from CD34$^+$ cells, samples from three patients were assayed at two times points; pre-treatment and following nine months of treatment (Table 3.3). One sample was removed due to QC issues (Section 3.2.2.2). Differential methylation testing identified 81,024 significant CpG probes (p-value <0.05), and 690 DMRs (q-value <0.15 and containing at least 4 CpGs) (Fig. 3.6A).

DNA methylation at gene promoters has been associated with repression of a small subset of genes during development. However, aberrant hypermethylation in some CGI promoters have shown to be associated with carcinogenesis (Esteller *et al.*, 2000). Accordingly, DMRs that overlapped CGI promoters were selected and correlated with gene expression (Fig. 3.6A). Of these regions (188), 35 were associated to DEGs (Table 3.5 and Fig. 3.6B). Loss of DNA methylation and increased gene expression were present in 22 genes, which included kinases (*FGR*, *CDKN1C*, *MAP4K4*) and transcription factors (*SPI1*, *RARA*, *JMJD1C*). Surprisingly, 13 genes presented with loss of DNA methylation and decreased gene expres-

sion. These included the *HOXB4* and *HOXA5* genes, a finding which is explained by DNA methylation changes concentrated in small areas within the CGI, not necessarily whithin the promoter region (Fig. 3.7). DMRs that did not overlap CGIs (214), and that could be considered as low CpG content promoters (Weber *et al.*, 2007), were also identified. Conversely, 35 DMRs did overlap with DEGs (Table 3.6 and Fig. 3.6C). Among these, 16 showed negative correlation with gene expression, while 19 were positively correlated (Fig. 3.6C). This confirmed previous findings on low CpG content promoters, where a negative correlation with gene expression is not expected, but some active promoters may be unmethylated (Weber *et al.*, 2007).

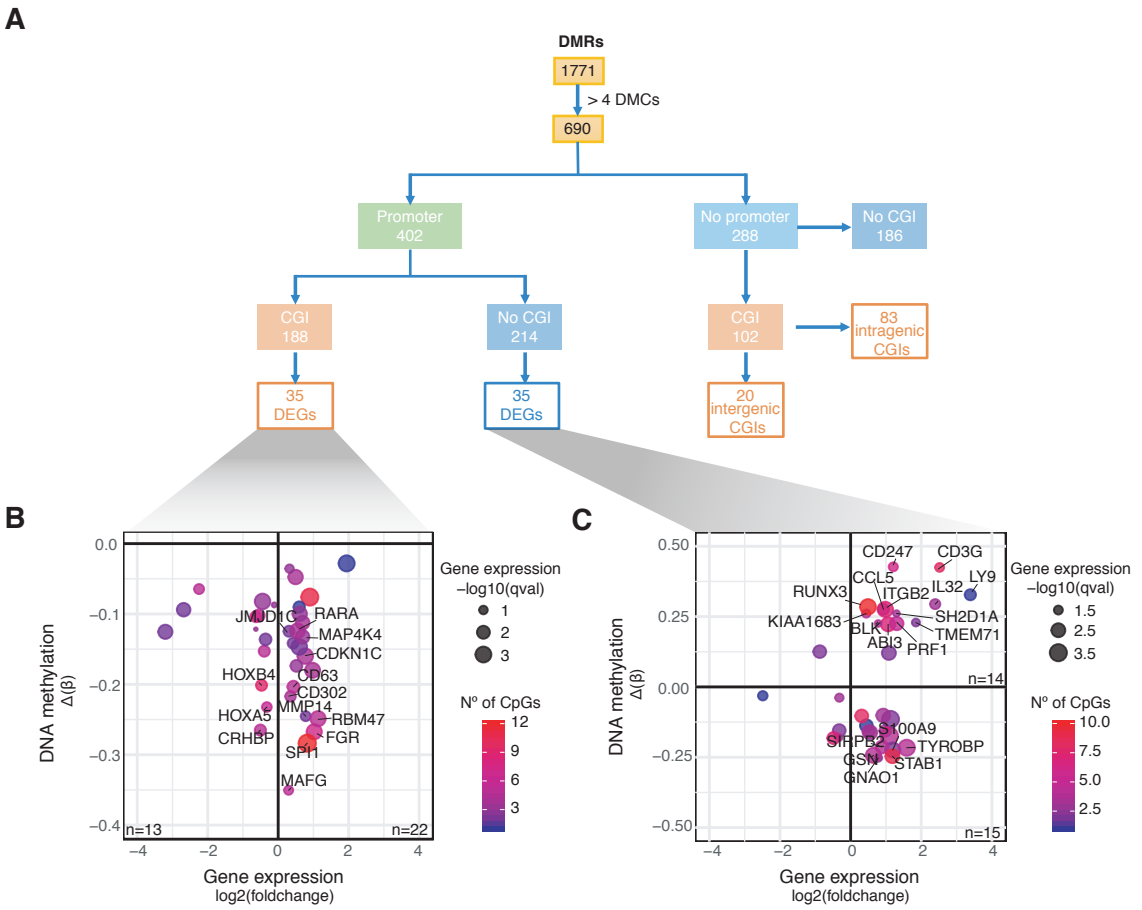

Figure. 3.6. Workflow of the DNA methylation analysis and correlation between differentially methylated regions and gene expression. A) Workflow of the analysis of the differentially methylated regions (DMRs) in $CD34^{+}$ cells. DMRs having more than 4 CpGs in the region where selected and divided between DMRs overlapping promoters and no promoters. DMRs overlapping promoter were separated between overlapping a CpG island (CGI) or not. DMRs not overlapping promoters were further separated between overlapping or not CGI and according to genomic region (intragenic or intergenic). B) DMRs that overlap CGI at gene promoters of differentially expressed genes (DEGs). C) DMRs that overlap gene promoters of DEGs, but where no CGIs have been described in the promoter region. DNA methylation of each region is expressed by the difference of the β-value prior to and following HU treatment. Gene expression change is expressed by the log2 fold-change prior to and following HU treatment (value obtained from Sleuth package). Color scale represents the number of differentially methylated CpGs in the region that were identified by limma package.

Among the DMRs negatively correlated with gene expression, *SPI1* was the most differentially expressed gene and with the highest number of differentially methylated CpG probes (Fig. 3.6B). Further analysis of *SPI1* revealed that its regulatory region, located -17kb away from the promoter (Okuno *et al.*, 2005), was also differen-

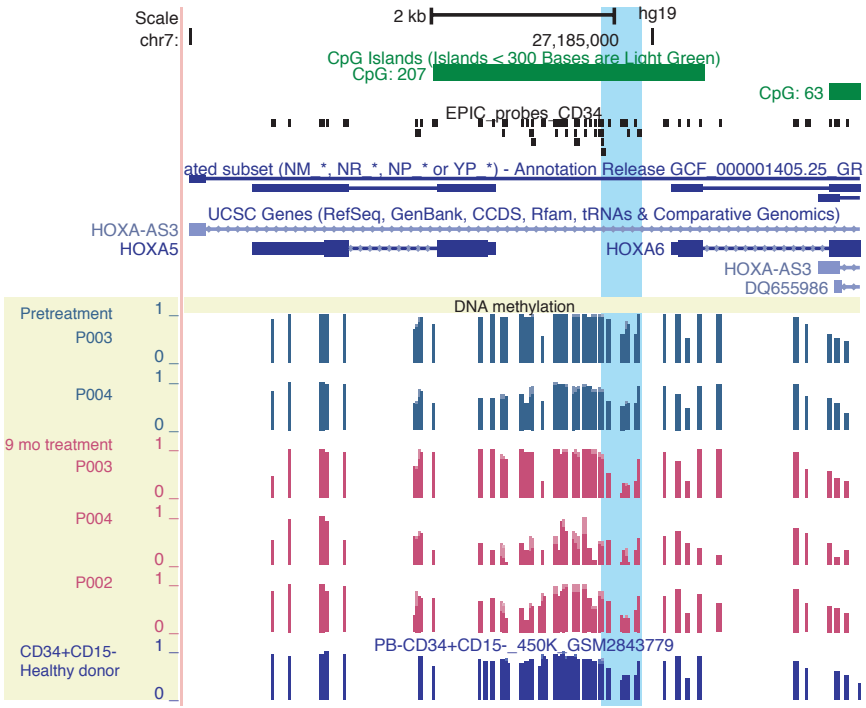

Figure. 3.7. HOXA5 locus. Example of a differentially methylated region that overlap CGI in gene promoters but the region that is differentially methylated (light blue shading) does not overlap the promoter of the differentially expressed gene, *HOXA5*. Yellow shading indicates the DNA methylation levels from $CD34^{+}$ cells from patients at two time points, and one healthy donor.

Table. 3.5. Differentially methylated regions overlapping CpG islands in gene promoters whose genes are differentially expressed

Symbol	Gene expression details			Differentially methylated region (DMR)						
	q-val (transcript)	q-val (gene)	b	Chr	Start	End	Width	Nº CpGs	DNA meth difference	Nº sig CpGs (limma)
CD302	0.043	0.784	0.35	chr2	160653686	160655179	1494	11	-0.217	4
CD63	0.018	0.002	0.434	chr12	56121099	56122120	1022	7	-0.203	5
CDKN1C	0.001	0.001	0.768	chr11	2907670	2908471	802	13	-0.16	4
CLCN5	0.021	0.019	0.518	chrX	49687006	49687331	326	11	-0.174	3
CTIF	0.038	0.027	0.422	chr18	46064542	46065292	751	9	-0.142	2
CTSZ	0	0	0.898	chr20	57582006	57583709	1704	27	-0.076	11
ELF4	0.001	0.004	0.599	chrX	129244725	129244816	92	5	-0.147	2
FGR	0.002	0	1.023	chr1	27960788	27962692	1905	13	-0.267	7
G0S2	0.001	0	1.943	chr1	209848479	209849006	528	8	-0.028	1
GK	0.028	0.037	0.597	chrX	30670939	30671488	550	12	-0.09	1
GPSM3	0.003	0	0.494	chr6	32164124	32165200	1077	23	-0.047	4
IFNGR2	0.006	0.001	0.693	chr21	34774627	34775459	833	8	-0.112	4
JMJD1C	0.034	0.166	0.308	chr10	65225648	65226682	1035	9	-0.125	2
KCNE3	0.004	0.002	0.612	chr11	74178114	74179208	1095	11	-0.1	3
MAFG	0.084	0.019	0.297	chr17	79880647	79882042	1396	9	-0.35	5
MAP4K4	0.006	0.001	0.69	chr2	102313069	102313888	820	6	-0.132	3
MMP14	0.056	0.041	0.781	chr14	23305153	23305957	805	9	-0.245	2
RARA	0.002	0	0.548	chr17	38498077	38499096	1020	10	-0.124	4
RBM47	0.002	0	1.131	chr4	40632362	40633572	1211	10	-0.249	5
SORT1	0.002	0.203	0.976	chr1	109940754	109941201	448	7	-0.18	4
SPI1	0	0	0.825	chr11	47398575	47401027	2453	15	-0.284	12
ZNF641	0.106	0.029	0.323	chr12	48744457	48745694	1238	13	-0.036	3
CD34	0.001	0	-0.444	chr1	208084071	208085889	1819	18	-0.082	3
CRHBP	0.03	0.005	-0.505	chr5	76248637	76249502	866	7	-0.265	5
DAPK1	0.055	0.033	-2.245	chr9	90112086	90113998	1913	19	-0.065	5
HOXA5	0.063	0.046	-0.322	chr7	27183591	27185393	1803	34	-0.232	6
HOXB4	0.033	0.023	-0.472	chr17	46656777	46660097	3321	19	-0.201	9
ITGA6	0.004	1	-3.202	chr2	173292978	173294093	1116	9	-0.125	2
MECP2	0.234	0.035	-0.111	chrX	153362589	153363472	884	6	-0.087	3
MORC2	0.017	1	-0.558	chr22	31317914	31318373	460	8	-0.102	5
MSRB3	0.253	0.032	-0.633	chr12	65671924	65672696	773	9	-0.122	3
SLC17A9	0.032	0.029	-0.395	chr20	61583686	61584850	1165	11	-0.153	5
TRIM9	0.006	0	-2.683	chr14	51562381	51563131	751	12	-0.094	2
UCK2	0.022	0.009	-0.358	chr1	165796134	165796797	664	7	-0.136	2
ZNF577	0.033	0.035	-0.632	chr19	52390810	52391789	980	14	-0.104	7

b: log(foldchange) computed by Sleuth package

Table. 3.6. Differentially methylated regions in gene promoters without CpG islands whose genes are differentially expressed

Symbol	Gene expression details			Differentially methylated region (DMR)						
	q-val (transcript)	q-val (gene)	b	Chr	Start	End	Width	Nº CpGs	DNA meth difference	Nº sig CpGs (limma)
AIF1	0	0	1.09	chr6	31582837	31584223	1387	6	-0.174	4
FCER1G	0.01	0	0.982	chr1	161184489	161185092	604	7	-0.156	4
GNAO1	0.048	0.084	0.785	chr16	56228511	56229180	670	5	-0.253	2
GSN	0.001	0.001	0.646	chr9	124048188	124048515	328	5	-0.243	4
ITGAM	0	0	1.129	chr16	31270808	31271240	433	5	-0.115	2
KDM2B	0.006	0.001	0.447	chr12	122018562	122021013	2452	19	-0.137	1
LST1	0.008	0	0.511	chr6	31554199	31555671	1473	10	-0.158	4
NFAM1	0.003	0.001	0.923	chr22	42828125	42828946	822	8	-0.101	3
RARA	0.002	0	0.548	chr17	38464679	38465510	832	11	-0.161	3
S100A9	0.001	0	1.192	chr1	153330310	153330776	467	5	-0.224	2
SIRPB2	0.005	0.003	0.888	chr20	1471884	1472419	536	7	-0.214	4
STAB1	0.003	0.001	1.184	chr3	52528714	52529524	811	10	-0.247	9
TNFSF13	0.073	0.015	1.053	chr17	7460690	7462249	1560	14	-0.149	7
TYROBP	0.001	0	1.587	chr19	36399185	36400049	865	7	-0.216	4
WDFY4	0.005	0.003	0.314	chr10	49892741	49893549	809	15	-0.103	8
SNORD33	0.006	0.003	-0.874	chr19	49992293	49994372	2080	19	0.125	2
CBFA2T3	0.006	0.002	-0.322	chr16	89042948	89043707	760	10	-0.155	2
IGLL1	0.052	0.03	-0.333	chr22	23922551	23923610	1060	6	-0.18	3
INPP5B	0.026	0.005	-2.486	chr1	38411997	38413334	1338	10	-0.031	1
MEIS1	0.038	0.39	-0.318	chr2	66664345	66667101	2757	16	-0.038	3
ROBO4	0.008	0.027	-0.484	chr11	124766927	124768950	2024	16	-0.182	7
ABI3	0.002	0.002	1.061	chr17	47287223	47289036	1814	15	0.223	7
BLK	0.053	0.015	0.78	chr8	11350853	11351846	994	8	0.226	4
CCL5	0.013	0.008	0.931	chr17	34207332	34207663	332	5	0.276	3
CD247	0.03	0.003	1.207	chr1	167486600	167487871	1272	8	0.427	6
CD3G	0.031	0.058	2.52	chr11	118213272	118215112	1841	11	0.425	8
IL32	0.02	0.197	2.394	chr16	3115133	3115809	677	9	0.294	3
ITGB2	0.001	0	0.981	chr21	46340351	46341918	1568	14	0.276	7
KIAA1683	0.038	0.002	0.443	chr19	18385244	18385672	429	5	0.261	4
LY9	0.013	0.24	3.389	chr1	160765225	160766535	1311	6	0.328	1
PRF1	0.004	0.002	1.306	chr10	72362292	72363272	981	11	0.225	5
RUNX3	0.001	0	0.492	chr1	25290947	25292412	1466	17	0.285	10
S100A8	0.003	0.001	1.083	chr1	153363264	153364020	757	7	0.121	2
SH2D1A	0.057	0.042	1.298	chrX	123479823	123480594	772	5	0.261	3
TMEM71	0.042	0.12	1.847	chr8	133772657	133773484	828	6	0.229	2

b: log(foldchange) computed by Sleuth package

tially methylated (Fig. 3.8). Comparison with healthy individual samples revealed that these regions are aberrantly methylated in MPN patients (Fig. 3.8). Interestingly, HU was able to rescue the DNA methylation state to a similar level of a healthy individual (Fig. 3.8). These observations suggest that there is an underlying mechanism where HU is able to modulate DNA methylation and affect gene expression or vice versa.

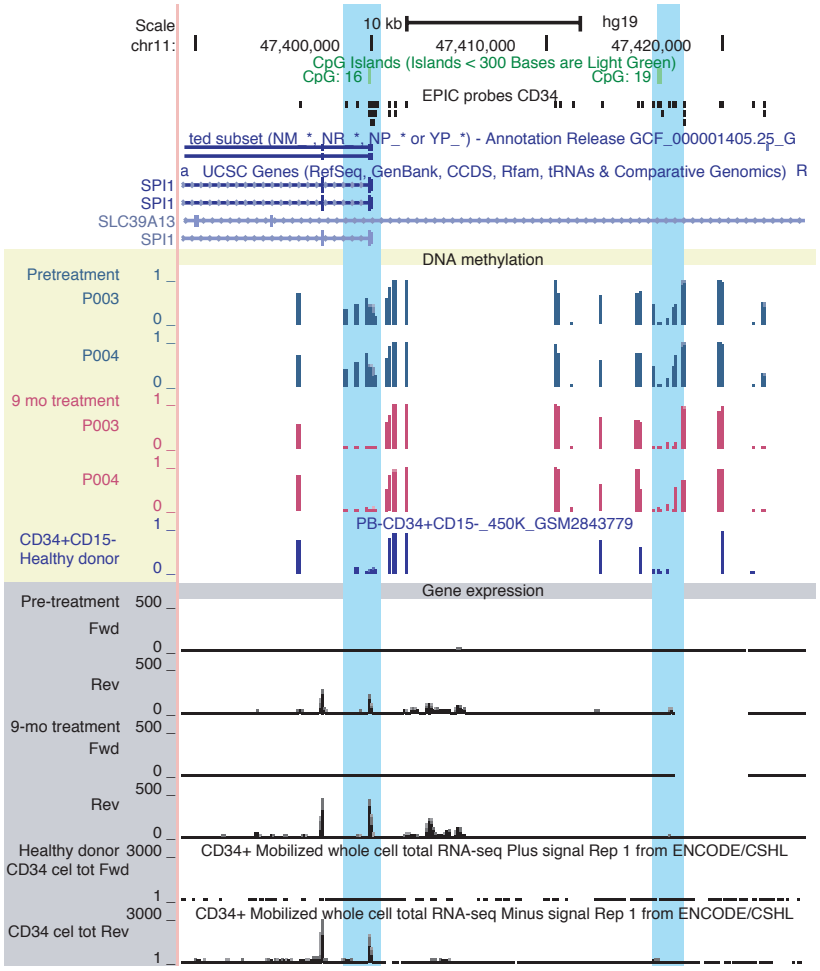

Figure. 3.8. Locus of the promoter and regulatory region of SPI1 with DNA methylation and gene expression levels before and after nine months of HU treatment. Yellow shading indicates the DNA methylation levels from $CD34^+$ cells from two patients at two time points, and one healthy donor. Blue shading indicates significantly differentially methylated regions (DMRs). Grey shading denotes gene expression levels in $CD34^+$ cells from one patient at two time points, and one healthy donor.

3.2.4.3 Analysis of intergenic and intragenic regions

A total of 288 regions that did not overlap promoters, such as intergenic and intragenic CGIs, were also analyzed.

Methylation of intergenic CGI regions Intergenic CGIs can represent regulatory regions of genes located at their proximities (Illingworth and Bird, 2009). Of the 288 regions mentioned above, 20 were shown to be intergenic CGIs (Table 3.8), of which three were hypermethylated, and 17 hypomethylated, following treatment. These 20 regions were screened using the online tool i-cisTarget (Imrichová *et al.*, 2015), to determine whether they were associated with regulatory features. This tool also generates a list of genes associated with the regions screened. The results of this analysis showed that the 17 hypomethylated regions were enriched for histone 3 lysine 27 tri-methylation (H3K27me3) in different tissues (Table 3.7). The H3K27me3 has been described as a repressor mark, since it is associated with regions of the polycomb repressive complex 2 (PRC2). Moreover, H3K27me3 co-locates with CGIs which are mostly hypomethylated (Rose and Klose, 2014). To evaluate the relevance of the enrichment analysis, DEGs were superimposed on the lists of genes associated with the regions screened with i-cisTarget, and it was observed that none of the listed genes were differentially expressed (Table 3.8). Therefore, it is possible that the 17 hypomethylated intergenic CGIs identified have no relevance to the effects of HU treatment. However, when three of the intergenic hypermethylated CGIs identified were analyzed in the same way (Table 3.8), it was observed that the transcription factor RUNX3 (also differentially methylated and expressed), bound to these regions. Moreover, one of these regions was associated with the *IRF8* gene, which was upregulated and differentially expressed (Table 3.8). RUNX3 plays a role in haematopoiesis since its deletion leads to a myeloproliferative phenotype (Wang *et al.*, 2013). In the case of *IRF8*, this gene is expressed mainly in haematopoietic cells, and its deletion has shown to perturb HSCs self-renewal of mice (Qiu *et al.*,

2015). In conclusion, it appears, the effect of HU at intergenic CGIs is limited, with few regions affected. However, one hypermethylated region bound by RUNX3 could be important for the regulation of the *IRF8* gene.

Table. 3.7. Top enrichment results from i-cisTarget of intergenic and intragenic CpG islands that are differentially methylated regions. Genes associated with the feature identified by i-cisTarget were overlapped with the list of differentially expressed genes (DEGs). The number of associated genes that were differentially expressed is indicated.

Type of DMR	Feature	Nº of genes associated	DEGs		Dataset used by i-cisTarget
			up	down	
Hypo intergenic	H3K27me3	6	0	0	H3K27me3 in Primary mononuclear cells from peripheral blood (E062,)
Hyper intergenic	**RUNX3**	3	1	0	RUNX3 ChIP-seq protocol v042211.1 on human GM12878
Hypo intragenic	**ZBTB7A**	23	6	6	ZBTB7A ChIP-seq protocol v042211.1 on human ECC-1
	H3K27me3	24	1	4	H3K27me3 ChIP-seq on human H7-hESC differentiated 14 days
Hyper intragenic	Possible TFs: **TFE3**	2	0	0	transfac_pro__M06957 Description: V$TFEA_01: TFEA
	Possible TFs: **FOS**, **FOSB**, **JUNB**, **JUN**, **JUND**, FOSL1, **FOSL2**	2	0	0	stark__TGANTCA Description: activating-protein 1 (FOS-JUN heterodimer)

Bold indicates differentially expressed genes

Table. 3.8. Differentially methylated intergenic CpG islands and associated genes identified by i-cisTarget.

Position of the DMR			Width	Nº	FDR	Mean methylation	Gene associated
Chr	Start	End	(bp)	CpGs		change	(i-cisTarget)
chr4	174428790	174430614	1825	12	1.15E-41	-0.072	*HAND2*
chr5	139088316	139089549	1234	7	5.68E-38	-0.203	
chr5	43000303	43001210	908	8	1.14E-33	-0.282	
chr2	241562085	241562650	566	5	9.23E-33	-0.218	
chr6	156717406	156718546	1141	6	3.94E-31	0.283	*ARID1B*
chr2	85640762	85641985	1224	13	4.21E-31	-0.158	
chr21	45575014	45576085	1072	5	2.63E-30	0.443	*C21orf33*
chr16	85981336	85981947	612	5	4.49E-26	0.387	***IRF8***
chr11	32008659	32009163	505	5	5.34E-24	-0.081	*RCN1*
chr6	27647843	27648605	763	7	2.64E-23	-0.059	*HIST1H2BL*
chr19	57149436	57149813	378	7	1.03E-22	-0.028	
chr22	36806001	36806655	655	6	2.80E-22	-0.197	
chr11	47416109	47417205	1097	6	4.26E-21	-0.126	
chr1	38941882	38942644	763	8	8.53E-20	-0.086	
chr4	85403167	85403915	749	5	4.17E-19	-0.066	
chr5	77253544	77253990	447	5	7.31E-19	-0.142	*TBCA*
chr2	223185049	223185848	800	11	1.10E-18	-0.125	
chr6	29521499	29521803	305	16	1.59E-18	-0.032	*UBD*
chr8	8820844	8821258	415	5	3.81E-18	-0.129	
chr11	20618001	20618651	651	6	4.39E-16	-0.028	*SLC6A5*

Bold indicates differentially expressed genes

Methylation of individual CpGs at intergenic regions In addition to the DMR analysis, single CpG analysis was carried out since some TFs can be affected by DNA methylation at one CpG in their binding site. An example of this is CEBPB, which is involved in inflammation and preferably binds to methylated CpGs (Rahman *et al.*, 2012; Mann *et al.*, 2013). Therefore, all the differentially methylated CpGs (± 5 bp) at intergenic regions were selected and subjected to i-cisTarget screening to identify enrichment for the binding of certain TFs. The CpGs were divided into "open sea" CpGs, which are isolated from CGIs (Sandoval *et al.*, 2011), and in CpGs overlapping CGIs. To identify the relevance of the enriched TFs, the genes associated with the TF binding provided by i-cisTarget were overlapped with DEGs (Table 3.9). "Open sea" CpGs were enriched for TFs playing a role in haematopoietic differentiation, such as *GATA1*, *GATA2* and *SPI1*. Interestingly, the TFs associated with the highest number of DEGs were also differentially expressed (Table 3.9, shown in bold). Finally, CpGs overlapping CGIs were enriched

for two members of the PRC2 (SUZ12 and EZH2) and for CTBP2, which was differentially expressed. In conclusion, DNA methylation analysis at intergenic CpGs identified distal regulatory regions that explain the involvement of key TFs in the regulation of gene expression following HU treatment. In addition, this analysis identified CEBPB, whose binding is known to be affected by DNA methylation, thus confirming that the regions identified here could be relevant to the regulation of gene expression.

Table. 3.9. Enrichment results of transcription factors that bind differentially methylated CpGs located in intergenic regions.

Type of CpG	T. Factor	Nº of genes associated	DEGs		Dataset used by i-cisTarget
			up	down	
Open sea hypo (7,791 CpGs)	GATA2	233	33	15	GATA2 ChIP-seq protocol PCR1x on human K562
	GATA1	133	16	4	GATA1 ChIP-seq on human PBDE Fetal
	SPI1	439	92	13	SPI1 ChIP-seq protocol PCR1x on human K562
Open sea hyper (8,034 CpGs)	**CEBPB**	165	43	5	CEBPB ChIP-seq protocol v042211.1 on human GM12878
	TCF12	490	126	21	TCF12 ChIP-seq protocol PCR1x on human GM12878
	BCL11A	463	110	20	BCL11A ChIP-seq protocol PCR1x on human GM12878
CGI hypo (3,049 CpGs)	SUZ12	168	10	4	SUZ12 ChIP-seq on human NT2-D1
	EZH2	305	15	6	ChIP-seq on human H1-hESC EZH2
	CTBP2	479	49	9	CTBP2 ChIP-seq on human H1-hESC
CGI hyper (1,974 CpGs)	SUZ12	210	1	0	SUZ12 ChIP-seq on human NT2-D1
	EZH2	178	9	0	ChIP-seq on human H1-hESC EZH2
	CTBP2	210	20	3	CTBP2 ChIP-seq on human H1-hESC

CGI: CpG island; T. Factor: transcription factor; hyper: hypermethylated; hypo: hypomethylated; DEGs: differentially expressed gene; bold indicates differentially expressed genes

Methylation at intragenic CGI Analysis of intragenic CGIs was also carried out. These regions are more prone to bmethylation than promoter CGIs, where methylation is usually associated with transcription from the host gene. However, intragenic CGIs can exhibit transcriptional activity as in the case of imprinted

genes (Edwards and Ferguson-Smith, 2007). Moreover, methylation in these regions has also been associated with disease (Tufarelli *et al.*, 2003). In this study, among the intragenic CGIs identified, 13 were hypermethylated and 70 hypomethylated. These were analyzed using i-cisTarget, to identify potential regulatory features. Hypomethylated regions were enriched for the transcription factor ZBTB7A, which is differentially expressed and upregulated. Among the genes associated with ZBTB7A, 12 were also differentially expressed (Table 3.7). The binding sites of ZBTB7A from this analysis were added to the UCSC genome browser, to check whether these were properly annotated as intragenic CGIs. Of the 12 genes analyzed, only one was a purely intragenic CGI (*F2RL1*, Fig. 3.9); the remainder were CGIs in alternative promoters, that overlapped introns in the longest isoform. In one case, *F2RL1* was downregulated, and the intragenic CGI lost methylation following treatment (Fig. 3.9). Another regulatory feature enriched was H3K27me3, which was associated with five DEGs. *F2RL1* intragenic CGI was also identified in these five DEGs (Fig. 3.9). Another gene, HOXB3, was shown to be downregulated and hypomethylated after treatment. These observations confirm what has already been described; Intragenic CGIs are usually methylated when the gene is transcribe, following expression of the host gene where the CGI lies. In this case, hypomethylation of the intragenic CGI follows the downregulation of the gene within which it lies. On the other hand, hypermethylated intragenic CGIs were enriched for one potential transcription factor, TFE3: this was differentially expressed, but its associated genes were not (Table 3.7). Similarly, other possible transcription factors were the members of the Fos family: most of these were differentially expressed, but the associated genes did not lead to any DEGs (Table 3.7 in bold). In conclusion, HU treatment affects binding sites for ZBTB7A, which may act as a negative regulator of *F2RL1* expression.

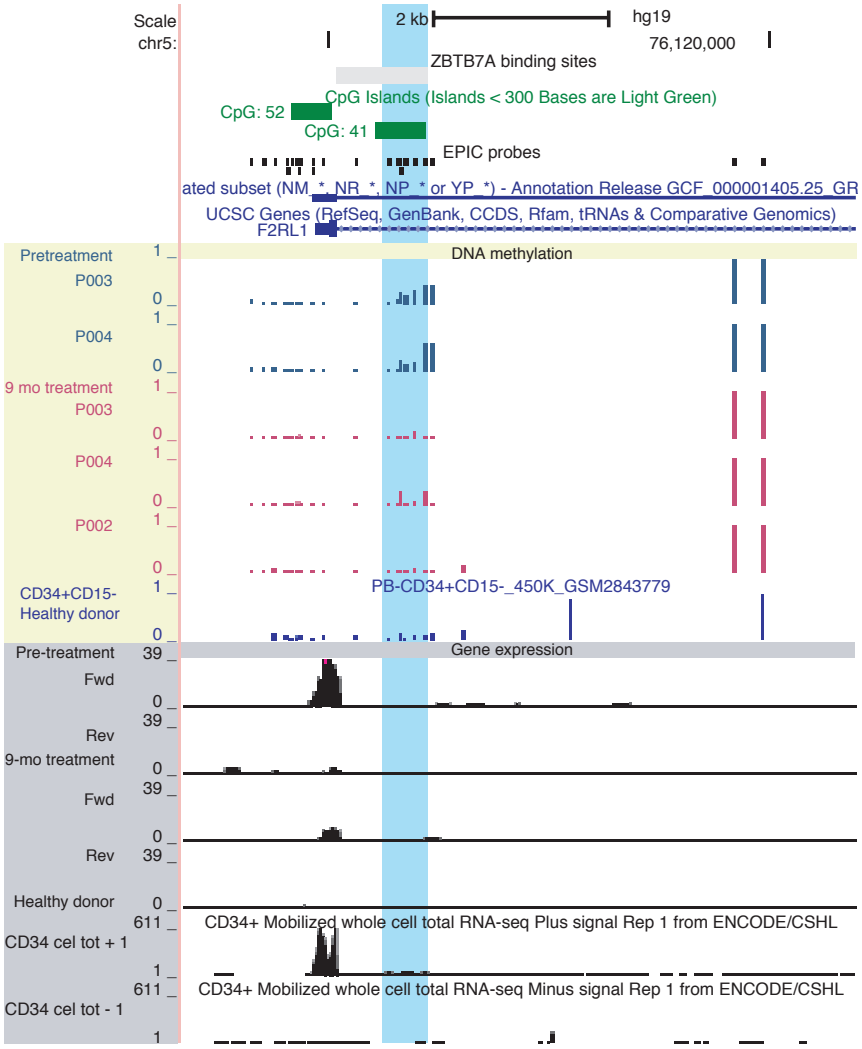

Figure. 3.9. Intragenic CpG island of the *F2RL1* gene. Intragenic CpG islands (CGI) were identified and compared with gene expression of the host gene. Among the genes identified, the *F2RL1* gene showed hypomethylation of its intragenic CGI following treatment and was also downregulated. Yellow shading indicates DNA methylation levels from patients before and after HU treatment. Light blue shading indicates the differentially methylated CGI. Grey shading indicates gene expression data from one patient before and after HU treatment.

3.3 Conclusions

Gene expression and DNA methylation were assayed in samples prior to and following HU treatment in two human cell types: $CD34^+$ cells and neutrophils. Gene expression analysis revealed that the DEGs which responded to HU treatment were mainly involved in immune system functions, and that these were regulated by key transcription factors for haematopoiesis.

In $CD34^+$cells, the differentially expressed TFs, SPI1 (or PU.1) and RUNX1 were found to both upregulate and downregulate target genes. These TFs are important regulators of haematopoiesis, specifically during myeloid differentiation. Moreover, in the case of neutrophils, SPI1 and CEBPD were also identified as potential regulators of DEGs and both were differentially expressed. Interestingly, SPI1 is crucial for granulopoiesis and interacts with CEBPD. These observations highlight SPI1, RUNX1 and CEBPD as potential candidate genes to understand HU mechanism of action.

DNA methylation in neutrophils was highly stable and HU treatment did not produced any significant changes during any of the time-points investigated. However, the effect of HU treatment in $CD34^+$ was more pronounced. Interestingly, HU was able to affect DNA methylation at the *SPI1* promoter and its regulatory region. Moreover, analysis in intergenic and intragenic CpG islands also led to relevant transcription factors such as ZBTB7A, RUNX3 and CEBPB, which were differentially expressed and were identified as regulators of DEGs. Although, not all the changes in gene expression were explained by DNA methylation, this study identified relevant transcription factors that regulated a large proportion of the DEGs. It is also probable that in neutrophils other mechanisms are involved in the regulation of gene expression.

Chapter 4

The effect of hydroxyurea treatment on gene expression in the Jak2V617F - knock - in mouse

4.1 Introduction

Mouse studies have several advantages over human studies; for example, they permit the evaluation of controlled drug intake, diet and environment against a single genetic background. Mouse studies also permit the study of a greater number of subjects, and across a wider range of samples. Overall, these features result in less inter-individual variation, and increase the variety of analyses that can be performed.

This chapter presents a comparative study across species, mirroring the previous clinical study in a mouse model of myeloproliferative neoplasm (MPN). Comparative studies, by observing similarities and differences between two different species, provide additional information about features that are relevant to the research question being investigated. This comparative study was designed to confirm my previous findings in MPN patients, and to identify stronger candidate genes, that were conserved in the two species, for further validation. Overall, this analysis led to the

identification of candidate genes which explain some of the clinical outcomes following HU treatment, which may be informative in discerning potential therapies.

The study utilised a previously described heterozygous Jak2V617F-knock-in (Jak2VF) mouse resembling the MPN phenotype (Chen and Mullally, 2014). Briefly, these mice were generated by crossing floxed heterozygous Jak2V617F mice (Mullally *et al.*, 2010) with *vav*Cre$^+$ transgenic mice (Georgiades *et al.*, 2002), in order to switch expression from Jak2 to the mutant form Jak2V617F in the haematopoietic compartment alone. The generated Jak2VF mouse model shows similar clinical features to human MPN, including extra-medullary haematopoiesis that results in splenomegaly (Chen and Mullally, 2014).

The Jak2VF mice were treated with HU or vehicle (NaCl) for a total of six weeks. Wild-type C57Bl/6 and *vav*Cre mice were used as controls, and were also treated with HU or vehicle. As for the human study, two cell types were isolated. LSK cells, which represent the stem cell population, were isolated from bone marrow, and their gene expression compared to that of human CD34$^+$ cells. Neutrophils were isolated from blood samples drawn from the retro-orbital sinus, and their gene expression was compared to that in their counterparts isolated from human blood.

Unfortunately, in this mouse study, DNA methylation could not be assayed due to time constraints and the low concentrations of DNA isolated, particularly from neutrophil samples. DNA methylation analysis would have required time-consuming optimization of the experimental protocols, which were all designed for higher concentrations of DNA (>100 ng). Accordingly, this mouse study focused on the gene expression analysis of LSK cells and neutrophils.

Similarly to the previous chapter, RNA libraries were prepared and sequenced and the same pipelines were used for data generation. This chapter presents a brief report of these analyses. It also provides a detailed analysis of all the different groups of mice (mutant, wild-type, treated, non-treated) and finally, a comparative analysis of the two species.

4.2 Results

4.2.1 HU treatment in Jak2VF and wild-type mice

This study was funded by a Junior Collaboration Award from the European Hematology Association, which covered the travel and experimental costs of a three-month attachment in the laboratory of Dr Ann Mullally (Department of Hematology, Harvard Medical School, Boston, USA). During this time I applied HU treatment to a Jak2VF mouse model of MPN previously generated by Dr Mullaly's laboratory (Chen and Mullally, 2014).

To minimize factors which might lead to data variation, the mice were sex- and age-matched wherever possible. However, this largely dependeded on the availability of mice, and the time requiered to conduct the study (Table 4.1). A total of 27 mice were treated, of which 23 were included in the final data analysis. Three mice died before the treatment protocol was completed, and one did not develop the expected Jak2VF mutant phenotype. The mice included in the study were distributed as shown in Table 4.2. Wild-type C57Bl/6 and *vav*Cre^{+} transgenic mice were used as controls for the mutant genotype; both groups are referred as "WT" in the text, without any distinction between them.

Table. 4.1. Mice treated in the study, matching group, phenotypic details and exclusion details.

Mouse ID	Match	Genotype	Sex	Treatment	Age at starting treatment	Included	Early death
RB_431	A	Jak2VF	F	HU	20.714	YES	NO
MM_5322	B	Jak2VF	F	HU	6.571	YES	NO
RB_461	C	Jak2VF	M	HU	6.571	YES	NO
RB_490	B	Jak2VF	F	HU	6.429	YES	NO
RB_445	D	Jak2VF	M	HU	11.857	NO	YES
RB_423	A	Jak2VF	F	HU	17.286	NO	YES
RB_444	D	WT	M	HU	10.857	YES	NO
MM_5276	B	WTcre	F	HU	6.571	YES	NO
MM_5277	B	WTcre	F	HU	6.571	YES	NO
RB_429	A	WTcre	F	HU	20.714	YES	NO
RB_447	D	WTcre	M	HU	10.857	YES	NO
RB_458	C	WTcre	M	HU	6.571	YES	NO
RB_489	B	WTcre	F	HU	6.429	YES	NO
SX_581		WTcre	M	HU	17	YES	NO
RB_430	A	Jak2VF	F	VEH	20.714	YES	NO
RB_462	C	Jak2VF	M	VEH	6.571	YES	NO
RB_491	F	Jak2VF	F	VEH	10.857	YES	NO
RB_481	B	Jak2VF	F	VEH	5.571	YES	NO
MM_5396		Jak2VF	M	VEH	6.571	NO	NO
RB_438	F	WTcre	F	VEH	12	YES	NO
RB_460	C	WTcre	M	VEH	6.571	YES	NO
RB_441	D	WT	M	VEH	12	YES	NO
RB_427	A	WT	F	VEH	20.714	YES	NO
MM_4964	D	WTcre	M	VEH	10.857	YES	NO
MM_5327	C	WTcre	M	VEH	6.571	YES	NO
RB_470	B	WTcre	F	VEH	5.429	YES	YES

Jak2VF: Jak2V617F-knock-in mouse; F: female, M: male; HU: Hydroxyurea; VEH: vehicle

Table. 4.2. Mice included in the study by genotype and type of treatment received.

Treatment	WT	Jak2VF-knock-in
	wild-type or *vav*Cre^{+}	
Vehicle (NaCl)	2/6	4
HU	1/6	4
Total	3/12	8

The HU dose selection was based on previous studies (Kubovcakova *et al.*, 2013). The starting dose was set at 50 mg/kg/day (equivalent to 4 mg/kg/day in human) (Reagan-Shaw *et al.*, 2008), delivered in the form of intraperitoneal injections, 5 days a week with weekly monitoring of the CBC parameters using blood samples drawn from the retro-orbital sinus (Fig. 4.1A). After two weeks of treatment, no

noticeable changes in the haematocrit were observed in WT or Jak2VF mutant mice, and therefore the dose was increased to 100 mg/kg/day for the following four weeks of treatment. A reduction in the white blood cells (WBC) and platelet counts, and the haematocrit, was observed after one week of the higher HU dose (Fig. 4.1A). Interestingly, reduction of platelet counts in Jak2VF mutant mice demonstrated a rapid reduction in platelet count compared to the other parameters, and also compared to WT mice. Body weight was monitored at the beginning of every weekly treatment period to determine whether toxicity was present. Toxicity is denoted by a gradual weight loss over time, which was not the case in these mice (Fig. 4.1A). The body weight was also used for dose adjustment (Fig. 4.1A).

Like MPN patients, Jak2VF mutant mice develop splenomegaly. Accordingly, following completion of the treatment program, the spleens were isolated in order to evaluate the effects of HU treatment. A significant decrease in spleen size was observed in treated mice, compared to controls (Fig. 4.1B, C).

Finally, bone marrow was isolated and enriched for LSK cells as described in Section 2.2.3.1. Blood from the retro-orbital sinus was collected and neutrophils isolated as described in Section 2.2.3.2.

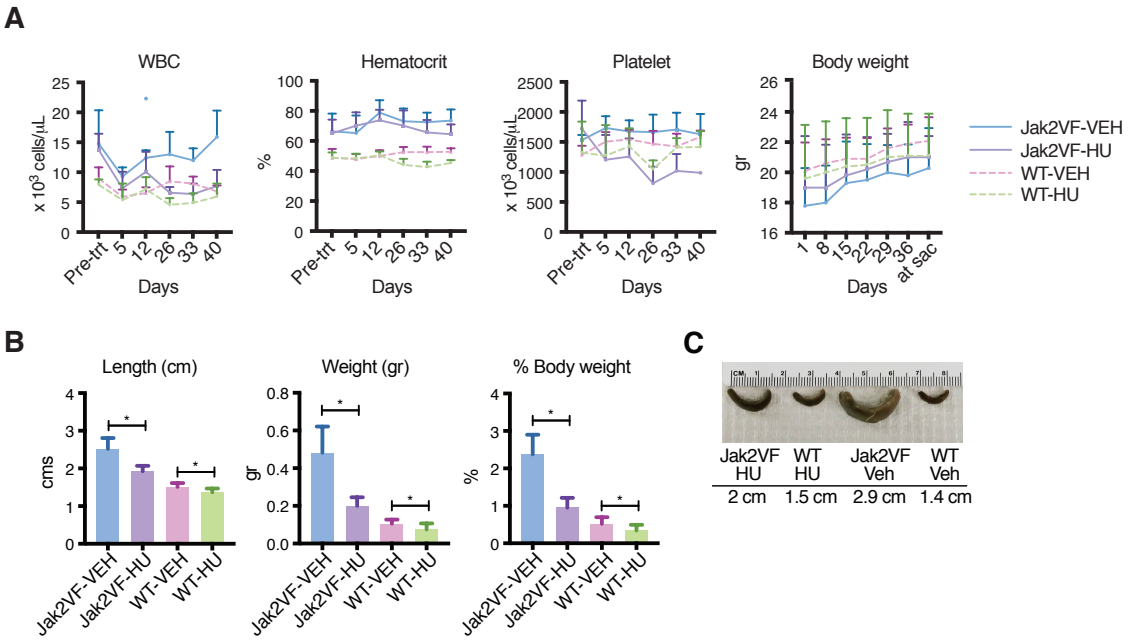

Figure. 4.1. Treatment responses in Jak2VF-knock-in and wild-type mice. A) CBC, including white blood cells (WBC), platelets and haematocrit, were measured on the last day of each weekly treatment period. The mean of each cell count, and mean weight, were calculated per mouse group. B) Spleen weight and length following six weeks of hydroxyurea (HU) or vehicle (VEH) treatment. Mann Whitney unpaired test non-parametric. Error bars indicate mean with standard deviation. C) Spleens were dissected after six weeks of treatment and conserved in 10% formaldehyde. Length given corresponds to fresh tissue.

4.2.2 RNA-sequencing processing

LSK cells and neutrophils were isolated and RNA was purified (Section 2.3). RNA-sequencing was performed in both cell types to identify the effects of HU treatment on gene expression. Details of the library preparation steps, and QC results, are described in detail in Methods Section 2.4.

To avoid high levels of duplication, as was observed in the human neutrophil samples, a low sequencing depth (15 million reads per library) was used for the mouse neutrophil samples, also low in RNA (<5 ng). Alternatively, LSK libraries were sequenced at a higher depth (30 million reads per library) since more RNA was available.

The raw reads were processed as described in Method Section 2.4.6. QC was

performed with FastQC (Appendix D.2). Reads were aligned using STAR (mm10) (Table 4.3) and "RNAseqMetrics" was used to assess others QC parameters (Appendix D.3). In terms of duplication and rRNA percentages, libraries exhibited the expected values and no samples were removed at this stage (Fig. 4.2A and B).

Table. 4.3. Percentages of uniquely mapped reads of LSK and neutrophil libraries. Raw reads were processed and aligned to the reference genome (mm10) using the STAR package.

ID	Uniquely mapped reads	ID	Uniquely mapped reads
MM-4964-LSK	70.27%	MM-4964-NEU	73.30%
MM-5276-LSK	71.23%	MM-5276-NEU	66.81%
MM-5277-LSK	65.51%	MM-5277-NEU	74.72%
MM-5322-LSK	64.15%	MM-5322-NEU	69.11%
MM-5327-LSK	70.09%	MM-5327-NEU	71.15%
RB-427-LSK	68.74%	RB-427-NEU	74.75%
RB-429-LSK	65.28%	RB-429-NEU	52.48%
RB-430-LSK	67.92%	RB-430-NEU	65.24%
RB-431-LSK	65.20%	RB-431-NEU	69.75%
RB-438-LSK	70.92%	RB-438-NEU	64.54%
RB-441-LSK	68.04%	RB-441-NEU	66.87%
RB-444-LSK	66.05%	RB-444-NEU	69.91%
RB-447-LSK	73.94%	RB-447-NEU	71.97%
RB-458-LSK	70.00%	RB-458-NEU	73.58%
RB-460-LSK	69.80%	RB-460-NEU	72.98%
RB-461-LSK	65.70%	RB-461-NEU	69.99%
RB-462-LSK	67.86%	RB-462-NEU	70.26%
RB-470-LSK	69.48%	RB-470-NEU	21.77%
RB-481-LSK	64.43%	RB-481-NEU	68.25%
RB-489-LSK	64.74%	RB-489-NEU	74.45%
RB-490-LSK	75.49%	RB-490-NEU	68.91%
RB-491-LSK	69.87%	RB-491-NEU	79.30%
SX-581-LSK	62.58%	SX-581-NEU	73.14%

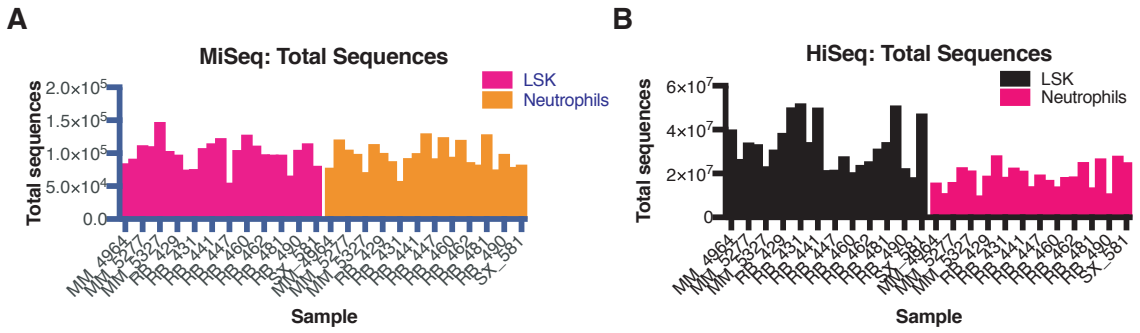

Figure. 4.2. Sequencing QC. A) FastQC results of duplication percentage from HiSeq 4000 sequencing runs. All duplication percentages fell below the expected ranges. B) RNAseqMetrics results of ribosomal RNA (rRNA) percentage from HiSeq 4000 sequencing runs. All percentage levels were shown to be low. In light of these parameters, no samples were removed at this point

Quantification and transcript analysis were performed using Kallisto and Sleuth, respectively. It is expected that mice treated under the same conditions will exhibit more homogenous gene expression levels, since less inter-variability exists. However, it is good practice to determine whether samples are outliers, and whether known factors are driving variance in the data. For this purpose, principal component analysis (PCA) was used. PCAs were plotted using treatment (HU or vehicle) and genotype (Jak2VF mutant and WT) as known factors which could drive variance.

LSK samples in PC2 showed (Fig. 4.3A, y-axis) clustering according to the treatment. However, in PC1 (x-axis) two samples were outliers (RB-447 and RB-490) from the two main clusters, and two samples were clustered with the HU samples despite receiving vehicle treatment (MM-5327 and RB-430) (Fig. 4.3A). In the case of neutrophils, a single cluster with six samples could be observed, separated from the main cluster, and these six samples were considered outliers (Fig. 4.3A). All outliers from both LSK cells and neutrophils were removed from the analysis, and the PCA was re-plotted (Fig. 4.3B). Re-analysis of the LSK cells revealed two distinct clusters in PC1 (x-axis), which reflected the type of treatment. Similarly, in neutrophils (but not as strongly as in LSK cells), two clusters were observed in PC1 (x-axis), in accordance with the type of treatment. In both cell types, the genotype appears to be driving PC2 (y-axis). Overall, LSK cells were more affected by the

treatment than neutrophils: they showed a larger separation between the vehicle cluster and the HU cluster, compared to neutrophils (Fig. 4.3B). Distribution of the samples analyzed after removal in LSK cells and neutrophils is shown in Table 4.4.

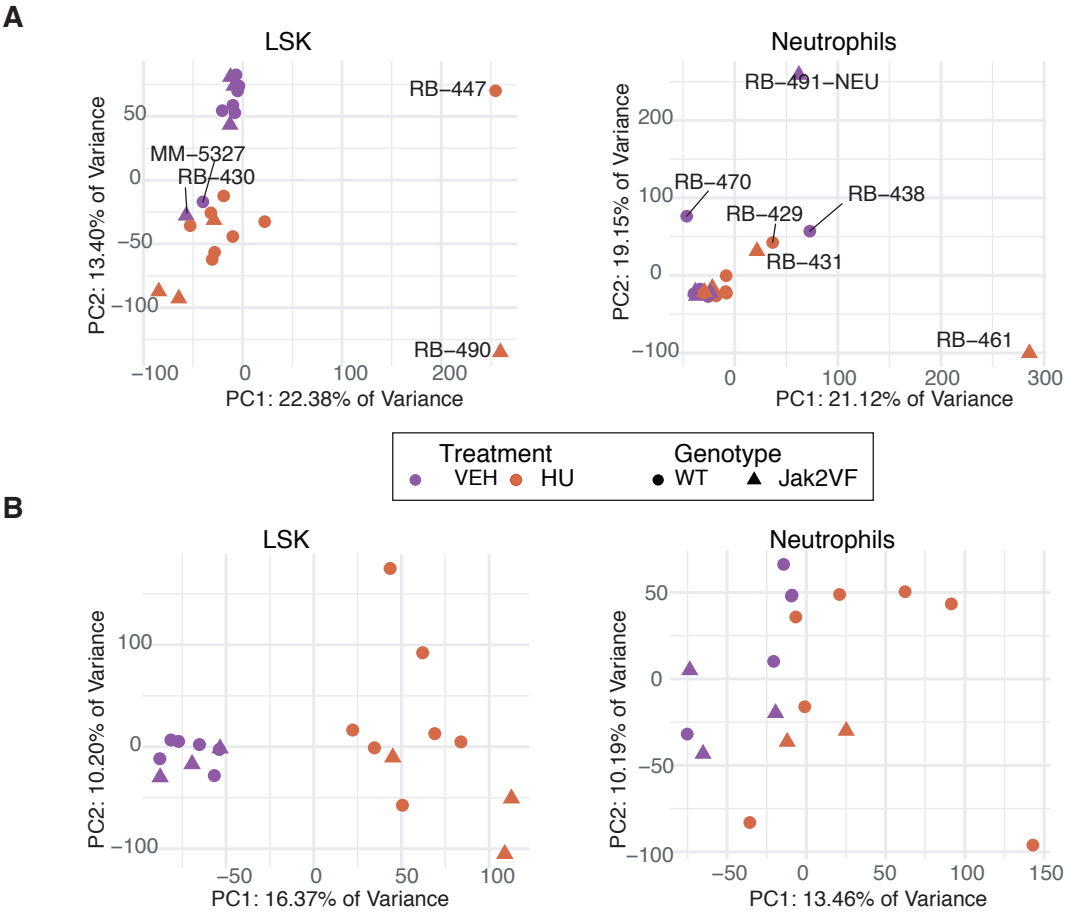

Figure. 4.3. Principal component analysis (PCA) of LSK cells and neutrophils. PCA was used to identify factors capable of driving variance among the samples sequenced. A) PCA of all samples per cell type. Samples that were considered outliers are labeled. In LSK, cells two samples from the vehicle group clustered with the HU-treated group, and two samples from the HU group did not cluster with any of the other samples. In neutrophils, six samples - three from the vehicle group and three from the HU group – were grouped separately from one main cluster. B) PCA following removal of outliers. In LSK a clear separation between HU and vehicle treated groups was observed. Similarly, in neutrophils, there is a separation between the samples from each treatment.

Table. 4.4. Total number of mice after removal per cell type, genotype and treatment

Treatment	LSK		Neutrophils	
	WT	Jak2VF-knock-in	WT	Jak2VF-knock-in
	wild-type or *vav*Cre$^+$		wild-type or *vav*Cre$^+$	
Vehicle (NaCl)	2/5	3	2/4	3
HU	1/6	3	1/5	2
Total	14	6	12	5

4.2.3 Overall effect of HU treatment on gene expression in mice

Differentially expressed genes (DEGs) were identified using Sleuth to test whether the expression levels of any given gene across mouse samples was significantly better explained by a linear model comprising genotype (WT or mutant) and treatment (HU vs vehicle) factors, versus a model with only the genotype (WT or mutant) factor. The DEGs (qval <0.05) were hierarchically clustered (Euclidean correlation distance and ward.D2 agglomeration method) according to their expression. Gene set enrichment analysis using MouseMine (Motenko *et al.*, 2015) and enrichR (Kuleshov *et al.*, 2016), were performed for each gene cluster, with gene sets defined by biological pathway or by the target of transcription factors (TFs).

In LSK cells, hierarchical clustering in columns was mainly divided by the type of treatment. Samples from the same genotype (mutant or WT) were also grouped. These observations suggested that DEGs affected by HU were also affected by the genotype. The rows (DEGs) were divided into six clusters, and each gene set was screened using the enrichment tools. Downregulated genes (cluster 1 and 2) were enriched for gene expression pathways with a high representation of genes involved in transcription regulation (Fig. 4.4A). MouseMine phenotype enrichment analysis identified abnormal haematopoietic phenotypes, including morphology and development, revealing that a large proportion of the genes from those clusters are involved in haematopoietic functions (Fig. 4.4A). This was also demonstrated in the TF en-

richment, where downregulated genes were controlled mainly by RUNX1, which is an important transcription factor for haematopoiesis. The *Runx1* gene was also differentially expressed and downregulated (Fig. 4.4A). Conversely, upregulated genes (clusters 3, 4, 5 and 6) were enriched for four main pathways, including pyruvate metabolism, cell cycle, immune system and RNA metabolism. MYC and MAX were identified as TFs controlling the DEGs. Among the MouseMine phenotypes, general terms, such as abnormal survival and preweaning lethality, were identified (Fig. 4.4A).

In neutrophils, hierarchical clustering in columns was mainly divided by the type of treatment. In the rows, two clusters that were defined by upregulated genes (cluster 1) and downregulated genes (cluster 2) were identified, but only cluster 2 was significantly enriched for pathways, MouseMine phenotype and transcription factors (Fig. 4.4B). Among pathways, enrichment in interleukin signalling and neutrophil degranulation were observed. MouseMine phenotype enrichment analysis identified terms related to abnormal haematopoietic system physiology. Similarly, TFs that control haematopoiesis, including RUNX1 and SPI1 (Fig. 4.4B), were enriched, although these genes were not differentially expressed in this dataset

These results were consistent with the observations in the human study. Two main pathways were affected by HU treatment: immune system and protein translation. However, using this mouse model of MPN, the effect of HU on the cell cycle was more evident than in the human study (Fig. 3.4).

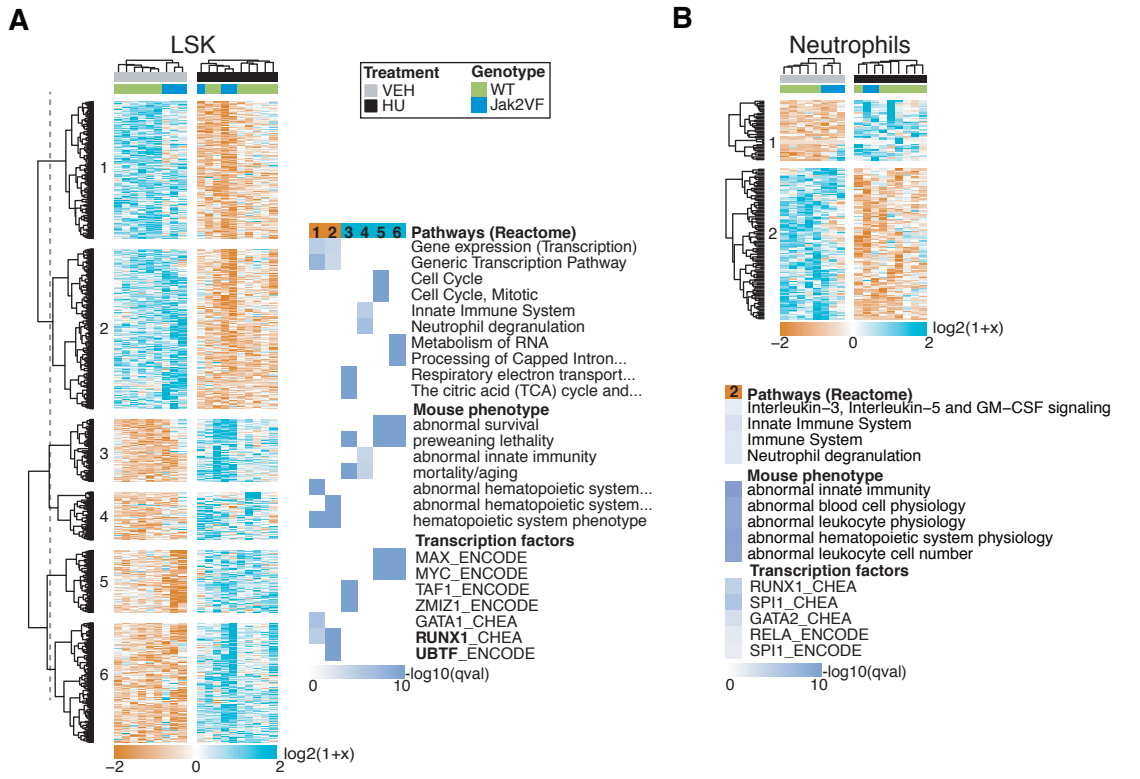

Figure. 4.4. Heatmap plot of gene expression analysis. Significant differentially expressed genes (DEGs) (q-value <0.05) were clustered according to their log2-fold change in expression. Clusters were divided at the same branch heights, and numbered. Enrichment analyses for pathways, mouse phenotype and transcription factors were performed in each cluster. The top significantly enriched results from each cluster are represented as a heatmap plot, where colour scale indicates significance. A) DEGs from LSK cells were divided into six clusters: clusters 1 and 2 represent downregulated genes and clusters 3 to 6 represent upregulated genes. B) DEGs from neutrophils were divided into two clusters: cluster 1 represents upregulated genes and cluster 2 downregulated genes. VEH: vehicle; HU: hydroxyurea; WT: wild-type; Jak2VF: Jak2V617F-knock-in mouse.

4.2.4 HU partially reverts deregulation of gene expression in Jak2VF-knock-in mice

In addition to the overall effects of HU treatment assessed in the previous section, the study also sought to determine whether HU treatment is able to rescue the genes disturbed by the Jak2V617F mutation. Therefore, comparisons were performed between Jak2VF mutant and WT mice, to identify genes that were affected by the mutation. The effect of the treatment was then evaluated in the Jak2VF mutant

genotype. Finally, to determine whether the genes disturbed by the Jak2V617F mutation can be rescued by HU, comparisons of the two datasets were performed.

4.2.4.1 Jak2VF mutant versus wild-type mice

To identify those genes which were altered in Jak2VF mutant mice, they were compared with WT mice in the vehicle group. The identified DEGs were screened using enrichment tools to obtain an overview of the data.

A total of 461 genes were differentially expressed in LSK cells, and 394 genes were differentially expressed in neutrophils. According to the enrichment analysis, in both cell types the effect of the mutation was an upregulation of genes involved in the immune system functions, which were mostly modulated by RUNX1 (Fig. 4.5 A and B). Accordingly, *Runx1*, *Stat3* and *Cebpb*, which participate in the transcriptional regulation of granulopoiesis, were upregulated in LSK cells. *Gata1,* which plays an important role during erythroid development, was also upregulated. In neutrophils, MouseMine phenotype enrichment analysis revealed that downregulated genes were also enriched in genes which are affected in abnormal immune cell proliferation and physiology (Fig. 4.5 B).

In conclusion, the main effect of Jak2VF mutation is the upregulation of genes involved in immune system pathways and which are regulated by the transcription factor RUNX1. Interestingly, downregulated genes in neutrophils were also involved in haematopoiesis.

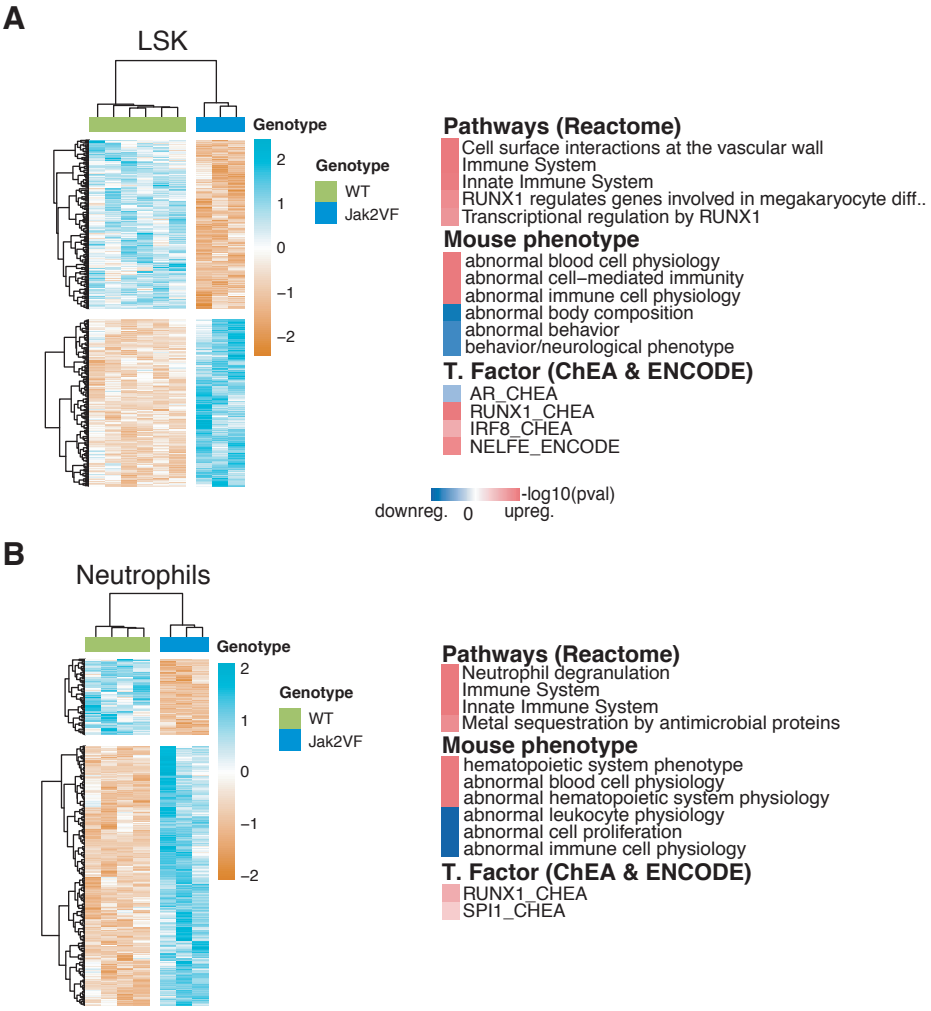

Figure. 4.5. Comparison of Jak2V617F-knock-in and wild-type mice in the vehicle group. Differentially expressed genes (DEGs) in Jak2V617F-knock-in mutant and wild-type mice were selected and enrichment analysis was performed for pathways, mouse phenotypes and transcription factors in A) LSK cells and B) neutrophils.

4.2.4.2 Jak2VF mutant: HU vs vehicle

To identify the effect of HU treatment on Jak2VF mutant mice, Jak2VF treated and untreated samples were compared, and enrichment analysis was performed.

The major effect in the two cell types was at stem cell level (1,135 DEGs in LSK cells as opposed to 76 DEGs in neutrophils). In LSK cells, enrichment analysis identified upregulation of pathways involved in RNA metabolism, the cell cycle,

and pyruvate metabolism. Downregulated pathways included cytokine signalling, where *Jak3*, *Stat3* and *Junb* genes were downregulated. Another pathway involved NOTCH1, where the *Notch1* gene was differentially expressed and downregulated. MouseMine phenotype enrichment analysis showed that downregulated genes were enriched for phenotypes associated with reduced immune cell numbers (Fig. 4.6A). Transcription factors enriched in downregulated genes were RUNX1 and GATA1 (Fig. 4.6A). In neutrophils, due to low number of DEGs, poor enrichment was observed. However, downregulated genes were enriched for abnormal cell physiology and cellular phenotype. Transcription factor enrichment identified CEBPD as the most relevant for neutrophil functions (Fig. 4.6B).

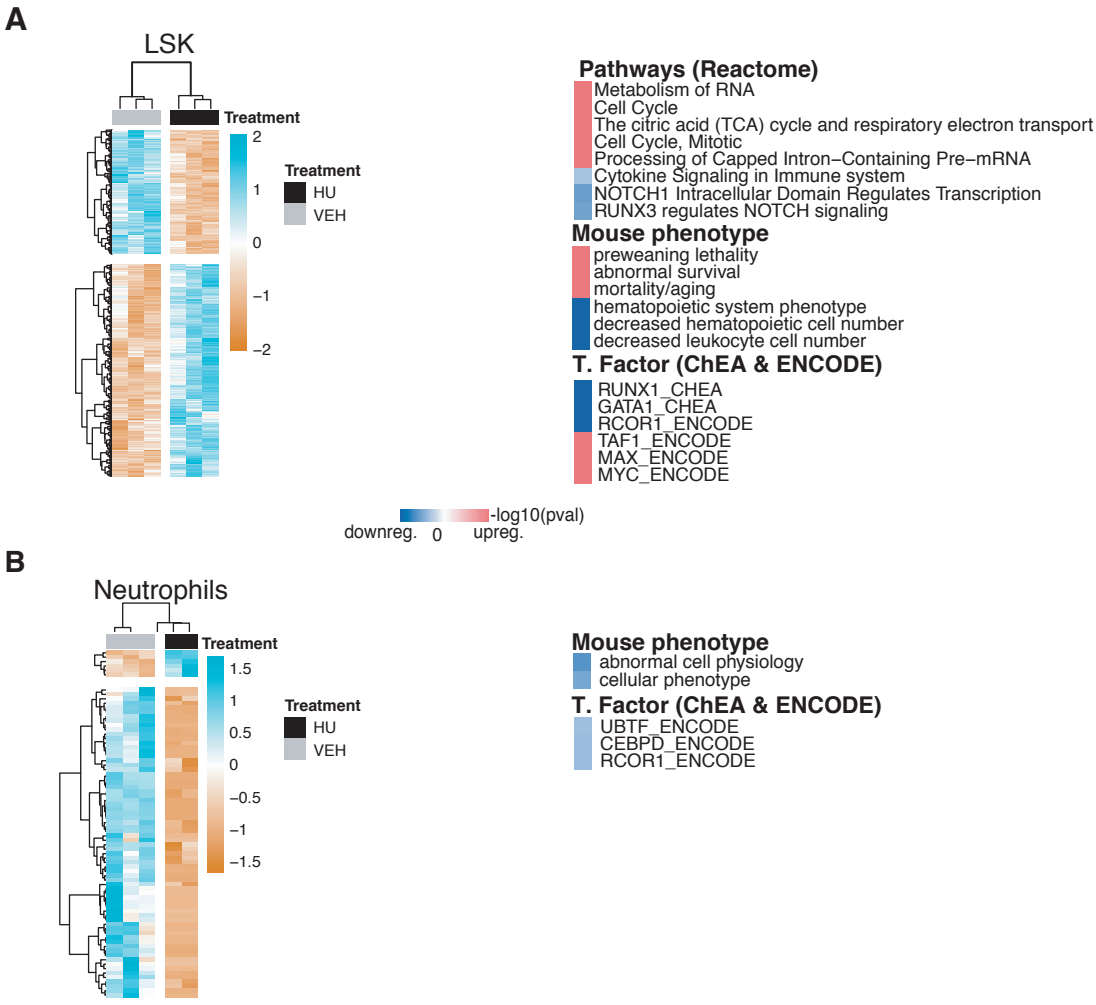

Figure. 4.6. Comparison of Jak2VF mutant treated and non-treated samples. Differentially expressed genes (DEGs) in Jak2VF mutant treated and non-treated mice were selected, and enrichment analysis was performed for pathways, mouse phenotypes and transcription factors in A) LSK cells and B) neutrophils.

4.2.5 HU reverses dysregulation of gene expression in Jak2VF mutant mice

These two analyses demonstrated that HU treatment downregulated the same pathways that were upregulated in Jak2VF mutant mice. Therefore, the effect of HU treatment on these genes was investigated by analysing DEGs from both sets of comparisons (WT vs Jak2VF mutant and Jak2VF mutant untreated vs untreated).

The mean expression of each gene was calculated per group: that is, Jak2VF treated, Jak2VF untreated, and WT untreated. Next, the log2-fold change per comparison was calculated using the mean expression value, and the results were plotted (Fig. 4.7). Enrichment analyses were performed, and the top significantly enriched terms were reported. Genes with the highest-fold change, or those involved in enriched pathways, were also selected (Fig. 4.7).

4.2.5.1 LSK cells

All the DEGs identified in the two previous analyses (WT vs Jak2VF mutant, and Jak2VF mutant untreated vs untreated) were selected. In LSK cells, a total of 1,596 genes showed differential expression between these two comparisons. The log2-fold change of these genes was calculated per comparison, and plotted (Fig. 4.7A). Most of the genes that were upregulated in the Jak2VF mutant (x-axis) were downregulated in the Jak2VF mutant mice treated with HU (y-axis) (Group 1) (Fig. 4.7A). Similarly, downregulated genes in the Jak2VF mutant (x-axis) were upregulated in the Jak2VF mutant mice treated with HU (y-axis) (Group 2). Therefore, among all genes analyzed, the expression of 1,170 genes was negatively correlated (Fig. 4.7A). Among the genes in Group 1, enrichment for immune system pathways, including IL-4 and IL-13 signalling and abnormal haematological phenotypes, were identified. Genes with the highest level of changes encoded for immunoglobin proteins (*Igll1*, *Ighv11-2*, *Igkv14-126* and *Vpreb1*) (Fig. 4.7A). Moreover, changes in genes involved in the JAK-STAT pathway, such as *Jak3*, *Stat3*, *Stat2* and *Socs3,* were observed. The most significant differentially expressed gene was *Serpina3g,* where loss of function has been associated with a deficit of granulocyte progenitors and increased HSC proliferation (Li *et al.*, 2014). In the case of *Myc*, two isoforms were differentially expressed in opposite directions, one in Group 1 and the other in Group 2 (Fig. 4.7A). Genes from Group 2 were enriched mostly for cell cycle, and metabolism of RNA pathways (Fig. 4.7A). In the cell cycle enrichment, genes that participate in

cell cycle checkpoints *(Wee1*, *Rpa1*, *Rpa2*), DNA synthesis (*Pold1*, *Fen1*), and G_1/S transition (*Ccnd2*, *Cdk4*), among others, were present (Fig. 4.7A). Metabolism of RNA included genes involved in pre-mRNA processing and mRNA (*Snrpb*, *Snrpd1*, *Snrpd2*) (Fig. 4.7A). Finally, the genes with the highest-fold change were *Cmc2* and *Gtf2i*. Transcription factor enrichment identified similar proteins, as in the Jak2VF mutant treated versus non-treated comparison (Fig. 4.6 A).

These findings demonstrate that HU downregulates genes which were upregulated in the Jak2VF mutant mice. These genes were enriched for immune system-related pathways, which is in addition to the known impact of HU on the cell cycle. Interestingly, some of these genes were also members of the JAK-STAT pathway, and involved in interleukin signalling, revealing that the direct effect of HU was to downregulate pathways which were altered in the Jak2VF mutant mice, and which constituted their MPN phenotype. Moreover, transcription factors known to be important for haematopoiesis were shown to be regulating the genes downregulated by HU treatment.

The effect of HU treatment on the cell cycle was also evident. Genes that were downregulated in the Jak2VF mutant mice, and upregulated by HU treatment, corresponded to genes participating in the cell cycle, and were controlled by the transcription factor MYC. As expected, these findings confirmed the main mechanism of action of HU.

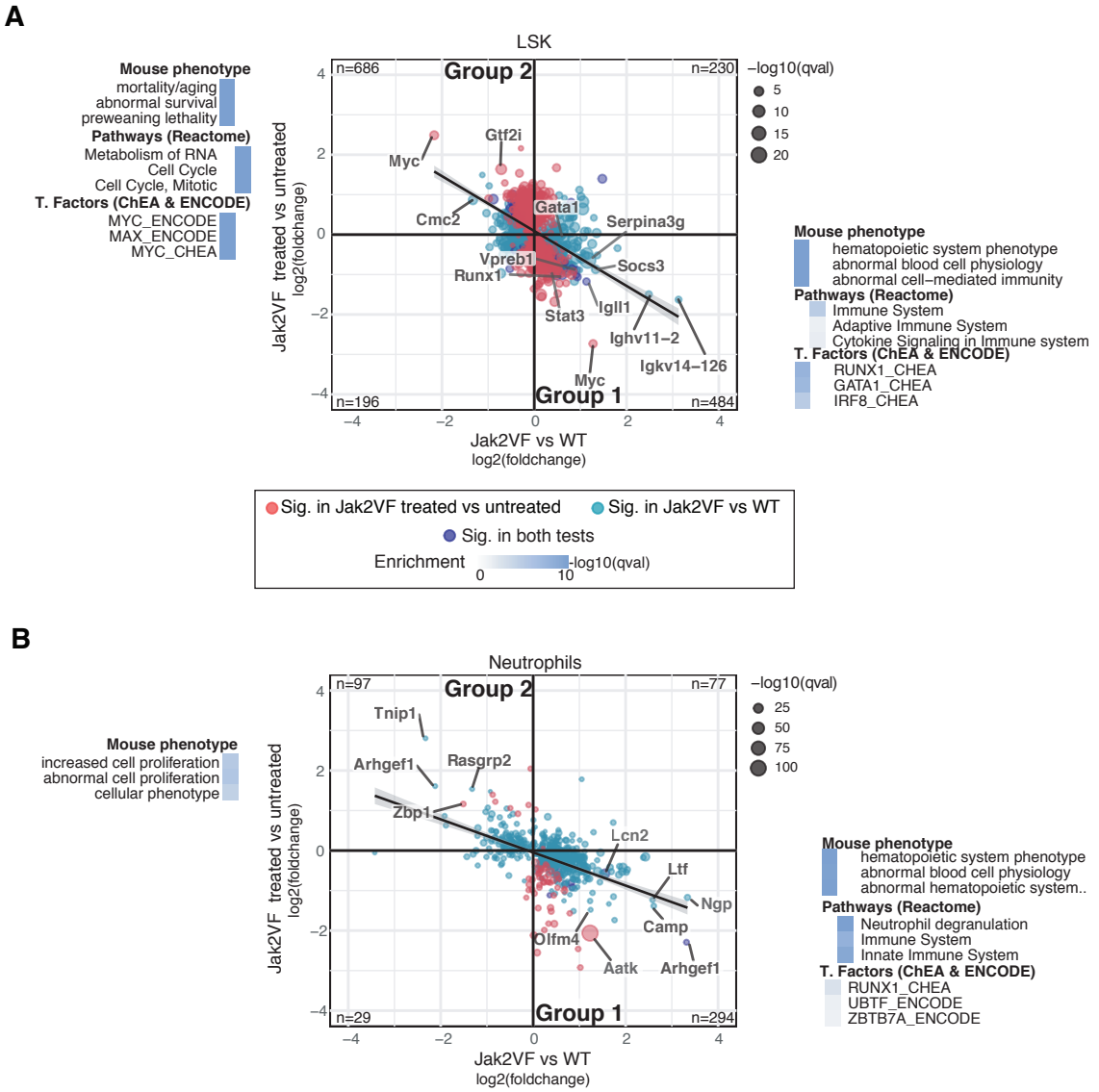

Figure. 4.7. Comparison between wild-type vs Jak2VF, and Jak2VF treated vs untreated. All the differentially expressed genes (DEGs) identified in the analysis of wild-type (WT) vs Jak2VF (x-axis), plus all the DEGs identified in the analysis of Jak2VF treated vs untreated (y-axis), were selected, and the log2-fold change was calculated between each group comparison. Group 1 represents genes that were upregulated between WT vs Jak2VF, and downregulated in the treated Jak2VF mutant compared to untreated. Group 2 represents genes that were downregulated between WT vs Jak2VF, and upregulated in Jak2VF mutant untreated vs treated. Enrichment analysis was performed in Group 1 and Group 2. Colour scale indicates the significance of each enriched term. Number of DEGs in each section are indicated. Size of the symbol represents significance of the gene expression change. These analyses were performed in A) LSK cells and B) neutrophils.

4.2.5.2 Neutrophils

In neutrophils, a total of 497 genes were differentially expressed in the two previous comparisons (WT vs Jak2VF, plus Jak2VF treated vs untreated). These genes were plotted according to the log2-fold change between each comparison (Fig. 4.7B). Most of the DEGs that were upregulated in the Jak2VF mutant mice (x-axis) were downregulated following HU treatment (y-axis) (Group 1) (Fig. 4.7B). Similarly, genes that were downregulated in the Jak2VF mutant mice (x-axis), were upregulated following HU treatment (y-axis) (Group 2). Genes from Group 1 were enriched for immune system functions. In this case, genes involved in neutrophil degranulation (*Camp*, *Itgam*, *Ltf* and *Olfm4*) and IL-4 and IL-13 signalling (*Lcn2*) were significantly enriched. *Aatk*, which is induced during apoptosis, was the gene which showed the most differential expression between Jak2VF mutant untreated and treated (red dot) samples (Fig. 4.7B). Neutrophilic granule protein (*Ngp*) exhibited the highest-fold change in Jak2VF mutants as compared to WT (blue dot). This gene is controlled by CEBPE and SPI1, and its protein dysregulation has been linked to metastatic tissues in myeloid-derived suppressor cells (Boutté *et al.*, 2011; Gombart *et al.*, 2003). *Arhgef1* (or *Lsc*) had the highest-fold change among downregulated genes by HU treatment. Studies in *Arhgef1* knock-out mice have shown that it is necessary for normal migration and adhesion in neutrophils (Francis *et al.*, 2006). However, a different isoform of *Arhgef1* was upregulated in Jak2VF mutant treated mice (y-axis) (Group 2). Further upregulated genes in treated Jak2VF mutant mice (Group 2) were *Tnip1,* which has been associated with autoimmune and inflammatory diseases when its protein (ABIN-1) levels are low (Kuriakose *et al.*, 2019), and *Zbp1*, encoding the protein DLM-1, which acts as a cytosolic DNA sensor, and triggers the activation of the innate immune system (Wang *et al.*, 2008). *Zbp1* is also involved in programmed cell death and inflammation (Kuriakose *et al.*, 2016). *Rasgrp2* was also differentially expressed; mutations in this gene have been associated with bleeding disorders due to decreased platelet aggregation (Sevivas *et al.*,

2018). Transcription factor enrichment analysis identified RUNX1 as the principal regulator of downregulated genes following HU treatment.

As with LSK cells, in neutrophils, HU downregulated genes involved in immune system pathways, and, in this case, genes that are relevant for neutrophil functions. Interestingly, genes upregulated by HU treatment were also of relevance in the haematopoietic system. Both analyses, in LSK and neutrophils, identified RUNX1 as the principal regulator of genes affected by HU treatment, confirming the previous observations (Fig. 3.4).

4.2.6 Effect of HU treatment in human and mouse: a comparison

Identifying genes affected by HU in two different species is a good strategy for the detection of candidate genes which explain the mechanism of action of HU, or the clinical outcomes of HU treatment. Accordingly, all DEGs homologous in human and mouse were selected (Fig. 4.8A) and compared, and those DEGS which were changing in the same direction (i.e. up- or downregulated) were subjected to further analysis.

At the stem cell level ($CD34^+$ and LSK cells), 425 genes were found to be up/downregulated in the two species, but only 21 genes in neutrophils (Fig. 4.8A). These genes were subjected to enrichment analysis, and potential candidate genes which might elucidate the mechanism of action of HU, or its clinical outcomes, were reported.

Common upregulated genes from stem cells were enriched for neutrophil degranulation and innate immune system pathways (Fig. 4.8B), which was in accordance with the results previously observed in the gene expression analysis of MPN patients and Jak2VF mutant mice. Enrichment analysis for regulating TFs revealed that SPI1 and GATA1 act as regulators for up- and downregulated genes, respectively. No significant pathways were identified in downregulated genes. However, enrich-

ment for genes involved in "transcription regulation" (which included a large number of TF encoding genes) was observed, although this was not significant (Table. 4.5). Some of these, such as *HOXB4* and *BCL11A*, are involved in haematopoiesis. Interestingly, some of the differentially expressed TFs such as *RUNX1*, *ZBTB7A*, *ETS1* and *ATF3,* were also enriched (Table. 4.5 in bold). Other genes classed as transcriptional regulators, such as *KDM5B* and *TET1*, are known to be involved in epigenetic mechanisms (Table 4.6).

In the previous Chapter, it was shown that DNA methylation from $CD34^+$ cells was affected by HU treatment. Accordingly, all genes common to human and mouse, and annotated as epigenetic modifiers (Medvedeva *et al.*, 2015), were retrieved in order to identify genes which might explain the effect of HU in DNA methylation through epigenetic mechanisms of gene modulation (Table 4.6). Interestingly, most of the epigenetic modifiers identified have functions in histone modification rather than in DNA methylation, with the exception of the *TET1* gene, which encodes a protein involved in the oxidation of 5-methylcytosine to 5-hydroxymethylcytosine and further oxidation (Table 4.6).

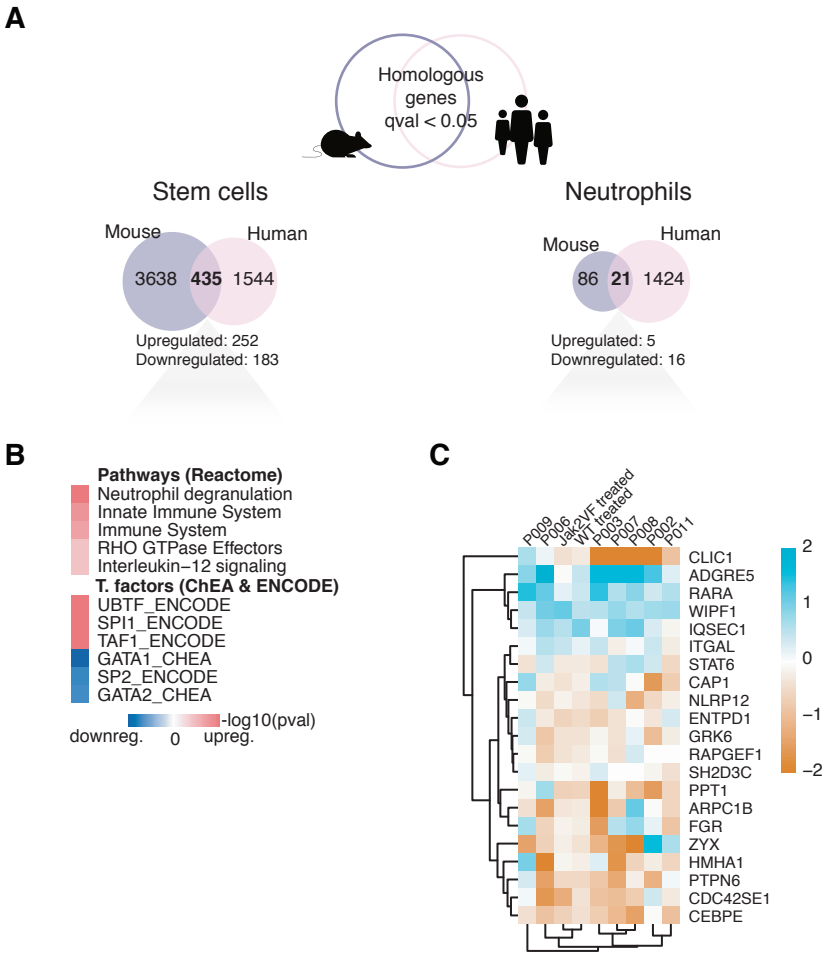

Figure. 4.8. Comparison between genes differentially expressed in human and mouse, following HU treatment A) Significantly differentially expressed genes shared between the two species were selected for each cell type. A total of 435 differentially expressed genes were shared in $CD34^{+}$/LSK cells, and 21 in neutrophils. B) Enrichment analysis for pathways and regulating transcription factors was performed at the stem cell level. C) Heatmap plot of log2-fold change of 21 differentially expressed genes shared at neutrophil level between human and mouse. Columns represent individual patients, and their corresponding rows are the log2-fold change of "after treatment" versus "before treatment". Mice group comparison (untreated vs treated).

In neutrophils, enrichment analysis was not performed because of the low number of shared DEGs (Fig. 4.8A). However, the log2-fold changes per patient (after HU treatment vs before HU treatment) and per mice group comparison were plotted. Some genes having homogenous expression between patients and mice were identified (Fig. 4.8B). This was the case for *RARA* and *WIPF1* among the downregulated genes and *CEBPE* among the upregulated genes (Fig. 4.8B). Interestingly, the

transcription factors RARA and CEBPE are involved in neutrophil maturation.

In conclusion, at the stem cell level, there was a greater overlap of DEGs between human and mouse Interestingly, SPI1, which was differentially expressed in $CD34^+$ cells was also enriched in the upregulated genes in mouse, although *Spi1* was not differentially expressed in mouse. Overall, this analysis identified important TFs, which confirmed findings in the separate analysis of human and mouse, including *RUNX1*, *BCL11A* and *ZBTB7A*. In neutrophils, poor overlap was observed, probably due to the low number of DEGs identified in mouse.

Table. 4.5. Common differentially expressed transcription factors between human and mouse

Symbol	Direction	Human			Mouse		
		b	q-val trans.	q-val gene	b	q-val trans.	q-val gene
NR4A2	upregulated	0.841	0.046	0.039	0.559	0.008	0.024
MAFB	upregulated	1.591	0.003	0.001	1.467	0.005	0.003
MEF2D	upregulated	0.372	0.008	0	0.076	0.396	0.029
ATF3	upregulated	1.392	0.005	0	1.521	0.007	0.005
ZBTB7A	upregulated	0.432	0.001	0	0.134	0.529	0
FOXJ2	upregulated	0.229	0.071	0.028	0.111	1	0.008
MAFG	upregulated	0.297	0.084	0.019	0.084	0.408	0.026
ETS1	upregulated	0.486	0.013	0.001	0.08	1	0
ELK4	upregulated	0.28	0.101	0.05	0.118	0.532	0.003
FOXN3	upregulated	0.127	0.169	0.023	0.651	0.11	0
ARID5A	upregulated	0.578	0.006	0.003	0.051	1	0
ID2	upregulated	1.245	0.003	0	0.591	0.01	0.014
PLAG1	downregulated	-1.04	0.002	0	-1.317	0	0
HOXA5	downregulated	-0.322	0.063	0.046	-0.299	0.018	0.013
RUNX1	downregulated	-0.225	0.089	0.001	-0.305	0.001	0.001
BCL11A	downregulated	-0.516	0.018	0	-0.298	0	0
MXD1	downregulated	-0.595	1	0.008	-0.434	0	0
STAT6	downregulated	-1.268	0.342	0.005	-0.129	0.012	0.008
ERG	downregulated	-0.549	0.005	0	-0.465	0.001	0.001
ZNF652	downregulated	-0.22	0.223	0.002	-0.3	0	0
HOXB4	downregulated	-0.472	0.033	0.023	-0.226	0.058	0.042
IKZF2	downregulated	-0.33	0.128	0.029	-0.297	0.002	0
MGA	downregulated	-0.226	0.144	0.023	-0.192	0.007	0.045
TFDP2	downregulated	-0.588	0.003	0	-0.279	0.001	0.013
KAT6A	downregulated	-1.101	0.21	0.011	-0.174	0.024	0.048
KAT6B	downregulated	-0.56	0.057	0.029	-0.351	0.006	0.029
KDM5B	downregulated	-0.29	0.018	0.001	-0.495	0	0

b: log(foldchange) computed by Sleuth; trans.: transcript level. Transcription factors enriched denoted in bold
Dataset of transcription factor from Chawla *et al.* (2013)

Table. 4.6. Differentially expressed genes in $CD34^+$ and LSK cells that are classified as epigenetic regulators. Gene expression details are indicated for human and mouse.

Symbol	Direction	Human			Mouse			Function	Modification
		b	q-value trans.	q-value gene	b	q-value trans.	q-value gene		
PRKCD	upregulated	1.075	0.002	0	2.074	0.028	0	Histone modification	#
PKM	upregulated	1.516	0.002	0	0.307	0	0	Histone modification write cofactor	Histone phosphorylation
RPS6KA4	upregulated	0.437	0.014	0	0.191	0.069	0.009	Histone modification write	Histone phosphorylation
MBD2	upregulated	0.518	0.003	0.001	0.131	0.017	0.011	Histone modification write cofactor, Histone modification erase cofactor, TF	Histone methylation, Histone acetylation, TF repressor
SETD1B	upregulated	0.252	0.028	0.007	0.131	0.222	0.002	Histone modification write	Histone methylation
PPP4C	upregulated	0.351	0.014	0.009	0.499	0.021	0.005	Histone modification erase	Histone phosphorylation
TLE1	upregulated	0.471	0.026	0.012	0.669	0.009	0.01	Chromatin remodeling, Histone modification cofactor	#
UCHL5	upregulated	2.274	0.036	0.018	0.303	0.005	0	Histone modification erase cofactor	Histone ubiquitination
CDYL	upregulated	0.336	0.071	0.023	0.07	0.076	0.049	Histone modification write	Histone acetylation
GATAD2A	upregulated	0.309	0.048	0.032	0.304	0.001	0.001	Histone modification read	#
SYNCRIP	upregulated	0.062	0.253	0.035	0.253	0.001	0	RNA modification	mRNA editing
MASTL	upregulated	1.266	0.079	0.035	0.289	0.009	0.006	Histone modification write	Histone phosphorylation
BRD2	upregulated	0.342	0.05	0.039	0.134	0.804	0.001	Histone modification read	#
MGA	downregulated	-0.226	0.144	0.023	-0.192	0.007	0.045	Histone modification write cofactor, TF	Histone methylation, Histone acetylation, TF activator, TF repressor
KDM5B	downregulated	-0.29	0.018	0.001	-0.495	0	0	Histone modification erase	Histone methylation
CBX2	downregulated	-0.539	0.008	0.002	-0.249	0.008	0.005	Histone modification read	#
TET1	downregulated	-0.408	0.005	0.002	-0.354	0.003	0.001	DNA modification	DNA hydroxymethylation
PRDM11	downregulated	-2.678	0.024	0.01	-0.343	0.002	0	Histone modification write	Histone methylation
KAT6A	downregulated	-1.101	0.21	0.011	-0.174	0.024	0.048	Histone modification write	Histone acetylation
EPC1	downregulated	-0.254	0.245	0.028	-0.371	0.001	0.012	Polycomb group (PcG) protein	#
KAT6B	downregulated	-0.56	0.057	0.029	-0.351	0.006	0.029	Histone modification write	Histone acetylation
BPTF	downregulated	-0.161	0.099	0.038	-0.162	0.008	0	Chromatin remodeling	#
STK4	downregulated	-0.838	1	0.039	-0.215	0.001	0.001	Histone modification write	Histone phosphorylation
ZMYM2	downregulated	-0.209	0.164	0.04	-0.193	0.01	0.001	Histone modification erase cofactor, TF	Histone acetylation
ASH1L	downregulated	-0.342	0.091	0.046	-0.192	0.009	0.007	Histone modification write	Histone methylation

b: log(foldchange) computed by Sleuth; trans.: transcript level. Dataset of epigenetic modifiers from Medvedeva *et al.* (2015)

4.3 Conclusions

A Jak2V617F-knock-in (Jak2VF) mouse model of MPN, and wild-type mice, were successfully treated with HU, as demonstrated by the reduction of CBC and spleen size. From these mice, two comparable counterparts in human, LSK cells and neutrophils, were isolated and whole gene expression was assessed. Gene expression analysis showed that LSK cells were more affected by the treatment than neutrophils, due to the high number of differentially expressed genes (DEGs) identified in LSK cells. According to the enrichment analysis of LSK cells, HU affected genes involved in the cell cycle, the innate immune system, RNA metabolism, the citric acid cycle, and transcription. RUNX1 was identified as the principal transcription factor regulating the DEGs. RUNX1 was also differentially expressed and down-regulated. Specifically, immune system pathways in Jak2VF mutant mice, were the most de-regulated compared to WT mice in both cell types, LSK and neutrophils. Interestingly, changes in those pathways were reversed by HU treatment. An effect of HU on the cell cycle – mostly an up-regulation of genes involved in the checkpoint pathways – was also observed.

The DEGs in the patient study were compared to DEGs affected by HU in mice. At the stem cell level ($CD34^+$ and LSK cells), transcription factors important for haematopoiesis were shown to be differentially expressed in both species, and some of these were also identified as regulators of DEGs using enrichment analysis. Interestingly, several epigenetic modifiers were differentially expressed between the two species: these were mostly involved in histone modification rather than DNA methylation.

Chapter 5

Discussion

HU is the first-line treatment for high-risk PV and ET patients. PV and ET are MPNs which give rise to a high number of blood cells increasing a patient's risk of thrombotic events and bleeding. HU effectively reduces blood cell counts, and improves symptoms such as headache and fatigue (Löfvenberg and Wahlin, 1988; Martínez-Trillos *et al.*, 2010). HU has been in use for more than 30 years but its only well known mechanism of action involves the inhibition of the RNR that is necessary for DNA synthesis. This mechanism leads to cell cycle arrest, which can result in cell death. Interestingly, HU can be used to treat sickle cell anaemia patients as it is able to induce HbF expression. This is a paradoxical effect, however, since the main effect of HU is to reduce the number of blood cells. Several mechanisms have been proposed to explain HbF induction, including a pathway involving nitric oxide production by HU and the soluble guanylate cyclase in erythroid progenitors cells (Platt *et al.*, 1984; Cokic *et al.*, 2006; Lou *et al.*, 2009). Other mechanisms propose that HU produces hypomethylation of the γ-globin promoter, facilitating the transcription of HbF (Walker *et al.*, 2011). Moreover, 5-azacytidine and 5-aza-2´-deoxycytidine (decitabine), which inhibit DNA methylation, are also known to induce HbF expression (Sauntharajah *et al.*, 2003; Ley *et al.*, 1983). However, these mechanisms are still not well understood, and no studies have assessed the

global effect of HU on DNA methylation in MPN patients.

Previous work suggests HU may affect gene expression, by affecting DNA methylation. Accordingly, in the study described here, I assayed and analyzed the effect of HU treatment on DNA methylation and gene expression in MPN patients, and in a MPN mouse model. For this, two clinically relevant and differentially developed cells (stem cells and neutrophils) were isolated from each species. Through the work described in this thesis I sought to identify genes that are affected by HU treatment in two species, with the aim of achieving a better understanding of its mechanism of action.

5.1 The Jak2V617F-knock-in mouse: a good model to test HU treatment?

Recognition of a single gain-of-function mutation in the *JAK2* gene (JAK2V617F) has been one of the most important discoveries in MPN patients, which has permitted a better understanding of the pathogenesis of MPN. To investigate the impact of the JAK2V617F mutation on the haematopoietic system, several groups have generated MPN mouse models through induction of the same mutation.

Indeed, mouse studies have certain advantages over human studies. Mice have the same genetic background, are inbreed and are housed in controlled environments where food and treatment are consistent. Mouse studies also permit the study of a large number of individuals, and provide a wider range of samples for analysis. However, they also impose some limitations on potential studies, which is not surprising since the two species diverged around 65-75 million years ago (Mestas and Hughes, 2004). In terms of clinical trials, these differences have meant that only 8% of the drugs tested in mouse are translated to human patients (Mak *et al.*, 2014). Interestingly, this has not been the case for HU, since its test in a mouse model of leukemia led to its current use in the treatment of myeloproliferative diseases

(Stearns *et al.*, 1963; Thurman *et al.*, 1963). Similarly, MPN mouse models have been used to test the efficacy of novel drugs such as ruxolitinib, which inhibits the JAK1/JAK2 proteins (Kubovcakova *et al.*, 2013; Vaddi *et al.*, 2010; Mullally *et al.*, 2010). However, these studies were not designed to elucidate the mechanism of action of HU, nor determine whether gene expression is affected by drugs under investigation. Instead, research focussed on the analysis of the modulation of the JAK-STAT pathway by drugs at the molecular level, and the resulting effects on the phenotype of the MPN mouse model. Therefore, in this study I have chosen to compare a mouse model of the disease, with all the advantages that this brings, with human MPN patients treated with HU, in an experimental design which allows for the similarities and differences between the two species to be compared.

This study utilised a Jak2V617F-knock-in (Jak2VF) mouse which expressed the mutation only in the haematopoietic compartment (Chen *et al.*, 2015; Mullally *et al.*, 2010). Phenotypically these mice exhibited splenomegaly, with high levels of white blood cells, erythrocytes and platelets, thus recapitulating the MPN phenotype (Chen *et al.* 2015 and this work). The knock-in mutant mice, together with a group of wild-type C57Bl/6 and *vav*Cre^{+} transgenic mice (referred as WT), were treated with HU or vehicle for a period of six weeks. CBC were measured every week to assess the effectiveness of the treatment. Congruent with the findings from the human section of this study, CBCs were reduced by HU treatment in WT and Jak2VF mutant mice. Interestingly, the number of platelets decreased more rapidly in Jak2VF mutant mice than in WT mice treated with HU. Other studies have shown that platelet counts in patients harboring the JAK2V617F mutation who were treated with HU were more significantly reduced than those in JAK2V617F-negative patients (Sirhan *et al.*, 2008; Panova-Noeva *et al.*, 2011). For this reason, lower doses of HU are usually required to control the platelet count in JAK2V617F patients than in other MPN patients (Sirhan *et al.*, 2008). Moreover, platelet levels in the two JAK2V617F-negative patients included in the present study remained

higher than in the JAK2V617F patients. Accordingly, these similarities between human and mouse at the levels of disease development and phenotypic response to treatment, make the mouse a good model in which to study the effect of HU on DNA methylation and gene expression.

At the end of the treatment programme in mouse, spleens, LSK cells and neutrophils were isolated. LSK cells are HSCs capable of giving rise to all cell lineages (Spangrude *et al.*, 1988; Uchida, N, and Weissman, 1992), while neutrophils represent a fully differentiated cell type from the myeloid lineage. These cells were selected as counterparts of the cell types used in the human study, which were $CD34^{+}$cells (haematopoietic stem and progenitor cells) and neutrophils (Civin *et al.*, 1984). Gene expression was measured in all these cell types from human and mouse. Overall, enrichment analysis of differentially expressed genes (DEGs) demonstrated that similar pathways were altered in both species. These pathways were involved in RNA metabolism and immune system functions such as interleukin signalling and neutrophil degranulation. However, in LSK cells a higher number of DEGs were identified which also led to enrichment in other pathways not seen in the human study e.g. the cell cycle and the citric acid cycle. One reason for this could be that the mouse model carries the Jak2VF mutation in every haematopoietic cell, in contrast with human patients, where the cellular mutation status is mosaic (Butcher *et al.*, 2007). This difference might have led to a more severe phenotype in mouse than in human, thus magnifying the effects of HU treatment. Moreover, other biological differences exist between the two cell types. LSK cells are a more homogenous population of stem cells, whereas $CD34^{+}$cells are HSCs and progenitor cells. In this sense, the purity of the cell population is important for gene expression analysis since the transcriptomic landscape can vary between cell types. The two cell types also exhibit technical differences in RNA sequencing. In this study, RNA quality and concentration were better in LSK cells than in $CD34^{+}$ cells, which led to the sequencing of a wider range of RNA molecules. Unfortunately, $CD34^{+}$cell isola-

tion was performed using peripheral blood, where the CD34 marker is expressed in around 0.1% of cells (Bender *et al.*, 1992). However, overlap of DEGs in both species was still observed, which led to the identification of transcription factors already described in the previous analyses from human and mouse in isolation (Chapter 3 and 4, respectively).

Conversely, in neutrophils, the number of DEGs responding to HU treatment in mouse displayed an opposite trend. Only 155 genes were differentially expressed in mouse neutrophils, compared to 2,223 DEGs in human neutrophils. This also led to a low number of common DEGs between the two species. This might be due to the low concentration of RNA isolated from mouse neutrophils, and the lower sequencing depth used, which meant that only highly expressed genes were identified. More importantly, clear biological differences exist between the two species. For example, neutrophils are rarer in mouse than in human: they account for only 10-30% of blood cells in mouse, as opposed to 50-70% of blood cells in human (Doeing *et al.*, 2003; Hidalgo *et al.*, 2019). Mouse neutrophils do not express some genes present in human neutrophils: these include several chemokines (CXCL8 and its receptor CXCR1) and immunoglobin receptors (FcαRI and FcγRIIA), which are involved in important functions of the neutrophil response against pathogens (Otten *et al.*, 2005; Olson and Ley, 2002). These differences in immune response can be attibuted to the pathogen-free environment where mice reside, which are obviously very different from human environments. All of these differences made it difficult to compare neutrophils from the two species at the gene expression level.

Overall, the MPN mouse model used in this study recapitulated the phenotype of the patients during HU treatment. The mouse model also led to a stronger response at the stem cell level. However, the findings observed in mouse must be interpreted cautiously. Human patients have a much more complex disease since they are exposed to many different stimuli that cannot be recreated in mouse, and it is very likely that other factors also contribute to disease status in patients.

5.2 Comparative response to HU treatment in stem cells and neutrophils

Two cell populations, at two stages of differentiation, were analyzed to determine whether the effects of HU treatment were maintained through differentiation, and whether the effects of HU were of the same intensity in two differentially developed cell types. HSCs were selected since they harbour the JAK2V617F mutation, and identifying the effects of HU treatment at the stem cell level could allow the prediction of haematopoietic outcomes. Neutrophils were selected because they also harbour the JAK2V617F mutation, are found in high levels in MPN patients, and are significantly affected by HU treatment, and play an important role in inflammation. Both cell types, therefore, constitute interesting models in which to assess the effects of HU treatment.

In terms of the number of DEGs affected in HSCs and neutrophils, the effect was slightly higher in $CD34^+$ cells compared to human neutrophils. However, the difference between LSK and mouse neutrophils was striking. These observations suggest that the stem cell populations ($CD34^+$ and LSK cells) were more sensitive to HU treatment than neutrophils. This could be explained by the high plasticity of stem cells comparing to differentiated cells. However, when comparing the effect of HU treatment in the Jak2VF mutant mice, a similar response was observed in both LSK and neutrophils. Specifically, in both cell types HU was able to rescue the genes that were dysregulated by the Jak2V617F mutation. Moreover, the same transcription factor (RUNX1) was shown to be involved in the regulation of those genes. Similar enrichment pathways were identified in both $CD34^+$ cells and neutrophils in human, and the same transcription factor, SPI1 (also known as PU.1), was shown to be involved in the regulation of DEGs in both cell types. Therefore, although the effect of HU on stem cells and neutrophils differs in intensity, it is still comparable.

5.3 Distinct role of HU in DNA methylation in neutrophils and CD34$^+$ cells

Studies that have attempted to assess the effects of HU treatment have measured DNA methylation both in a site-specific manner, and relative to sickle cell anaemia patients (Walker *et al.*, 2011; Chondrou *et al.*, 2018). One study assayed whole-genome DNA methylation in MPN patients using neutrophil samples, but the aim was to discern whether DNA methylation was capable of determining clustering according to the type of MPN, which it was (Nischal *et al.*, 2013). Therefore, DNA methylation has not been examined to determine the effects of HU treatment either in a genome-wide manner, or in MPN patients.

In this study DNA methylation was assayed in CD34$^+$ cells and neutrophils from MPN patients, and comparisons were made prior to, and following HU treatment. In neutrophils, very small changes in DNA methylation were observed. These were more frequent in individual CpG sites than in differentially methylated regions (DMRs), and at later time-points of treatment (six and nine months). These results suggest that DNA methylation in neutrophils is stable, and HU is not able to affect it at the time-points analyzed in this study. Neutrophils are fully differentiated cells, the final product of granulopoiesis. Therefore, it is possible that the DNA methylation machinery is not as active here as it is in HSCs, which need to be highly plastic in order to give rise to many different types of cell. In line with this, studies have demonstrated that the lymphoid lineage makes more use of DNA methylation than the myeloid lineage (Farlik *et al.*, 2016). These observations were based on DNA methylation assays using whole genome bisulfite sequencing in several purified human cell lineages. These revealed that lymphoid lineages tend to be more methylated in regulatory regions. In the myeloid lineage by contrast, regions with lower DNA methylation than in the lymphoid lineage were enriched for binding sites of transcription factors (TFs) that are key for myeloid differentiation

(Farlik *et al.*, 2016). These findings suggest that the lymphoid lineage requires DNA methylation to protect it from myeloid specification (Farlik *et al.*, 2016; Bock *et al.*, 2012)and is consistent with tissue-specific differences in DNA methylation patterns during development and differentiation. This might, therefore, explain why HU treatment at a very late differentiation stage, and in the myeloid lineage, cannot perturb DNA methylation as seen in HSCs. However, in this study, gene expression was still affected, indicating that other DNA methylation-independent mechanisms are involved.

5.4 *SPI1* is an important target of HU treatment at DNA methylation and gene expression level

The effect of HU on $CD34^+$ cells was different. Following nine months of HU treatment, several CpG sites and DMRs were differentially methylated when comparisons were made with their pre-treatment state. Detailed analyses were performed at DMRs overlapping promoters, intergenic and intragenic regions. These analyses identified several binding sites for TF involved in haematopoiesis, such as RUNX3 and ZBTB7A. However, the gene most highly affected at the DNA methylation and gene expression levels was *SPI1*. *SPI1* was upregulated following HU treatment and two CGIs, in its promoter and regulatory region (located -17kb from the promoter), were shown to be hypomethylated. *SPI1* encodes the TF SPI1 (also known as PU.1), which has different roles during haematopoiesis. In the mouse, germline disruption of *SPI1* blocks the formation of all cell lineages, leading to embryonic death (Scott *et al.*, 1994). However, its major role has been described in myeloid progenitor cells, where its conditional deletion impairs differentiation into myeloid and monocyte cells (Iwasaki *et al.*, 2005). Interestingly, in acute myeloid leukaemia (AML) impaired expression of *SPI1*, caused by mutations in the regulatory region of *SPI1*, has been observed (Bonadies *et al.*, 2010; Mueller *et al.*, 2002). These in-

vestigations suggest that SPI1 is necessary for the induction of cell differentiation, moving from the stem cell stage to myeloid specification. Interestingly, SPI1 expression is also impaired in chronic myeloid leukemia (CML), where its expression can be recovered following drug treatment with interferon-α and imatinib (Albajar *et al.*, 2008). Aberrant methylation has also been identified in the *SPI1* promoter in whole blood from CML patients (Yang *et al.*, 2012).

In the case of neutrophils, *SPI1* was also upregulated, but no change was observed in DNA methylation following HU treatment, suggesting that other mechanisms are involved in the regulation of *SPI1* expression. Indeed, SPI1 is essential for neutrophil maturation, and its loss impairs their ability to kill pathogens (Fischer *et al.*, 2019). Although SPI1 is widely known as an activator transcription factor, repressive functions have also been described. Accordingly, SPI1 is able to interact with the histone deacetylate 1 (HDAC1) protein which results in decreased acetylation of histone H3 at Lys27 (H3K27ac) (Fischer *et al.*, 2019). This mechanism is thought to repress genes involved in positive regulation of the immune system and cellular metabolic processes, which are known to be over-expressed during pathogen infection in SPI1 null neutrophils (Fischer *et al.*, 2019). These findings suggest that SPI1 is able to modulate the epigenome of neutrophils to repress genes that could lead to an over-reactive immune response (Fischer *et al.*, 2019).

Taken together, these investigations support the role of SPI1 in abnormal haematopoiesis, and, accordingly, HU could prove to be an important factor in restoring *SPI1* expression. At the stem cell level, SPI1 would be inducing differentiation instead of self-renewal, while in neutrophils, it would repress genes which might be involved in the pro-inflammatory process. To assist in elucidating the effect of HU on *SPI1* expression, and its interplay with DNA methylation at the stem cell level, a more detailed analysis might be carried out, by performing a time-course study of HU treatment to observe how DNA methylation and gene expression change over time, which occurs first, and what other mechanisms are involved. In neutrophils, it

would be interesting to correlate the DEGs that are targeted by SPI1 following HU treatment with the dataset of Fischer *et al.* (2019), to determine whether the genes affected are involved in the pro-inflammatory pathways regulated by SPI1.

5.5 Another possible mechanism involved in HU therapeutic effect

Aberrant hypermethylation has been described in several cancer cells and mutations that affect epigenetic regulators such as *TET2* and *IDH1/2* have also being observed in haematological disorders. Investigations of DNA methylation in cancer cells have also led to interesting findings, such as hypermethylation of the *BRCA1* promoter in breast cancer (Esteller *et al.*, 2000). However, this is not always the case and several DEGs identified in neutrophils and $CD34^+$ cells did not correlate with DNA methylation. As mentioned previously, this indicates that other mechanisms are involved in the regulation of those DEGs.

Studies have demonstrated that the majority of hypermethylated CGIs located at gene promoters identified in cancer are normally not expressed in healthy cells from the same individual (Sproul *et al.*, 2012; Hinoue *et al.*, 2012). This suggests that other factors, independent of DNA methylation, contribute to the deregulated gene expression in disease. For instance, a mechanism involving an altered redistribution of the repressive histone mark H3K27me3, due to deregulation of the polycomb group proteins, has been observed, which produces gene silencing irrespectively of promoter methylation (Kondo *et al.*, 2008; Court *et al.*, 2019). In terms of MPN, a non-canonical function of JAK2 has been described. JAK2 is able to localize to the nucleus and phosphorylates the histone H3 which blocks the binding of the heterochromatin protein 1α (HP1α) that promotes transcriptional silencing (Dawson *et al.*, 2009). Furthermore, JAK2V617F strongly binds and phosphorylates PRMT5, an arginine methyltransferase, impairing PRMT5 ability to methylate histones H2A

and H4, resulting in expansion of HSCs and erythroid differentiation (Liu *et al.*, 2011). Therefore, the constitutive kinase activity of JAK2V617F results in epigenetic aberrations that contribute to the MPN phenotype. Interestingly, in this study, upregulation of *SPI1* was observed, and its protein has been shown to interact with the H3K27ac machinery, as previously mentioned. Moreover, several genes with roles in histone modification in human or mouse were affected by HU treatment. One example is *KDM5B*, which has been shown to be downregulated following HU treatment. KDM5B protein is able to demethylate tri-, di- and monomethylated lysine 4 of histone H3 and its over-expression has been observed in cancer cells (Wang *et al.*, 2016). It is therefore proposed that histone modification may be an interesting epigenetic mechanism worthy of further investigation in the context of HU treatment in MPN.

5.6 Limitations and future perspectives

- It could be argued that the effects observed in this study may be driven by the ability of HU to selectively reduce JAK2V617F mutant cells in preference to wild-type cells. However, evidence for this is contradictory. Some studies have demonstrated a decrease of the JAK2V617F allele burden, but others have observed the opposite (Antonioli *et al.*, 2010; Spanoudakis *et al.*, 2009; Ricksten *et al.*, 2008). Therefore it is recommended that the JAK2V617F allele burden in the patient samples from this study be assessed, in order to identify confounding factors that could lead to misinterpretation of the results.

- Another important omission, when assessing the patient samples, is that they were not screened for mutations in the *TET2*, *IDH1/1* and *DNMT3A* genes. Although the mutation rate is low in MPN patients, and most of the changes in DNA methylation observed here involved loss of methylation, this information could help to confirm the robustness of the findings, and that they were linked

to HU treatment.

- DNA methylation from mouse LSK samples could also constitute another source of important information. To bypass the problem of low DNA concentration, locus-specific analysis could be carried out on selected candidate genes from the human study (Fig 5.1). This methods usually employs less input material and standardization can be performed in a short period of time.

- Further validation of the genes identified here should also be performed using a larger cohort of patients. A list of candidate genes to the response of HU treatment and their description can be found in Table 5.1 (Fig 5.1).

- A more mechanistic analysis of the *SPI1* locus would reveal how HU is able to modulate DNA methylation and gene expression. Moreover, this would constitute an interesting model for an understanding of the interplay between gene expression and DNA methylation.

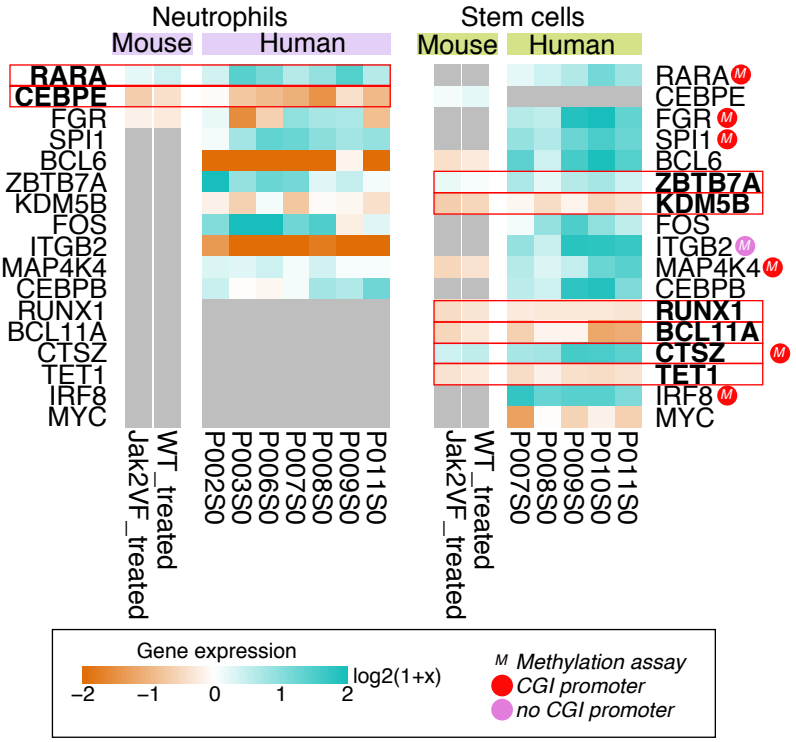

Figure. 5.1. Selected candidate genes to the response of HU treatment for further validation analyses. Heatmap plot of log2-fold change of DEGs of patients and mice group comparison (untreated vs treated). Genes that need methylation assays are represented with an "M". Genes shared between human and mouse are indicated in bold. CGI: CpG island.

Table. 5.1. Description of the selected candidate genes to the response of HU treatment for further validation analyses

Symbol	Description
RARA	Encodes transcription factor involved in neutrophils maturation. It can promote granulopoiesis in pluripotent progenitor cells (Kastner and Chan, 2001). Promoter hypomethylated and upregulated in $CD34^{+}$cells. Differentially expressed in neutrophils.
CEBPE	Encodes transcription factor that is required for the terminal differentiation of neutrophils (Bedi *et al.*, 2009). Gain of function mutation cause inflammatory disease (Göös *et al.*, 2019). Differentially expressed in neutrophils.
FGR	Encodes tyrosine kinase involved in immune responses, neutrophil functions. Acts downstream of ITGB1 and ITGB2 (Gutkind and Robbins, 1989). Promoter hypomethylated and upregulated in $CD34^{+}$cells.
SPI1	Encodes transcription factor required for neutrophil maturation and induce differentiation into myeloid lineage at stem cell level (Fischer *et al.*, 2019; Chen *et al.*, 1995; Dakic *et al.*, 2005). Promoter hypomethylated and upregulated in $CD34^{+}$cells. Upregulated in human neutrophils.
BCL6	Encodes transcription factor that acts mainly as a repressor. Suppresses macrophage proliferation (Yu *et al.*, 2005). Top differentially expressed gene in human neutrophils.
ZBTB7A	Encodes transcription factor involved in foetal haemoglobin synthesis (Chondrou *et al.*, 2018). Binds to intragenic CpG island of *F2RL1* gene. Upregulated in both species in stem cells.
KDM5B	Encodes a lysine-specific histone demethylase (Zhang *et al.*, 2014). Upregulated in both species in stem cells.
FOS	FOS proteins have been implicated as regulators of cell proliferation, differentiation, and transformation. Downregulated in proliferative HSCs(McKinney-Freeman *et al.*, 2012).
ITGB2	The encoded protein plays an important role in immune response and defects in this gene cause leukocyte adhesion deficiency. Acts upstream of FGR. Promoter hypermethylated and upregulated in $CD34^{+}$cells.
MAP4K4	Encodes serine/threonine kinase. Study in T cells of type 2 diabetes patients have shown hypomethylation of CpG island in promoter (Chuang *et al.*, 2016). Promoter hypomethylated and upregulated in $CD34^{+}$cells.
CEBPB	Encodes transcription factor involved in immune and inflammatory responses (Screpanti *et al.*, 1995). Transcription factor bound to differentially methylated CpG sites in $CD34^{+}$cells.
RUNX1	Encodes transcription factor essential for definitive haematopoiesis. Interacts with SPI1 (Goyal *et al.*, 2017; Imperato *et al.*, 2015). Downregulated in both species in stem cells.
BCL11A	Encodes transcription factor involved in foetal haemoglobin synthesis (Grieco *et al.*, 2015). Downregulated in both species in stem cells.
CTSZ	It exhibits both carboxy-monopeptidase and carboxy-dipeptidase activities. Promoter hypomethylated and upregulated. Upregulated in both species in stem cells.
TET1	Encodes protein with demethylase activity (Tahiliani *et al.*, 2009). Downregulated in both species in stem cells.
IRF8	Involved in self-renewal of HSCs (Qiu *et al.*, 2015). Involved in the macrophage versus granulocyte differentiation with SPI1. (Kurotaki *et al.*, 2014, 2018)
MYC	Encodes transcription factor necessary for balance of self-renewal and differentiation of haematopoietic stem ells (Wilson *et al.*, 2004).

Concluding remarks

- HU affects gene expression at the stem cell and neutrophil level, in both human and mouse.
- The Jak2V617F-knock-in mouse completely recapitulated the MPN phenotype with and without HU treatment.
- In addition to the well-documented effect of HU on cell cycle inhibition, HU treatment in the Jak2V617F-knock-in mouse was shown to reverse over-expressed cytokine signalling.
- HU has a distinct effect on DNA methylation in cells at different stages of differentiation, which is much stronger at the stem cell level.
- Both PV and ET patients exhibit aberrant DNA methylation in the *SPI1* gene, which can contribute to the disease state.
- The *SPI1* gene is a strong candidate to study to facilitate a better understanding of the HU effect through DNA methylation, which may lead to its therapeutic effect.
- Other DNA methylation-independent mechanisms are involved in the regulation of gene expression following HU treatment.
- Several genes enconding transcription factors affected in human and mouse, such as *BCL11A*, *ZBTB7A* and *RUNX1*, are proposed as potential targets for the effects of HU treatment.

Bibliography

Hajime Akada, Dongqing Yan, Haiying Zou, Steven Fiering, Robert E. Hutchison, and M. Golam Mohi. Conditional expression of heterozygous or homozygous Jak2V617F from its endogenous promoter induces a polycythemia vera-like disease. *Blood*, 115(17):3589–3597, 2010. ISSN 00064971. doi: 10.1182/blood-2009-04-215848.

Altuna Akalin, Francine E. Garrett-Bakelman, Matthias Kormaksson, Jennifer Busuttil, Lu Zhang, Irina Khrebtukova, Thomas A. Milne, Yongsheng Huang, Debabrata Biswas, Jay L. Hess, C. David Allis, Robert G. Roeder, Peter J.M. Valk, Bob Löwenberg, Ruud Delwel, Hugo F. Fernandez, Elisabeth Paietta, Martin S. Tallman, Gary P. Schroth, Christopher E. Mason, Ari Melnick, and Maria E. Figueroa. Base-pair resolution DNA methylation sequencing reveals profoundly divergent epigenetic landscapes in acute myeloid leukemia. *PLoS Genet.*, 8(6), 2012. ISSN 15537390. doi: 10.1371/journal.pgen.1002781.

Koichi Akashi, David Traver, Toshihiro Miyamoto, Irving L. Weissman, Akashi K, Traver D, Miyamoto T, and Weissman IL. A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. *Nature*, 404(6774):193–197, 2000. ISSN 00280836. doi: 10.1038/35004599.

Marta Albajar, Pilar Gutierrez, Carlos Richard, Manuel Rosa-Garrido, M. Teresa Gómez-Casares, Juan L. Steegmann, Javier León, and M. Dolores Delgado. PU.1

expression is restored upon treatment of chronic myeloid leukemia patients. *Cancer Lett.*, 270(2):328–336, 2008. ISSN 03043835. doi: 10.1016/j.canlet.2008.05.024.

Alberto Alvarez-Larrán, Luz Martínez-Avilés, Juan Carlos Hernández-Boluda, Francisca Ferrer-Marín, María Luisa Antelo, Carmen Burgaleta, M. Isabel Mata, Blanca Xicoy, Alejandra Martínez-Trillos, M. Teresa Gómez-Casares, M. Antonia Durán, Bárbara Marcote, Agueda Ancochea, Alicia Senín, Anna Angona, Montse Gómez, Vicente Vicente, Francisco Cervantes, Beatriz Bellosillo, and Carles Besses. Busulfan in patients with polycythemia vera or essential thrombocythemia refractory or intolerant to hydroxyurea. *Ann. Hematol.*, 93(12):2037–2043, dec 2014. ISSN 0939-5555. doi: 10.1007/s00277-014-2152-7.

Jennifer Antonchuk, Guy Sauvageau, and R. Keith Humphries. HOXB4-induced expansion of adult hematopoietic stem cells ex vivo. *Cell*, 109(1):39–45, 2002. ISSN 00928674. doi: 10.1016/S0092-8674(02)00697-9.

Elisabetta Antonioli, Alessandra Carobbio, Lisa Pieri, Alessandro Pancrazzi, Paola Guglielmelli, Federica Delaini, Vanessa Ponziani, Niccolò Bartalucci, Lorenzo Tozzi, Alberto Bosi, Alessandro Rambaldi, Tiziano Barbui, and Alessandro M. Vannucchi. Hydroxyurea does not appreciably reduce JAK2 v617F allele burden in patients with polycythemia vera or essential thrombocythemia. *Haematologica*, 95(8):1435–1438, 2010. ISSN 03906078. doi: 10.3324/haematol.2009.021444.

D.A. Arber, A. Orazi, R. Hasserjian, M.J. Borowitz, M.M. Beau Le, C.D. Bloomfield, M. Cazzola, and J.W. Vardiman. The 2016 revision to the World Health Organization classi fi cation of myeloid neoplasms and acute leukemia. *Blood*, 127 (20):2391–2406, 2016. ISSN 1528-0020. doi: 10.1182/blood-2016-03-643544.

William P Arnold, Chandra K Mittal, Shoji Katsuki, and F. Murad. Nitric oxide activates guanylate cyclase and increases guanosine 3':5'-cyclic monophosphate

levels in various tissue preparations. *Proc. Natl. Acad. Sci.*, 74(8):3203–3207, aug 1977. ISSN 0027-8424. doi: 10.1073/pnas.74.8.3203.

Taruna Arora, Bin Liu, Hongchin He, Jenny Kim, Theresa L. Murphy, Kenneth M. Murphy, Robert L. Modlin, and Ke Shuai. PIASx is a transcriptional co-repressor of signal transducer and activator of transcription 4. *J. Biol. Chem.*, 278(24): 21327–21330, 2003. ISSN 00219258. doi: 10.1074/jbc.C300119200.

Martin J Aryee, Andrew E Jaffe, Hector Corrada-Bravo, Christine Ladd-Acosta, Andrew P Feinberg, Kasper D Hansen, and Rafael A Irizarry. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics*, 30(10):1363–1369, may 2014. ISSN 1367-4811 (Electronic). doi: 10.1093/bioinformatics/btu049.

Yassen Assenov, Fabian Müller, Pavlo Lutsik, Jörn Walter, Thomas Lengauer, and Christoph Bock. Comprehensive analysis of DNA methylation data with RnBeads. *Nat Methods*, 11(11):1138–40, 2014. ISSN 1548-7105. doi: 10.1038/nmeth.3115.

Joanne L Attema, Peter Papathanasiou, E Camilla Forsberg, Jian Xu, Stephen T Smale, and Irving L Weissman. Epigenetic characterization of hematopoietic stem cell differentiation using miniChIP and bisulfite sequencing analysis. *Proc. Natl. Acad. Sci. U. S. A.*, 104(30):12371–12376, 2007. ISSN 00278424. doi: 10.1073/pnas.0704468104.

Chris M. Bacon, P. Justin Tortolani, Akihiro Shimosaka, Robert C. Rees, Dan L. Longo, and John J. O'Shea. Thrombopoietin (TPO) induces tyrosine phosphorylation and activation of STAT5 and STAT3. *FEBS Lett.*, 370(1-2):63–68, 1995. ISSN 00145793. doi: 10.1016/0014-5793(95)00796-C.

C. M. Baum, I. L. Weissman, A. S. Tsukamoto, A. M. Buckle, and B. Peault. Isolation of a candidate human hematopoietic stem-cell population. *Proc. Natl.*

Acad. Sci. U. S. A., 89(7):2804–2808, 1992. ISSN 00278424. doi: 10.1073/pnas.89.7.2804.

Richa Bedi, Jian Du, Arun K. Sharma, Ignatius Gomes, and Steven J. Ackerman. Human C/EBP-ϵ activator and repressor isoforms differentially reprogram myeloid lineage commitment and differentiation. *Blood*, 113(2):317–327, 2009. ISSN 00064971. doi: 10.1182/blood-2008-02-139741.

James G. Bender, L. Bik To, Stephanie Williams, and Lee S. Schwartzberg. Defining a Therapeutic Dose of Peripheral Blood Stem Cells. *J. Hematother.*, 1(4):329–341, jan 1992. ISSN 1061-6128. doi: 10.1089/scd.1.1992.1.329.

Shelley L Berger, Tony Kouzarides, Ramin Shiekhattar, and Ali Shilatifard. An operational definition of epigenetics. *Genes Dev.*, 23(7):781–783, 2009. ISSN 08909369. doi: 10.1101/gad.1787609.

P Bernasconi, M Boni, P M Cavigliano, S Calatroni, E Brusamolino, F Passamonti, G Volpe, A Pistorio, I Giardini, B Rocca, M Caresana, M Lazzarino, and C Bernasconi. Acute myeloid leukemia (AML) having evolved from essential thrombocythemia (ET): distinctive chromosome abnormalities in patients treated with pipobroman or hydroxyurea. *Leukemia*, 16(10):2078–83, 2002. ISSN 0887-6924. doi: 10.1038/sj.leu.2402638.

Adrian Bird. DNA methylation patterns and epigenetic memory. *Genes Dev.*, 16 (1):6–21, jan 2002. ISSN 08909369. doi: 10.1101/gad.947102.

Christoph Bock, Isabel Beerman, Wen Hui Lien, Zachary D. Smith, Hongcang Gu, Patrick Boyle, Andreas Gnirke, Elaine Fuchs, Derrick J. Rossi, and Alexander Meissner. DNA Methylation Dynamics during In Vivo Differentiation of Blood and Skin Stem Cells. *Mol. Cell*, 47(4):633–647, 2012. ISSN 10972765. doi: 10.1016/j.molcel.2012.06.019.

M N Boddy and P Russell. DNA replication checkpoint. *Curr. Biol.*, 11(23):R953–6, nov 2001. ISSN 0960-9822. doi: 10.2741/Boddy.

Nicola Bonadies, Thomas Pabst, and Beatrice U Mueller. Heterozygous deletion of the PU.1 locus in human AML. *Blood*, 115(2):331–4, jan 2010. ISSN 1528-0020. doi: 10.1182/blood-2009-03-212225.

Angela M. Boutté, David B. Friedman, Matthew Bogyo, Yongfen Min, Li Yang, and P. Charles Lin. Identification of a myeloid-derived suppressor cell cystatin-like protein that inhibits metastasis. *FASEB J.*, 25(8):2626–2637, 2011. ISSN 08926638. doi: 10.1096/fj.10-180604.

Nicolas L Bray, Harold Pimentel, Pall Melsted, and Lior Pachter. Near-optimal probabilistic RNA-seq quantification. *Nat Biotech*, 34(5):525–527, may 2016. ISSN 1087-0156.

Mandy Brecqueville, Jérôme Rey, François Bertucci, Emilie Coppin, Pascal Finetti, Nadine Carbuccia, Nathalie Cervera, Véronique Gelsi-Boyer, Christine Arnoulet, Olivier Gisserot, Denis Verrot, Borhane Slama, Norbert Vey, Marie-Joelle Mozziconacci, Daniel Birnbaum, and Anne Murati. Mutation analysis of ASXL1, CBL, DNMT3A, IDH1, IDH2, JAK2, MPL, NF1, SF3B1, SUZ12, and TET2 in myeloproliferative neoplasms. *Genes, Chromosom. Cancer*, 51(8):743–755, aug 2012. ISSN 10452257. doi: 10.1002/gcc.21960.

Hana Bruchova, Tereza Borovanova, Hana Klamova, and Radim Brdicka. Gene expression profiling in chronic myeloid leukemia patients treated with hydroxyurea. *Leuk. Lymphoma*, 43(6):1289–1295, 2002. ISSN 10428194. doi: 10.1080/10428190290026358.

B Brüne and V Ullrich. Inhibition of platelet aggregation by carbon monoxide is mediated by activation of guanylate cyclase. *Mol. Pharmacol.*, 32(4):497 LP – 504, oct 1987. ISSN 0026-895X.

Thomas G P Bumm, Collin Elsea, Amie S. Corbin, Marc Loriaux, Daniel Sherbenou, Lisa Wood, Jutta Deininger, Richard T. Silver, Brian J. Druker, and Michael W N Deininger. Characterization of murine JAK2V617F-positive myeloproliferative disease. *Cancer Res.*, 66(23):11156–11165, 2006. ISSN 00085472. doi: 10.1158/0008-5472.CAN-06-2210.

Edward R. Burns, L. Juden Reed, and Barry Wenz. Volumetric Erythrocyte Macrocytosis Induced by Hydroxyurea. *Am. J. Clin. Pathol.*, 85(3):337–341, mar 1986. ISSN 1943-7722. doi: 10.1093/ajcp/85.3.337.

C. M. Butcher, J. F. Hutton, U. Hahn, L. B. To, P. Bardy, I. Lewis, and R. J. D'Andrea. Cellular origin and lineage specificity of the JAK2V617F allele in polycythemia vera. *Blood*, 109(1):386–387, jan 2007. ISSN 0006-4971. doi: 10.1182/blood-2006-07-036426.

Nina Cabezas-Wallscheid, Daniel Klimmeck, Jenny Hansson, Daniel B Lipka, Alejandro Reyes, Qi Wang, Dieter Weichenhan, Amelie Lier, Lisa Von Paleske, Simon Renders, Peer Wünsche, Petra Zeisberger, David Brocks, Lei Gu, Carl Herrmann, Simon Haas, Marieke A.G. Essers, Benedikt Brors, Roland Eils, Wolfgang Huber, Michael D Milsom, Christoph Plass, Jeroen Krijgsveld, and Andreas Trumpp. Identification of regulatory networks in HSCs and their immediate progeny via integrated proteome, transcriptome, and DNA methylome analysis. *Cell Stem Cell*, 15(4):507–522, 2014. ISSN 18759777. doi: 10.1016/j.stem.2014.07.005.

N. Carbuccia, A. Murati, V. Trouplin, M. Brecqueville, J. Adélaïde, J. Rey, W. Vainchenker, O. A. Bernard, M. Chaffanet, N. Vey, D. Birnbaum, and M. J. Mozziconacci. Mutations of ASXL1 gene in myeloproliferative neoplasms. *Leukemia*, 23(11):2183–2186, 2009. ISSN 14765551. doi: 10.1038/leu.2009.141.

Grant a Challen, Deqiang Sun, Mira Jeong, Min Luo, Jaroslav Jelinek, Jonathan S Berg, Christoph Bock, Aparna Vasanthakumar, Hongcang Gu, Yuanxin Xi,

Shoudan Liang, Yue Lu, Gretchen J Darlington, Alexander Meissner, Jean-Pierre J Issa, Lucy a Godley, Wei Li, and Margaret a Goodell. Dnmt3a is essential for hematopoietic stem cell differentiation. *Nat. Genet.*, 44(1):23–31, 2011. ISSN 1061-4036. doi: 10.1038/ng.1009.

Grant A. Challen, Deqiang Sun, Allison Mayle, Mira Jeong, Min Luo, Benjamin Rodriguez, Cates Mallaney, Hamza Celik, Liubin Yang, Zheng Xia, Sean Cullen, Jonathan Berg, Yayun Zheng, Gretchen J. Darlington, Wei Li, and Margaret A. Goodell. Dnmt3a and Dnmt3b have overlapping and distinct functions in hematopoietic stem cells. *Cell Stem Cell*, 15(3):350–364, 2014. ISSN 18759777. doi: 10.1016/j.stem.2014.06.018.

Kai-Hsin Chang, Timothy Sullivan, Mei Liu, Xiao Yang, Chao Sun, Benjamin Vieira, Ming Zhang, Vu P Hong, Kai Chen, Sarah Smith, Siyuan Tan, Andreas Reik, Fyodor D Urnov, Edward J Rebar, Olivier Danos, and Haiyan Jiang. Clonal Analysis of Human Bone Marrow CD34+ Cells Edited By BCL11A-Targeting Zinc Finger Nucleases Reveals Clinically Relevant Levels of Fetal Globin Expression in Edited Erythroid Progeny. *Blood*, 126(23):3234, 2015. ISSN 0006-4971.

Konika Chawla, Sushil Tripathi, Liv Thommesen, Astrid Lægreid, and Martin Kuiper. TFcheckpoint: A curated compendium of specific DNA-binding RNA polymerase II transcription factors. *Bioinformatics*, 29(19):2519–2520, 2013. ISSN 14602059. doi: 10.1093/bioinformatics/btt432.

Edwin Chen and Ann Mullally. How does JAK2V617F contribute to the pathogenesis of myeloproliferative neoplasms? *Hematology*, 2014(1):268–276, 2014. ISSN 15204383. doi: 10.1182/asheducation-2014.1.268.

Edwin Chen, Rebekka K. Schneider, Lawrence J. Breyfogle, Emily A. Rosen, Luke Poveromo, Shannon Elf, Amy Ko, Kristina Brumme, Ross Levine, Benjamin L. Ebert, and Ann Mullally. Distinct effects of concomitant Jak2V617F expression

and Tet2 loss in mice promote disease progression in myeloproliferative neoplasms. *Blood*, 125(2):327–335, 2015. ISSN 15280020. doi: 10.1182/blood-2014-04-567024.

H. M. Chen, P. Zhang, M. T. Voso, S. Hohaus, D. A. Gonzalez, C. K. Glass, D. E. Zhang -, and D. G. Tenen. Neutrophils and monocytes express high levels of PU.1 (Spi-1) but not Spi- B. *Blood*, 85(10):2918–2928, 1995. ISSN 00064971.

Vasiliki Chondrou, Eleana F. Stavrou, Georgios Markopoulos, Alexandra Kouraklis-Symeonidis, Vasilios Fotopoulos, Argiris Symeonidis, Efthymia Vlachaki, Panagiota Chalkia, George P. Patrinos, Adamantia Papachatzopoulou, and Argyro Sgourou. Impact of ZBTB7A hypomethylation and expression patterns on treatment response to hydroxyurea. *Hum. Genomics*, 12(1):1–14, 2018. ISSN 14797364. doi: 10.1186/s40246-018-0177-z.

Judith K. Christman. 5-Azacytidine and 5-aza-2'-deoxycytidine as inhibitors of DNA methylation: Mechanistic studies and their implications for cancer therapy. *Oncogene*, 21(35 REV. ISS. 3):5483–5495, 2002. ISSN 09509232. doi: 10.1038/sj.onc.1205699.

Huai Chia Chuang, Jun Sing Wang, I. Te Lee, Wayne H.H. Sheu, and Tse Hua Tan. Epigenetic regulation of HGK/MAP4K4 in T cells of type 2 diabetes patients. *Oncotarget*, 7(10):10976–10989, 2016. ISSN 19492553. doi: 10.18632/oncotarget.7686.

Chan D Chung, Jiayu Liao, Bin Liu, Xiaoping Rao, Philippe Jay, Philippe Berta, and Ke Shuai. Specific Inhibition of Stat3 Signal Transduction by PIAS3. *Science (80-.).*, 278(5344):1803–1805, 1997. ISSN 0036-8075. doi: 10.1126/science.278.5344.1803.

Aldo Ciau-Uitz, Roger Patient, and Alexander Medvinsky. Ontogeny of the Hematopoietic System. In *Encycl. Immunobiol.*, volume 1, pages 1–14. 2016. ISBN 9780080921525. doi: 10.1016/B978-0-12-374279-7.01002-X.

Curt I Civin, Lewis C Strauss, Charlotte Brovall, M J Fackler, Jill F Schwartz, and Joel H Shaper. Antigenic analysis of hematopoiesis. III. A hematopoietic progenitor cell surface antigen defined by a monoclonal antibody raised against KG-1a cells. *J. Immunol.*, 133(1):157–65, jul 1984. ISSN 0022-1767.

Vladan P. Cokic, Bojana B. Beleslin-Cokic, Melanija Tomic, Stanko S. Stojilkovic, Constance T. Noguchi, and Alan N. Schechter. Hydroxyurea induces the eNOS-cGMP pathway in endothelial cells. *Blood*, 108(1):184–191, 2006. ISSN 00064971. doi: 10.1182/blood-2005-11-4454.

Sergio Cortelazzo, Guido Finazzi, Marco Ruggeri, Oscar Vestri, Monica Galli, Francesco Rodeghiero, and Tiziano Barbui. Hydroxyurea for patients with essential thrombocythemia and a high risk of thrombosis. *N. Engl. J. Med.*, 332 (17):1132–6, apr 1995. ISSN 0028-4793. doi: 10.1056/NEJM199504273321704.

Franck Court, Elisa Le Boiteux, Anne Fogli, Mélanie Müller-Barthélémy, Catherine Vaurs-Barrière, Emmanuel Chautard, Bruno Pereira, Julian Biau, Jean-Louis Kemeny, Toufic Khalil, Lucie Karayan-Tapon, Pierre Verrelle, and Philippe Arnaud. Transcriptional alterations in glioma result primarily from DNA methylation-independent mechanisms. *Genome Res.*, page 516997, sep 2019. ISSN 1088-9051. doi: 10.1101/gr.249219.119.

Richard Dahl, Jonathan C. Walsh, David Lancki, Peter Laslo, Sangeeta R. Iyer, Harinder Singh, and M. Celeste Simon. Regulation of macrophage and neutrophil cell fates by the PU.1:C/EBPα ratio and granulocyte colony-stimulating factor. *Nat. Immunol.*, 4(10):1029–1036, 2003. ISSN 15292908. doi: 10.1038/ni973.

Aleksandar Dakic, Donald Metcalf, Ladina Di Rago, Sandra Mifsud, Li Wu, and Stephen L. Nutt. PU.1 regulates the commitment of adult hematopoietic progenitors and restricts granulopoiesis. *J. Exp. Med.*, 201(9):1487–1502, 2005. ISSN 00221007. doi: 10.1084/jem.20050075.

Mark A Dawson, Andrew J Bannister, Berthold Göttgens, Samuel D Foster, Till Bartke, Anthony R Green, and Tony Kouzarides. JAK2 phosphorylates histone H3Y41 and excludes HP1alpha from chromatin. *Nature*, 461(7265):819–22, 2009. ISSN 1476-4687. doi: 10.1038/nature08448.

M Deaton, Shaun Webb, Alastair R W Kerr, Robert S Illingworth, Jacky Guy, Robert Andrews, and Adrian Bird. Cell type - specific DNA methylation at intragenic CpG islands in the immune system. *Genome Researh*, 21(7):1074–1086, 2011. ISSN 1549-5469. doi: 10.1101/gr.118703.110.Freely.

Rodney P. DeKoter and Harinder Singh. Regulation of B lymphocyte and macrophage development by graded expression of PU.1. *Science (80-.).*, 288 (5470):1439–1442, 2000. ISSN 00368075. doi: 10.1126/science.288.5470.1439.

Nicolas Delhomme, Niklas Mahler, Bastian Schiffthaler, David Sundell, Chanaka Mannapperuma, Torgeir R. Hvidsten, and Nathaniel R. Street. Guidelines for RNA-Seq data analysis (prot 67) | Bioinformatics. pages 1–24, 2015.

François Delhommeau, Sabrina Dupont, Véronique Della Valle, Chloé James, Severine Trannoy, Aline Massé, Olivier Kosmider, Jean-Pierre Le Couedic, Fabienne Robert, Antonio Alberdi, Yann Lécluse, Isabelle Plo, François J. Dreyfus, Christophe Marzac, Nicole Casadevall, Catherine Lacombe, Serge P. Romana, Philippe Dessen, Jean Soulier, Franck Viguié, Michaela Fontenay, William Vainchenker, and Olivier A. Bernard. Mutation in TET2 in Myeloid Cancers. *N. Engl. J. Med.*, 360(22):2289–2301, may 2009. ISSN 0028-4793. doi: 10.1056/NEJMoa0810069.

Alexander Dobin, Carrie A. Davis, Felix Schlesinger, Jorg Drenkow, Chris Zaleski, Sonali Jha, Philippe Batut, Mark Chaisson, and Thomas R. Gingeras. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, 29(1):15–21, 2013. ISSN 13674803. doi: 10.1093/bioinformatics/bts635.

Diana C. Doeing, Jessica L. Borowicz, and Elahé T. Crockett. Gender dimorphism in differential peripheral blood leukocyte counts in mice using cardiac, tail, foot, and saphenous vein puncture methods. *BMC Clin. Pathol.*, 3:1–6, 2003. ISSN 14726890. doi: 10.1186/1472-6890-3-1.

Sergei Doulatov, Faiyaz Notta, Elisa Laurenti, and John E. Dick. Hematopoiesis: A human perspective. *Cell Stem Cell*, 10(2):120–136, 2012. ISSN 19345909. doi: 10.1016/j.stem.2012.01.006.

Pankaj Dwivedi and Kenneth D. Greis. Granulocyte colony-stimulating factor receptor signaling in severe congenital neutropenia, chronic neutrophilic leukemia, and related malignancies. *Exp. Hematol.*, 46:9–20, 2017. ISSN 18732399. doi: 10.1016/j.exphem.2016.10.008.

Elaine Dzierzak and Nancy A Speck. Of lineage and legacy: the development of mammalian hematopoietic stem cells. *Nat. Immunol.*, 9(2):129–136, feb 2008. ISSN 1529-2916 (Electronic). doi: 10.1038/ni1560.

Carol A. Edwards and Anne C. Ferguson-Smith. Mechanisms regulating imprinted genes in clusters. *Curr. Opin. Cell Biol.*, 19(3):281–289, 2007. ISSN 09550674. doi: 10.1016/j.ceb.2007.04.013.

Takaho A. Endo, Masaaki Masuhara, Masahiro Yokouchi, Ritsu Suzuki, Hiroshi Sakamoto, Kaoru Mitsui, Akira Matsumoto, Shyu Tanimura, Motoaki Ohtsubo, Hiroyuki Misawa, Tadaaki Miyazaki, Nogueira Leonor, Tadatsugu Taniguchi, Takashi Fujita, Yuzuru Kanakura, Seturo Komiya, and Akihiko Yoshimura. A new protein containing an SH2 domain that inhibits JAK kinases. *Nature*, 387 (6636):921–924, 1997. ISSN 00280836. doi: 10.1038/43213.

K. Gunnar Engström and Eva Löfvenberg. Treatment of myeloproliferative disorders with hydroxyurea effects on red blood cell geometry and deformability. *Blood*, 91 (10):3986–3991, 1998. ISSN 0006-4971.

Thomas Ernst, Andrew J Chase, Joannah Score, Claire E Hidalgo-Curtis, Catherine Bryant, Amy V Jones, Katherine Waghorn, Katerina Zoi, Fiona M Ross, Andreas Reiter, Andreas Hochhaus, Hans G Drexler, Andrew Duncombe, Francisco Cervantes, David Oscier, Jacqueline Boultwood, Francis H Grand, and Nicholas C P Cross. Inactivating mutations of the histone methyltransferase gene EZH2 in myeloid disorders. *Nat. Genet.*, 42(8):722–726, 2010. ISSN 1061-4036. doi: 10.1038/ng.621.

M. Esteller, J M Silva, G Dominguez, F Bonilla, X Matias-Guiu, E Lerma, E Bussaglia, J Prat, I C Harkes, E A Repasky, E Gabrielson, M Schutte, S B Baylin, and J G Herman. Promoter hypermethylation and BRCA1 inactivation in sporadic breast and ovarian tumors. *J. Natl. Cancer Inst.*, 92(7):564–9, apr 2000. ISSN 0027-8874. doi: 10.1093/jnci/92.7.564.

Philip Ewels, Måns Magnusson, Sverker Lundin, and Max Käller. MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, 32(19):3047–3048, 2016. ISSN 14602059. doi: 10.1093/bioinformatics/btw354.

Antonio Fabregat, Konstantinos Sidiropoulos, Guilherme Viteri, Oscar Forner, Pablo Marin-Garcia, Vicente Arnau, Peter D'Eustachio, Lincoln Stein, and Henning Hermjakob. Reactome pathway analysis: a high-performance in-memory approach. *BMC Bioinformatics*, 18(1):142, 2017. ISSN 1471-2105. doi: 10.1186/s12859-017-1559-2.

Matthias Farlik, Florian Halbritter, Fabian Muller, Fizzah A. Choudry, Peter Ebert, Johanna Klughammer, Samantha Farrow, Antonella Santoro, Valerio Ciaurro, Anthony Mathur, Rakesh Uppal, Hendrik G. Stunnenberg, Willem H. Ouwehand, Elisa Laurenti, Thomas Lengauer, Mattia Frontini, Christoph Bock, Fabian Müller, Fizzah A. Choudry, Peter Ebert, Johanna Klughammer, Samantha Farrow, Antonella Santoro, Valerio Ciaurro, Anthony Mathur, Rakesh Uppal, Hendrik G. Stunnenberg, Willem H. Ouwehand, Elisa Laurenti, Thomas Lengauer,

Mattia Frontini, Christoph Bock, Fabian Muller, Fizzah A. Choudry, Peter Ebert, Johanna Klughammer, Samantha Farrow, Antonella Santoro, Valerio Ciaurro, Anthony Mathur, Rakesh Uppal, Hendrik G. Stunnenberg, Willem H. Ouwehand, Elisa Laurenti, Thomas Lengauer, Mattia Frontini, and Christoph Bock. DNA Methylation Dynamics of Human Hematopoietic Stem Cell Differentiation. *Cell Stem Cell*, 19(6):808–822, 2016. ISSN 18759777. doi: 10.1016/j.stem.2016.10.019.

Francesca Ficara, Mark J. Murphy, Min Lin, and Michael L. Cleary. Pbx1 Regulates Self-Renewal of Long-Term Hematopoietic Stem Cells by Maintaining Their Quiescence. *Cell Stem Cell*, 2(5):484–496, 2008. ISSN 19345909. doi: 10.1016/j.stem.2008.03.004.

Josephine Fischer, Carolin Walter, Alexander Tönges, Hanna Aleth, Marta Joana Costa Jordão, Mathias Leddin, Verena Gröning, Tabea Erdmann, Georg Lenz, Johannes Roth, Thomas Vogl, Marco Prinz, Martin Dugas, Ilse D. Jacobsen, and Frank Rosenbauer. Safeguard function of PU.1 shapes the inflammatory epigenome of neutrophils. *Nat. Immunol.*, 20(5):546–558, 2019. ISSN 15292916. doi: 10.1038/s41590-019-0343-z.

William N. Fishbein, Paul P. Carbone, Emil J. Freireich, Dwijendra Misra, and Emil Frei. Clinical trials of hydroxyurea in patients with cancer and leukemia. *Clin. Pharmacol. Ther.*, 5(5):574–580, sep 1964. ISSN 00099236. doi: 10.1002/cpt196455574.

Sanjeev A. Francis, Xun Shen, Jeffrey B. Young, Prashant Kaul, and Daniel J. Lerner. Rho GEF Lsc is required for normal polarization, migration, and adhesion of formyl-peptide-stimulated neutrophils. *Blood*, 107(4):1627–1635, 2006. ISSN 00064971. doi: 10.1182/blood-2005-03-1164.

Henrik Frederiksen, Waleed Ghanima, and Christina Roaldsnes. Myeloproliferative

neoplasms : trends in incidence , prevalence and survival in Norway. 98:85–93, 2016. doi: 10.1111/ejh.12788.

Jenna L. Galloway and Leonard I. Zon. Ontogeny of hematopoiesis: examining the emergence of hematopoietic cells in the vertebrate embryo. *Curr. Top. Dev. Biol.*, 53:139–58, 2003. ISSN 0070-2153.

Sheley Gambero, Andreia A. Canalli, Fabiola Traina, Dulcinéia M. Albuquerque, Sara T.O. Saad, Fernando F. Costa, and Nicola Conran. Therapy with hydroxyurea is associated with reduced adhesion molecule gene and protein expression in sickle red cells with a concomitant reduction in adhesive properties. *Eur. J. Haematol.*, 78(2):144–151, 2007. ISSN 09024441. doi: 10.1111/j.1600-0609.2006.00788.x.

M. Gardiner-Garden and M. Frommer. CpG Islands in vertebrate genomes. *J. Mol. Biol.*, 196(2):261–282, jul 1987. ISSN 00222836. doi: 10.1016/0022-2836(87)90689-9.

Pantelis Georgiades, Sarah Ogilvy, Hélène Duval, Diana R. Licence, D. Stephen Charnock-Jones, Stephen K. Smith, and Cristin G. Print. vavCre transgenic mice: A tool for mutagenesis in hematopoietic and endothelial lineages. *Genesis*, 34(4):251–256, 2002. ISSN 1526954X. doi: 10.1002/gene.10161.

Christopher J. Gibson, R. Coleman Lindsley, Vatche Tchekmedyian, Brenton G. Mar, Jiantao Shi, Alysia Bosworth, Liton Francisco, Jianbo He, Anita Bansal, Elizabeth A. Morgan, Ann S. Lacasce, S Arnold, David C. Fisher, Eric Jacobsen, Philippe Armand, Edwin P. Alyea, John Koreth, Vincent Ho, J Robert, Joseph H. Antin, Jerome Ritz, Sarah Nikiforow, Stephen J. Forman, Franziska Michor, Donna Neuberg, Ravi Bhatia, Smita Bhatia, Benjamin L. Ebert, Siddhartha Jaiswal, Alysia Bosworth, Liton Francisco, Jianbo He, Anita Bansal, Elizabeth A. Morgan, Ann S. Lacasce, Arnold S. Freedman, David C. Fisher, Eric Jacobsen,

Philippe Armand, Edwin P. Alyea, John Koreth, Vincent Ho, Robert J. Soiffer, Joseph H. Antin, Jerome Ritz, Sarah Nikiforow, Stephen J. Forman, Franziska Michor, Donna Neuberg, Ravi Bhatia, Smita Bhatia, and Benjamin L. Ebert. Clonal Hematopoiesis Associated With Adverse Outcomes After Autologous Stem-Cell Transplantation for Lymphoma. *J. Clin. Oncol.*, 35(14):JCO.2016.71.671, 2017. ISSN 0732-183X. doi: 10.1200/JCO.2016.71.6712.

Mark T. Gladwin and Alan N. Schechter. Nitric oxide therapy in sickle cell disease. *Semin. Hematol.*, 38(4):333–342, 2001. ISSN 00371963. doi: 10.1016/S0037-1963(01)90027-7.

Anna L. Godfrey, Edwin Chen, Francesca Pagano, Christina A. Ortmann, Yvonne Silber, Beatriz Bellosillo, Paola Guglielmelli, Claire N. Harrison, John T. Reilly, Frank Stegelmann, Fontanet Bijou, Eric Lippert, Mary F. McMullin, Jean Michel Boiron, Konstanze Döhner, Alessandro M. Vannucchi, Carlos Besses, Peter J. Campbell, and Anthony R. Green. JAK2V617F homozygosity arises commonly and recurrently in PV and ET, but PV is characterized by expansion of a dominant homozygous subclone. *Blood*, 120(13):2704–2707, 2012. ISSN 00064971. doi: 10.1182/blood-2012-05-431791.

Adrian F. Gombart, Scott H. Kwok, Karen L. Anderson, Yuji Yamaguchi, Bruce E. Torbett, and H. Phillip Koeffler. Regulation of neutrophil and eosinophil secondary granule gene expression by transcription factors C/EBPϵ and PU.1. *Blood*, 101 (8):3265–3273, 2003. ISSN 00064971. doi: 10.1182/blood-2002-04-1039.

Helka Göös, Christopher L. Fogarty, Biswajyoti Sahu, Vincent Plagnol, Kristiina Rajamäki, Katariina Nurmi, Xiaonan Liu, Elisabet Einarsdottir, Annukka Jouppila, Tom Pettersson, Helena Vihinen, Kaarel Krjutskov, Päivi Saavalainen, Asko Järvinen, Mari Muurinen, Dario Greco, Giovanni Scala, James Curtis, Dan Nordström, Robert Flaumenhaft, Outi Vaarala, Panu E. Kovanen, Salla Keskitalo, Annamari Ranki, Juha Kere, Markku Lehto, Luigi D. Notarangelo, Sergey Nejent-

sev, Kari K. Eklund, Markku Varjosalo, Jussi Taipale, and Mikko R.J. Seppänen. Gain-of-function CEBPE mutation causes noncanonical autoinflammatory inflammasomopathy. *J. Allergy Clin. Immunol.*, pages 1–13, 2019. ISSN 00916749. doi: 10.1016/j.jaci.2019.06.003.

Shubham Goyal, Takahiro Suzuki, Jing Ru Li, Shiori Maeda, Mami Kishima, Hajime Nishimura, Yuri Shimizu, and Harukazu Suzuki. RUNX1 induces DNA replication independent of active DNA demethylation at SPI1 regulatory regions. *BMC Mol. Biol.*, 18(1):1–7, 2017. ISSN 14712199. doi: 10.1186/s12867-017-0087-y.

Florian Grebien, Marc A Kerenyi, Boris Kovacic, Thomas Kolbe, Verena Becker, Helmut Dolznig, Klaus Pfeffer, U. Klingmuller, M. Muller, Hartmut Beug, E. W. Mullner, and R. Moriggl. Stat5 activation enables erythropoiesis in the absence of EpoR and Jak2. *Blood*, 111(9):4511–4522, may 2008. ISSN 0006-4971. doi: 10.1182/blood-2007-07-102848.

Christopher J. Greenhalgh, Donald Metcalf, Anne L. Thaus, Jason E. Corbin, Rachel Uren, Phillip O. Morgan, Louis J. Fabri, Jian Guo Zhang, Helene M. Martin, Tracy A. Willson, Nils Billestrup, Nicos A. Nicola, Manuel Baca, Warren S. Alexander, and Douglas J. Hilton. Biological evidence that SOCS-2 can act either as an enhancer or suppressor of growth hormone signaling. *J. Biol. Chem.*, 277(43):40181–40184, 2002. ISSN 00219258. doi: 10.1074/jbc.C200450200.

Amanda J. Grieco, Henny H. Billett, Nancy S. Green, M. Catherine Driscoll, and Eric E. Bouhassira. Variation in gamma-globin expression before and after induction with hydroxyurea associated with BCL11A, KLF1 and TAL1. *PLoS One*, 10 (6):1–14, 2015. ISSN 19326203. doi: 10.1371/journal.pone.0129431.

Stefan Gross, Rob A. Cairns, Mark D. Minden, Edward M. Driggers, Mark A. Bittinger, Hyun Gyung Jang, Masato Sasaki, Shengfang Jin, David P. Schenkein, Shinsan M. Su, Lenny Dang, Valeria R. Fantin, and Tak W. Mak. Cancer-

associated metabolite 2-hydroxyglutarate accumulates in acute myelogenous leukemia with isocitrate dehydrogenase 1 and 2 mutations. *J. Exp. Med.*, 207 (2):339–344, 2010. ISSN 15409538. doi: 10.1084/jem.20092506.

Chang Yun Gui, Chu Jiang, Heng Yue Xie, and Ruo Lan Qian. The apoptosis of HEL cells induced by hydroxyurea. *Cell Res.*, 7(1):91–97, jun 1997. ISSN 1001-0602. doi: 10.1038/cr.1997.10.

J. S. Gutkind and K. C. Robbins. Translocation of the FGR protein-tyrosine kinase as a consequence of neutrophil activation. *Proc. Natl. Acad. Sci. U. S. A.*, 86(22): 8783–8787, 1989. ISSN 00278424. doi: 10.1073/pnas.86.22.8783.

Claire N. Harrison, Peter J. Campbell, Georgina Buck, Keith Wheatley, Clare L. East, David Bareford, Bridget S. Wilkins, Jon D. van der Walt, John T. Reilly, Andrew P. Grigg, Paul Revell, Barrie E. Woodcock, Anthony R. Green, M R C Path, Peter J. Campbell, Georgina Buck, Keith Wheatley, D Phil, Clare L. East, David Bareford, Bridget S. Wilkins, F R C Path, Jon D. van der Walt, John T. Reilly, Andrew P. Grigg, B Sc, David Bareford, Bridget S. Wilkins, F R C Path, Jon D Van Der Walt, and F R C Path. Hydroxyurea compared with anagrelide in high-risk essential thrombocythemia. *N. Engl. J. Med.*, 353(1):33–45, jul 2005. ISSN 0028-4793. doi: 10.1056/NEJMoa043800.

Claire N. Harrison, David Bareford, Nauman Butt, Peter Campbell, Eibhlean Conneally, Mark Drummond, Wendy Erber, Tamara Everington, Anthony R. Green, Georgina W. Hall, Beverley J. Hunt, Christopher A. Ludlam, Richard Murrin, Catherine Nelson-Piercy, Deepti H. Radia, John T. Reilly, Jon Van Der Walt, Bridget Wilkins, and Mary F. McMullin. Guideline for investigation and management of adults and children presenting with a thrombocytosis. *Br. J. Haematol.*, 149(3):352–375, 2010. ISSN 00071048. doi: 10.1111/j.1365-2141.2010.08122.x.

Bryan Heit, Samantha Tavener, Eko Raharjo, and Paul Kubes. An intracellu-

lar signaling hierarchy determines direction of migration in opposing chemotactic gradients. *J. Cell Biol.*, 159(1):91–102, 2002. ISSN 00219525. doi: 10.1083/jcb.200202114.

Michael Heuser, Felicitas Thol, and Arnold Ganser. Clonal Hematopoiesis of Indeterminate Potential. *Dtsch. Arztebl. Int.*, 113(18):317–22, 2016. ISSN 1866-0452. doi: 10.3238/arztebl.2016.0317.

Andrés Hidalgo, Edwin R. Chilvers, Charlotte Summers, and Leo Koenderman. The Neutrophil Life Cycle. *Trends Immunol.*, 40(7):584–597, 2019. ISSN 14714981. doi: 10.1016/j.it.2019.04.013.

Toshinori Hinoue, Daniel J. Weisenberger, Christopher P.E. Lange, Hui Shen, Hyang Min Byun, David Van Den Berg, Simeen Malik, Fei Pan, Houtan Noushmehr, Cornelis M. Van Dijk, Rob A.E.M. Tollenaar, and Peter W. Laird. Genome-scale analysis of aberrant DNA methylation in colorectal cancer. *Genome Res.*, 22(2):271–282, 2012. ISSN 10889051. doi: 10.1101/gr.117523.110.

Hanno Hock, Melanie J. Hamblen, Heather M. Rooke, David Traver, Roderick T. Bronson, Scott Cameron, and Stuart H. Orkin. Intrinsic requirement for zinc finger transcription factor Gfi-1 in neutrophil differentiation. *Immunity*, 18(1): 109–120, 2003. ISSN 10747613. doi: 10.1016/S1074-7613(02)00501-0.

Hanno Hock, Melanie J Hamblen, Heather M Rooke, Jeffrey W Schindler, Shireen Saleque, Yuko Fujiwara, and Stuart H. Orkin. Gfi-1 restricts proliferation and preserves functional integrity of haematopoietic stem cells. *Nature*, 431(7011): 1002–1007, oct 2004. ISSN 0028-0836. doi: 10.1038/nature02994.

Gary C. Hon, Chun Xiao Song, Tingting Du, Fulai Jin, Siddarth Selvaraj, Ah Young Lee, Chia An Yen, Zhen Ye, Shi Qing Mao, Bang An Wang, Samantha Kuan, Lee E. Edsall, Boxuan Simen Zhao, Guo Liang Xu, Chuan He, and Bing Ren. 5mC

oxidation by Tet2 modulates enhancer activity and timing of transcriptome reprogramming during differentiation. *Mol. Cell*, 56(2):286–297, 2014. ISSN 10974164. doi: 10.1016/j.molcel.2014.08.026.

Kevin Huang and Guoping Fan. DNA methylation in cell differentiation and reprogramming: an emerging systematic view. *Regen. Med.*, 5:531–544, 2010. ISSN 1746-0751. doi: 10.2217/rme.10.35.

Motoshi Ichikawa, Takashi Asai, Toshiki Saito, Go Yamamoto, Sachiko Seo, Ieharu Yamazaki, Tetsuya Yamagata, Kinuko Mitani, Shigeru Chiba, Hisamaru Hirai, Seishi Ogawa, and Mineo Kurokawa. AML-1 is required for megakaryocytic maturation and lymphocytic differentiation, but not for maintenance of hematopoietic stem cells in adult hematopoiesis. *Nat. Med.*, 10(3):299–304, 2004. ISSN 10788956. doi: 10.1038/nm997.

P Ikonomi, C E Rivera, M Riordan, G Washington, A N Schechter, and C T Noguchi. Overexpression of GATA-2 inhibits erythroid and promotes megakaryocyte differentiation. *Exp. Hematol.*, 28(12):1423–1431, dec 2000. ISSN 0301-472X (Print).

Robert Illingworth, Alastair Kerr, Dina DeSousa, Helle Jørgensen, Peter Ellis, Jim Stalker, David Jackson, Chris Clee, Robert Plumb, Jane Rogers, Sean Humphray, Tony Cox, Cordelia Langford, and Adrian Bird. A novel CpG island set identifies tissue-specific methylation at developmental gene loci. *PLoS Biol.*, 6(1):0037–0051, 2008. ISSN 15449173. doi: 10.1371/journal.pbio.0060022.

Robert S. Illingworth and Adrian P. Bird. CpG islands - 'A rough guide'. *FEBS Lett.*, 583(11):1713–1720, 2009. ISSN 00145793. doi: 10.1016/j.febslet.2009.04.012.

Robert S. Illingworth, Ulrike Gruenewald-Schneider, Shaun Webb, Alastair R.W. Kerr, Keith D. James, Daniel J. Turner, Colin Smith, David J. Harrison, Robert Andrews, and Adrian P. Bird. Orphan CpG Islands Identify numerous conserved

promoters in the mammalian genome. *PLoS Genet.*, 6(9), 2010. ISSN 15537390. doi: 10.1371/journal.pgen.1001134.

Maria Rosaria Imperato, Pierre Cauchy, Nadine Obier, and Constanze Bonifer. The RUNX1-PU.1 axis in the control of hematopoiesis. *Int. J. Hematol.*, 101(4):319–329, 2015. ISSN 18653774. doi: 10.1007/s12185-015-1762-8.

Hana Imrichová, Gert Hulselmans, Zeynep Kalender Atak, Delphine Potier, and Stein Aerts. i-cisTarget 2015 update: generalized cis-regulatory enrichment analysis in human, mouse and fly. *Nucleic Acids Res.*, 43(W1):W57–64, jul 2015. ISSN 1362-4962 (Electronic). doi: 10.1093/nar/gkv395.

Ilya P. Ioshikhes and Michael Q. Zhang. Large-scale human promoter mapping using CpG islands. *Nat. Genet.*, 26(1):61–63, 2000. ISSN 10614036. doi: 10.1038/79189.

Alessandra Iurlo, Daniele Cattaneo, and Umberto Gianelli. Blast transformation in myeloproliferative neoplasms: Risk factors, biological findings, and targeted therapeutic options. *Int. J. Mol. Sci.*, 20(8):1–13, 2019. ISSN 14220067. doi: 10.3390/ijms20081839.

Andrejs Ivanovs, Stanislav Rybtsov, Lindsey Welch, Richard A Anderson, Marc L Turner, and Alexander Medvinsky. Highly potent human hematopoietic stem cells first emerge in the intraembryonic aorta-gonad-mesonephros region. *J. Exp. Med.*, 208(12):2417–2427, 2011. ISSN 00221007. doi: 10.1084/jem.20111688.

Hiromi Iwasaki, Chamorro Somoza, Hirokazu Shigematsu, Estelle A Duprez, Junko Iwasaki-Arai, Shin-Ichi Mizuno, Yojiro Arinobu, Kristin Geary, Pu Zhang, Tajhal Dayaram, Maris L Fenyus, Shannon Elf, Susan Chan, Philippe Kastner, Claudia S Huettner, Richard Murray, Daniel G Tenen, and Koichi Akashi. Distinctive and indispensable roles of PU.1 in maintenance of hematopoietic stem cells and their differentiation. *Blood*, 106(5):1590–600, sep 2005. ISSN 0006-4971. doi: 10.1182/blood-2005-03-0860.

Shankar Subramanian Iyer and Genhong Cheng. Role of interleukin 10 transcriptional regulation in inflammation and autoimmune disease. *Crit. Rev. Immunol.*, 32(1):23–63, 2012. ISSN 10408401.

Catriona H M Jamieson, Jason Gotlib, Jeffrey a Durocher, Mark P Chao, M Rajan Mariappan, Marla Lay, Carol Jones, James L Zehnder, Stan L Lilleberg, and Irving L Weissman. The JAK2 V617F mutation occurs in hematopoietic stem cells in polycythemia vera and predisposes toward erythroid differentiation. *Proc. Natl. Acad. Sci. U. S. A.*, 103(16):6224–9, 2006. ISSN 0027-8424. doi: 10.1073/pnas.0601462103.

Hong Ji, Lauren I R Ehrlich, Jun Seita, Peter Murakami, Akiko Doi, Paul Lindau, Hwajin Lee, Martin J. Aryee, Rafael A. Irizarry, Kitai Kim, Derrick J. Rossi, Matthew A. Inlay, Thomas Serwold, Holger Karsunky, Lena Ho, George Q. Daley, Irving L. Weissman, and Andrew P. Feinberg. Comprehensive methylome map of lineage commitment from haematopoietic progenitors. *Nature*, 467(7313):338–342, 2010. ISSN 14764687. doi: 10.1038/nature09367.

Catherine A. Johnson, Trevor H. Forster, Clay M. Winterford, and David J. Allan. Hydroxyurea induces apoptosis and regular DNA fragmentation in a Burkitt's lymphoma cell line. *Biochim. Biophys. Acta - Mol. Cell Res.*, 1136(1):1–4, jul 1992. ISSN 01674889. doi: 10.1016/0167-4889(92)90076-N.

Shweta Kambali and Asma Taj. Polycythemia vera masked due to severe iron deficiency anemia. *Hematol. Oncol. Stem Cell Ther.*, 11(1):38–40, 2018. ISSN 16583876. doi: 10.1016/j.hemonc.2016.08.007.

Leonie M Kamminga. The Polycomb group gene Ezh2 prevents hematopoietic stem cell exhaustion. *Blood*, 107(5):2170–2179, mar 2006. ISSN 0006-4971. doi: 10.1182/blood-2005-09-3585.

Young-Ju Kang, Seung-Jip Yang, Gyeongsin Park, Bin Cho, Chang-Ki Min, Tae-

Yoon Kim, Joon-Sung Lee, and Il-Hoan Oh. A Novel Function of Interleukin-10 Promoting Self-Renewal of Hematopoietic Stem Cells. *Stem Cells*, 25(7):1814–1822, 2007. ISSN 10665099. doi: 10.1634/stemcells.2007-0002.

P. Kastner and S. Chan. Function of RARα during the maturation of neutrophils. *Oncogene*, 20(49 REV. IIS. 6):7178–7185, 2001. ISSN 09509232. doi: 10.1038/sj.onc.1204757.

Benjamin T. Kile, Brenda A. Schulman, Warren S. Alexander, Nicos A. Nicola, Helene M.E. Martin, and Douglas J. Hilton. The SOCS box: A tale of destruction and degradation. *Trends Biochem. Sci.*, 27(5):235–241, 2002. ISSN 09680004. doi: 10.1016/S0968-0004(02)02085-6.

Maria Kleppe, Minsuk Kwak, Priya Koppikar, Markus Riester, Matthew Keller, Lennart Bastian, Todd Hricik, Neha Bhagwat, Anna Sophia McKenney, Efthymia Papalexi, Omar Abdel-Wahab, Raajit Rampal, Sachie Marubayashi, Jonathan J. Chen, Vincent Romanet, Jordan S. Fridman, Jacqueline Bromberg, J. Teruya-Feldstein, Masato Murakami, Thomas Radimerski, Franziska Michor, Rong Fan, and Ross L. Levine. JAK-STAT Pathway Activation in Malignant and Nonmalignant Cells Contributes to MPN Pathogenesis and Therapeutic Response. *Cancer Discov.*, 5(3):316–331, mar 2015. ISSN 2159-8274. doi: 10.1158/2159-8290.CD-14-0736.

Ursula Klingmüller, Ulrike Lorenz, Lewis C. Cantley, Benjamin G. Neel, and Harvey F. Lodish. Specific recruitment of SH-PTP1 to the erythropoietin receptor causes inactivation of JAK2 and termination of proliferative signals. *Cell*, 80(5):729–738, 1995. ISSN 00928674. doi: 10.1016/0092-8674(95)90351-8.

Ursula Klingmüller, Hong Wu, Jonathan G. Hsiao, Alex Toker, Brian C. Duckworth, Lewis C. Cantley, and Harvey F. Lodish. Identification of a novel pathway important for proliferation and differentiation of primary erythroid progenitors.

Proc. Natl. Acad. Sci. U. S. A., 94(7):3016–3021, 1997. ISSN 00278424. doi: 10.1073/pnas.94.7.3016.

Myunggon Ko, Yun Huang, Anna M Jankowska, Utz J Pape, Mamta Tahiliani, Hozefa S Bandukwala, Jungeun An, Edward D Lamperti, Kian Peng Koh, Rebecca Ganetzky, X Shirley Liu, L Aravind, Suneet Agarwal, Jaroslaw P Maciejewski, and Anjana Rao. Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature*, 468(7325):839–843, 2010. ISSN 0028-0836. doi: 10.1038/nature09586.

Motonari Kondo, Irving L. Weissman, and Koichi Akashi. Identification of clonogenic common lymphoid progenitors in mouse bone marrow. *Cell*, 91(5):661–672, 1997. ISSN 00928674. doi: 10.1016/S0092-8674(00)80453-5.

Yutaka Kondo, Lanlan Shen, Alfred S. Cheng, Saira Ahmed, Yanis Boumber, Chantale Charo, Tadanori Yamochi, Takeshi Urano, Koichi Furukawa, Bernard Kwabi-Addo, David L. Gold, Yoshitaka Sekido, Tim Hui Ming Huang, and Jean Pierre J. Issa. Gene silencing in cancer by histone H3 lysine 27 trimethylation independent of promoter DNA methylation. *Nat. Genet.*, 40(6):741–750, 2008. ISSN 10614036. doi: 10.1038/ng.159.

Tony Kouzarides. Chromatin Modifications and Their Function. *Cell*, 128(4):693–705, 2007. ISSN 00928674. doi: 10.1016/j.cell.2007.02.005.

Irwin H Krakoff, Neal C Brown, and Peter Reichard. Inhibition of ribonucleoside diphosphate reductase by hydroxyurea. *Cancer Res.*, 28(8):1559–65, aug 1968. ISSN 0008-5472.

Robert Kralovics, Francesco Passamonti, Andreas S. Buser, Soon Siong Teo, Ralph Tiedt, Jakob R Passweg, Andre Tichelli, Mario Cazzola, and Radek C Skoda. A gain-of-function mutation of JAK2 in myeloproliferative disorders. *N. Engl. J. Med.*, 352(17):1779–1790, 2005. ISSN 00284793. doi: 10.1056/NEJMoa051113.

Kandasamy Krishnaraju, Barbara Hoffman, and Dan A. Liebermann. Early growth response gene 1 stimulates development of hematopoietic progenitor cells along the macrophage lineage at the expense of the granulocyte and erythroid lineages. *Blood*, 97(5):1298–1305, 2001. ISSN 00064971. doi: 10.1182/blood.V97.5.1298.

Lucia Kubovcakova, Pontus Lundberg, Jean Grisouard, Hui Hao-Shen, Vincent Romanet, Rita Andraos, Masato Murakami, Stephan Dirnhofer, Kay-Uwe Uwe Wagner, Thomas Radimerski, and Radek C. Skoda. Differential effects of hydroxyurea and INC424 on mutant allele burden and myeloproliferative phenotype in a JAK2-V617F polycythemia vera mouse model. *Blood*, 121(7):1188–1199, 2013. ISSN 00064971. doi: 10.1182/blood-2012-03-415646.

Ralf Kühn, Frieder Schwenk, Michel Aguet, and Klaus Rajewsky. Inducible gene targeting in mice. *Science (80-.).*, 269(5229):1427–1429, 1995. ISSN 00368075. doi: 10.1126/science.7660125.

Maxim V. Kuleshov, Matthew R. Jones, Andrew D. Rouillard, Nicolas F. Fernandez, Qiaonan Duan, Zichen Wang, Simon Koplev, Sherry L. Jenkins, Kathleen M. Jagodnik, Alexander Lachmann, Michael G. McDermott, Caroline D. Monteiro, Gregory W. Gundersen, and Avi Ma'ayan. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.*, 44(W1):W90–W97, 2016. ISSN 13624962. doi: 10.1093/nar/gkw377.

Jeeba Kuriakose, Vanessa Redecke, Cliff Guy, Jingran Zhou, Ruiqiong Wu, Sirish K. Ippagunta, Heather Tillman, Patrick D. Walker, Peter Vogel, and Hans Häcker. Patrolling monocytes promote the pathogenesis of early lupus-like glomerulonephritis. *J. Clin. Invest.*, 129(6):2251–2265, 2019. ISSN 15588238. doi: 10.1172/JCI125116.

Teneema Kuriakose, Si Ming Man, R. K. Subbarao Malireddi, Rajendra Karki, Sannula Kesavardhana, David E. Place, Geoffrey Neale, Peter Vogel, and Thirumala-

Devi Kanneganti. ZBP1/DAI is an innate sensor of influenza virus triggering the NLRP3 inflammasome and programmed cell death pathways. *Sci. Immunol.*, 1(2): aag2045–aag2045, aug 2016. ISSN 2470-9468. doi: 10.1126/sciimmunol.aag2045.

Daisuke Kurotaki, Michio Yamamoto, Akira Nishiyama, Kazuhiro Uno, Tatsuma Ban, Motohide Ichino, Haruka Sasaki, Satoko Matsunaga, Masahiro Yoshinari, Akihide Ryo, Masatoshi Nakazawa, Keiko Ozato, and Tomohiko Tamura. IRF8 inhibits C/EBPα activity to restrain mononuclear phagocyte progenitors from differentiating into neutrophils. *Nat. Commun.*, 5:1–15, 2014. ISSN 20411723. doi: 10.1038/ncomms5978.

Daisuke Kurotaki, Jun Nakabayashi, Akira Nishiyama, Haruka Sasaki, Wataru Kawase, Naofumi Kaneko, Kyoko Ochiai, Kazuhiko Igarashi, Keiko Ozato, Yutaka Suzuki, and Tomohiko Tamura. Transcription Factor IRF8 Governs Enhancer Landscape Dynamics in Mononuclear Phagocyte Progenitors. *Cell Rep.*, 22(10): 2628–2641, 2018. ISSN 22111247. doi: 10.1016/j.celrep.2018.02.048.

Julie Lacombe, Sabine Herblot, Shanti Rojas-Sutterlin, André Haman, Stéphane Barakat, Norman N. Iscove, Guy Sauvageau, and Trang Hoang. Scl regulates the quiescence and the long-term competence of hematopoietic stem cells. *Blood*, 115 (4):792–803, 2010. ISSN 00064971. doi: 10.1182/blood-2009-01-201384.

Catherine Lacout, Didier F Pisani, Micheline Tulliez, and Françoise Moreau Gachelin. JAK2 V617F expression in murine hematopoietic cells leads to MPD mimicking human PV with secondary myelofibrosis JAK2 V617F expression in murine hematopoietic cells leads to MPD mimicking human PV with secondary myelofibrosis. 108(5):1652–1660, 2008. doi: 10.1182/blood-2006-02-002030.

Raffaele Landolfi, Roberto Marchioli, Jack Kutti, Heinz Gisslinger, Gianni Tognoni, Carlo Patrono, and Tiziano Barbui. Efficacy and Safety of Low-Dose Aspirin in

Polycythemia Vera. *N. Engl. J. Med.*, 350(2):114–124, 2004. ISSN 00284793. doi: 10.1056/NEJMoa035572.

Jens Langstein, Michael D. Milsom, and Daniel B. Lipka. Impact of DNA methylation programming on normal and pre-leukemic hematopoiesis. *Semin. Cancer Biol.*, 51(September 2017):89–100, 2018. ISSN 10963650. doi: 10.1016/j.semcancer.2017.09.008.

Peter Laslo, Chauncey J. Spooner, Aryeh Warmflash, David W. Lancki, Hyun Jun Lee, Roger Sciammas, Benjamin N. Gantner, Aaron R. Dinner, and Harinder Singh. Multilineage Transcriptional Priming and Determination of Alternate Hematopoietic Cell Fates. *Cell*, 126(4):755–766, 2006. ISSN 00928674. doi: 10.1016/j.cell.2006.06.052.

Elisa Laurenti and Berthold Göttgens. From haematopoietic stem cells to complex differentiation landscapes. *Nature*, 553(7689):418–426, jan 2018. ISSN 0028-0836. doi: 10.1038/nature25022.

Jeffrey T. Leek, W. Evan Johnson, Hilary S. Parker, Andrew E. Jaffe, and John D. Storey. The SVA package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics*, 28(6):882–883, 2012. ISSN 13674803. doi: 10.1093/bioinformatics/bts034.

Benjamin Lehne, Alexander W. Drong, Marie Loh, Weihua Zhang, William R. Scott, Sian Tsung Tan, Uzma Afzal, James Scott, Marjo Riitta Jarvelin, Paul Elliott, Mark I. McCarthy, Jaspal S. Kooner, and John C. Chambers. A coherent approach for analysis of the Illumina HumanMethylation450 BeadChip improves data quality and performance in epigenome-wide association studies. *Genome Biol.*, 16(1): 1–12, 2015. ISSN 1474760X. doi: 10.1186/s13059-015-0600-x.

Hong Lei, Suk P Oh, Masaki Okano, Ruth Jüttermann, Kendrick A Goss, Rudolf

Jaenisch, and En Li. De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. *Development*, 122(10):3195–3205, 1996. ISSN 09501991.

T J Ley, J DeSimone, C T Noguchi, P H Turner, A N Schechter, P Heller, and A W Nienhuis. 5-Azacytidine increases gamma-globin synthesis and reduces the proportion of dense cells in patients with sickle cell anemia. *Blood*, 62(2):370–80, aug 1983. ISSN 0006-4971.

Timothy J. Ley, Li Ding, Matthew J. Walter, Michael D. McLellan, Tamara Lamprecht, David E Larson, D Ph, Cyriac Kandoth, E Jacqueline, Jack Baty, John Welch, Christopher C Harris, Cheryl F Lichti, R Reid Townsend, Robert S. Fulton, David J Dooling, Daniel C Koboldt, Heather Schmidt, Qunyuan Zhang, R John, Sean D. McGrath, Lucinda a Fulton, Vincent J Magrini, Tammi L Vickery, Jasreet Hundal, Lisa L Cook, Joshua J Conyers, W Swift, Jerry P Reed, Patricia A Alldredge, Todd Wylie, Joelle Kalicki, Mark a Watson, Sharon Heath, D William, Nobish Varghese, Rakesh Nagarajan, Peter Westervelt, Jacqueline E Payton, Jack Baty, John Welch, Christopher C Harris, Cheryl F Lichti, R Reid Townsend, Robert S. Fulton, David J Dooling, Daniel C Koboldt, Heather Schmidt, Qunyuan Zhang, John R Osborne, Ling Lin, Michelle O'Laughlin, Joshua F. McMichael, Kim D Delehaunty, Sean D. McGrath, Lucinda a Fulton, Vincent J Magrini, Tammi L Vickery, Jasreet Hundal, Lisa L Cook, Joshua J Conyers, Gary W Swift, Jerry P Reed, Patricia A Alldredge, Todd Wylie, Jason Walker, Joelle Kalicki, Mark a Watson, Sharon Heath, William D Shannon, Nobish Varghese, Rakesh Nagarajan, Peter Westervelt, Michael H. Tomasson, Daniel C. Link, Timothy A. Graubert, John F. DiPersio, Elaine R Mardis, and Richard K Wilson. DNTM3A mutations in acute myeloid leukemia. *N. Engl. J. Med.*, 363(25):2424–2433, dec 2010. ISSN 1533-4406. doi: 10.1056/NEJMoa1005143.DNMT3A.

Juan Li, Dominik Spensberger, Jong Sook Ahn, Shubha Anand, Philip A. Beer, Cedric Ghevaert, Edwin Chen, Ariel Forrai, Linda M. Scott, Rita Ferreira, Peter J.

Campbell, Steve P. Watson, Pentao Liu, Wendy N. Erber, Brian J.P. P Huntly, Katrin Ottersbach, and Anthony R. Green. JAK2 V617F impairs hematopoietic stem cell function in a conditional knock-in mouse model of JAK2 V617F-positive essential thrombocythemia. *Blood*, 116(9):1528–38, sep 2010. ISSN 1528-0020. doi: 10.1182/blood-2009-12-259747.

Lei Li, Susan M. Byrne, Nicole Rainville, Su Su, Edward Jachimowicz, Anne Aucher, Daniel M. Davis, Philip G. Ashton-Rickardt, and Don M. Wojchowski. Brief report: Serpin Spi2A as a novel modulator of hematopoietic progenitor cell formation. *Stem Cells*, 32(9):2550–2556, sep 2014. ISSN 15494918. doi: 10.1002/stem.1778.

Wei Liao, Jian-xin Lin, and Warren J Leonard. IL-2 family cytokines: new insights into the complex roles of IL-2 as a broad regulator of T helper cell differentiation. *Curr. Opin. Immunol.*, 23(5):598–604, oct 2011. ISSN 09527915. doi: 10.1016/j.coi.2011.08.003.

B Liu, J Liao, X Rao, S A Kushner, C D Chung, D D Chang, and K Shuai. Inhibition of Stat1-mediated gene activation by PIAS1. *Proc. Natl. Acad. Sci. U. S. A.*, 95 (18):10626–10631, sep 1998. ISSN 0027-8424. doi: 10.1073/pnas.95.18.10626.

Fan Liu, Xinyang Zhao, Fabiana Perna, Lan Wang, Priya Koppikar, Omar Abdel-Wahab, Michael W. Harr, Ross L. Levine, Hao Xu, Ayalew Tefferi, Anthony Deblasio, Megan Hatlen, Silvia Menendez, and Stephen D. Nimer. JAK2V617F-Mediated Phosphorylation of PRMT5 Downregulates Its Methyltransferase Activity and Promotes Myeloproliferation. *Cancer Cell*, 19(2):283–294, feb 2011. ISSN 15356108. doi: 10.1016/j.ccr.2010.12.020.

Eva Löfvenberg and Anders Wahlin. Management of polycythaemia vera, essential thrombocythaemia and myelofibrosis with hydroxyurea. *Eur. J. Haematol.*, 41(4): 375–381, apr 1988. ISSN 16000609. doi: 10.1111/j.1600-0609.1988.tb00212.x.

Tzu-Fang Lou, Manisha Singh, Ashley Mackie, Wei Li, and Betty S. Pace. Hydroxyurea Generates Nitric Oxide in Human Erythroid Cells: Mechanisms for gamma-Globin Gene Activation. *Exp. Biol. Med.*, 234(11):1374–1382, 2009. ISSN 1535-3702. doi: 10.3181/0811-rm-339.

Xiaohui Lu, Lily Jun Shen Huang, and Harvey F. Lodish. Dimerization by a cytokine receptor is necessary for constitutive activation of JAK2V617F. *J. Biol. Chem.*, 283(9):5258–5266, 2008. ISSN 00219258. doi: 10.1074/jbc.M707125200.

Isabella W.Y. Mak, Nathan Evaniew, and Michelle Ghert. Lost in translation: Animal models and clinical trials in cancer treatment. *Am. J. Transl. Res.*, 6(2): 114–118, 2014. ISSN 19438141.

Ishminder K. Mann, Raghunath Chatterjee, Jianfei Zhao, Ximiao He, Matthew T. Weirauch, Timothy R. Hughes, and Charles Vinson. CG methylated microarrays identify a novel methylated sequence bound by the CEBPB|ATF4 heterodimer that is active in vivo. *Genome Res.*, 23(6):988–997, 2013. ISSN 10889051. doi: 10.1101/gr.146654.112.

Elaine R. Mardis, Li Ding, David J. Dooling, David E. Larson, Michael D. McLellan, Ken Chen, Daniel C. Koboldt, Robert S. Fulton, Kim D. Delehaunty, Sean D. McGrath, Lucinda A. Fulton, Devin P. Locke, Vincent J. Magrini, Rachel M. Abbott, Tammi L. Vickery, Jerry S. Reed, Jody S. Robinson, Todd Wylie, Scott M. Smith, Lynn Carmichael, James M. Eldred, Christopher C. Harris, Jason Walker, Joshua B. Peck, Feiyu Du, Adam F. Dukes, Gabriel E. Sanderson, Anthony M. Brummett, Eric Clark, Joshua F. McMichael, Rick J. Meyer, Jonathan K. Schindler, Craig S. Pohl, John W. Wallis, Xiaoqi Shi, Ling Lin, Heather Schmidt, Yuzhu Tang, Carrie Haipek, Madeline E. Wiechert, Jolynda V. Ivy, Joelle Kalicki, Glendoria Elliott, Rhonda E. Ries, Jacqueline E. Payton, Peter Westervelt, Michael H. Tomasson, Mark A. Watson, Jack Baty, Sharon Heath, William D. Shannon, Rakesh Nagarajan, Daniel C. Link, Matthew J.

Walter, Timothy A. Graubert, John F. DiPersio, Richard K. Wilson, and Timothy J. Ley. Recurring mutations found by sequencing an acute myeloid leukemia genome. *N. Engl. J. Med.*, 361(11):1058–1066, 2009. ISSN 15334406. doi: 10.1056/NEJMoa0903840.

Cecilia P. Marin Oyarzún, Agostina Carestia, Paola R. Lev, Ana C. Glembotsky, Miguel A. Castro Ríos, Beatriz Moiraghi, Felisa C. Molinas, Rosana F. Marta, Mirta Schattner, and Paula G. Heller. Neutrophil extracellular trap formation and circulating nucleosomes in patients with chronic myeloproliferative neoplasms. *Sci. Rep.*, 6(1):38738, 2016. ISSN 2045-2322. doi: 10.1038/srep38738.

Alejandra Martínez-Trillos, Anna Gaya, Margherita Maffioli, Eduardo Arellano-Rodrigo, Xavier Calvo, Marina Díaz-Beyá, and Francisco Cervantes. Efficacy and tolerability of hydroxyurea in the treatment of the hyperproliferative manifestations of myelofibrosis: Results in 40 patients. *Ann. Hematol.*, 89(12):1233–1237, 2010. ISSN 09395555. doi: 10.1007/s00277-010-1019-9.

C Marty, C Lacout, A Martin, S Hasan, S Jacquot, M C Birling, W Vainchenker, and J L Villeval. Myeloproliferative neoplasm induced by constitutive expression of JAK2V617F in knock-in mice. *Blood*, 116(5):783–787, 2010. ISSN 0006-4971. doi: 10.1182/blood-2009-12-257063.

Stéphanie Maupetit-Mehouas, Franck Court, Céline Bourgne, Agnès Guerci-Bresler, Pascale Cony-Makhoul, Hyacinthe Johnson, Gabriel Etienne, Philippe Rousselot, Denis Guyotat, Alexandre Janel, Eric Hermet, Sandrine Saugues, Juliette Berger, Philippe Arnaud, and Marc G. Berger. DNA methylation profiling reveals a pathological signature that contributes to transcriptional defects of CD34+CD15- cells in early chronic-phase chronic myeloid leukemia. *Mol. Oncol.*, 12(6):814–829, 2018. ISSN 18780261. doi: 10.1002/1878-0261.12191.

Shannon McKinney-Freeman, Patrick Cahan, Hu Li, Scott A. Lacadie, Hsuan-Ting

Huang, Matthew Curran, Sabine Loewer, Olaia Naveiras, Katie L. Kathrein, Martina Konantz, Erin M. Langdon, Claudia Lengerke, Leonard I. Zon, James J. Collins, and George Q. Daley. The Transcriptional Landscape of Hematopoietic Stem Cell Ontogeny. *Cell Stem Cell*, 11(5):701–714, nov 2012. ISSN 19345909. doi: 10.1016/j.stem.2012.07.018.

Mary F. Mcmullin, Bridget S. Wilkins, and Claire N. Harrison. Management of polycythaemia vera: A critical review of current data. *Br. J. Haematol.*, 172 (October 2015):337–349, feb 2015. ISSN 13652141. doi: 10.1111/bjh.13812.

Mary Frances McMullin, Claire N. Harrison, Sahra Ali, Catherine Cargo, Frederick Chen, Joanne Ewing, Mamta Garg, Anna Godfrey, S. Steven Knapper, Donal P. McLornan, Jyoti Nangalia, Mallika Sekhar, Frances Wadelin, and Adam J. Mead. A guideline for the diagnosis and management of polycythaemia vera. A British Society for Haematology Guideline. *Br. J. Haematol.*, 184(2):176–191, 2019. ISSN 13652141. doi: 10.1111/bjh.15648.

Yulia A. Medvedeva, Andreas Lennartsson, Rezvan Ehsani, Ivan V. Kulakovskiy, Ilya E. Vorontsov, Pouda Panahandeh, Grigory Khimulya, Takeya Kasukawa, and Finn Drabløs. EpiFactors: A comprehensive database of human epigenetic factors and complexes. *Database*, 2015:1–10, 2015. ISSN 17580463. doi: 10.1093/database/bav067.

A L Medvinsky, N L Samoylina, A M Muller, and E A Dzierzak. An early preliver intraembryonic source of CFU-S in the developing mouse. *Nature*, 364(6432): 64–67, jul 1993. ISSN 0028-0836 (Print). doi: 10.1038/364064a0.

Alexander Medvinsky, Stanislav Rybtsov, and Samir Taoudi. Embryonic origin of the adult hematopoietic system: advances and questions. *Development*, 138(6): 1017–1031, mar 2011. ISSN 1477-9129 (Electronic). doi: 10.1242/dev.040998.

Alexander Meissner, Tarjei S. Mikkelsen, Hongcang Gu, Marius Wernig, Jacob

Hanna, Andrey Sivachenko, Xiaolan Zhang, Bradley E. Bernstein, Chad Nusbaum, David B. Jaffe, Andreas Gnirke, Rudolf Jaenisch, Eric S. Lander, Jacob Hanna, Andrey Sivachenko, Xiaolan Zhang, Bradley E. Bernstein, Chad Nusbaum, David B. Jaffe, Andreas Gnirke, Rudolf Jaenisch, and Eric S. Lander. Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature*, 454 (7205):766–770, aug 2008. ISSN 0028-0836. doi: 10.1038/nature07107.

A Merlat, J L Lai, Y Sterkers, J L Demory, F Bauters, C Preudhomme, and P Fenaux. Therapy-related myelodysplastic syndrome and acute myeloid leukemia with 17p deletion. A report on 25 cases. *Leukemia*, 13(2):250–257, 1999. ISSN 08876924.

Javier Mestas and Christopher C. W. Hughes. Of Mice and Not Men: Differences between Mouse and Human Immunology. *J. Immunol.*, 172(5):2731–2738, 2004. ISSN 0022-1767. doi: 10.4049/jimmunol.172.5.2731.

Irene M. Min, Giorgio Pietramaggiori, Francis S. Kim, Emmanuelle Passegué, Kristen E. Stevenson, and Amy J. Wagers. The Transcription Factor EGR1 Controls Both the Proliferation and Localization of Hematopoietic Stem Cells. *Cell Stem Cell*, 2(4):380–391, 2008. ISSN 19345909. doi: 10.1016/j.stem.2008.01.015.

Yoshitaka Miyakawa, Atsushi Oda, Brian J. Druker, Hiroshi Miyazaki, Makoto Handa, Hideya Ohashi, and Yasuo Ikeda. Thrombopoietin induces tyrosine phosphorylation of Stat3 and Stat5 in human blood platelets. *Blood*, 87(2):439–446, 1996. ISSN 00064971.

M A Moore and J J Owen. Chromosome marker studies on the development of the haemopoietic system in the chick embryo. *Nature*, 208(5014):956 passim, dec 1965. ISSN 0028-0836 (Print). doi: 10.1038/208956a0.

M A Moore and J J Owen. Chromosome marker studies in the irradiated chick

embryo. *Nature*, 215(5105):1081–1082, sep 1967. ISSN 0028-0836 (Print). doi: 10.1038/2151081a0.

Daniel E. Morales-Mantilla and Katherine Y. King. The Role of Interferon-Gamma in Hematopoietic Stem Cell Development, Homeostasis, and Disease. *Curr. Stem Cell Reports*, 4(3):264–271, 2018. ISSN 21987866. doi: 10.1007/s40778-018-0139-3.

Kelly Moran-Crusio, Linsey Reavie, Alan Shih, Omar Abdel-Wahab, Delphine Ndiaye-Lobry, Camille Lobry, Maria E. Figueroa, Aparna Vasanthakumar, Jay Patel, Xinyang Zhao, Fabiana Perna, Suveg Pandey, Jozef Madzo, Chunxiao Song, Qing Dai, Chuan He, Sherif Ibrahim, Miloslav Beran, Jiri Zavadil, Stephen D. Nimer, Ari Melnick, Lucy A. Godley, Iannis Aifantis, and Ross L. Levine. Tet2 Loss Leads to Increased Hematopoietic Stem Cell Self-Renewal and Myeloid Transformation. *Cancer Cell*, 20(1):11–24, 2011. ISSN 15356108. doi: 10.1016/j.ccr.2011.06.001.

H. Motenko, S. B. Neuhauser, M. O'Keefe, and J. E. Richardson. MouseMine: a new data warehouse for MGI. *Mamm. Genome*, 26(7-8):325–330, aug 2015. ISSN 0938-8990. doi: 10.1007/s00335-015-9573-z.

Beatrice U. Mueller, Thomas Pabst, Motomi Osato, Norio Asou, Lisa M. Johansen, Mark D. Minden, Gerhard Behre, Wolfgang Hiddemann, Yoshiaki Ito, and Daniel G. Tenen. Heterozygous PU.1 mutations are associated with acute myeloid leukemia. *Blood*, 100(3):998–1007, 2002. ISSN 00064971. doi: 10.1182/blood.V100.3.998.

Ann Mullally, Steven W. Lane, Brian Ball, Christine Megerdichian, Rachel Okabe, Fatima Al-Shahrour, Mahnaz Paktinat, J. Erika Haydu, Elizabeth Housman, Allegra M. Lord, Gerlinde Wernig, Michael G. Kharas, Thomas Mercher, Jeffery L. Kutok, D. Gary Gilliland, and Benjamin L. Ebert. Physiological Jak2V617F Expression Causes a Lethal Myeloproliferative Neoplasm with Differential Effects on

Hematopoietic Stem and Progenitor Cells. *Cancer Cell*, 17(6):584–596, 2010. ISSN 15356108. doi: 10.1016/j.ccr.2010.05.015.

Claus Nerlov and Thomas Graf. PU.1 induces myeloid lineage commitment in multipotent hematopoietic progenitors. *Genes Dev.*, 12(15):2403–2412, 1998. ISSN 08909369. doi: 10.1101/gad.12.15.2403.

Zuqin Nie, Gangqing Hu, Gang Wei, Kairong Cui, Arito Yamane, Wolfgang Resch, Ruoning Wang, Douglas R Green, Lino Tessarollo, Rafael Casellas, Keji Zhao, and David Levens. c-Myc is a universal amplifier of expressed genes in lymphocytes and embryonic stem cells. *Cell*, 151(1):68–79, sep 2012. ISSN 1097-4172. doi: 10.1016/j.cell.2012.08.033.

Sangeeta Nischal, Sanchari Bhattacharyya, Maximilian Christopeit, Yiting Yu, Li Zhou, Tushar D. Bhagat, Davendra Sohal, Britta Will, Yongkai Mo, Masako Suzuki, Animesh Pardanani, Michael McDevitt, Jaroslaw P. Maciejewski, Ari M. Melnick, John M. Greally, Ulrich Steidl, Alison Moliterno, and Amit Verma. Methylome profiling reveals distinct alterations in phenotypic and mutational subgroups of myeloproliferative neoplasms. *Cancer Res.*, 73(3):1076–1085, 2013. ISSN 00085472. doi: 10.1158/0008-5472.CAN-12-0735.

Laura J. Norton, Alister P. W. Funnell, Jon Burdach, Beeke Wienert, Ryo Kurita, Yukio Nakamura, Sjaak Philipsen, Richard C. M. Pearson, Kate G. R. Quinlan, and Merlin Crossley. KLF1 directly activates expression of the novel fetal globin repressor ZBTB7A/LRF in erythroid cells . *Blood Adv.*, 1(11):685–692, 2017. ISSN 2473-9529. doi: 10.1182/bloodadvances.2016002303.

P. Nowell and D. Hungerford. A minute chromosome in human chronic granulocytic leukemia. *Landmarks Med. Genet. Class. Pap. with Comment.*, page 132:103, may 1985.

Masaki Okano, Shaoping Xie, and En Li. Cloning and characterization of a family

of novel mammalian DNA (cytosine-5) methyltransferases Non-invasive sexing of preimplantation stage mammalian embryos. *Nat. Am. Inc.*, 19(july):219–220, 1998. ISSN 1061-4036. doi: 10.1038/890.

Tsukasa Okuda, Jan Van Deursen, Scott W Hiebert, Gerard Grosveld, and James R Downing. AML1, the target of multiple chromosomal translocations in human leukemia, is essential for normal fetal liver hematopoiesis. *Cell*, 84(2):321–330, 1996. ISSN 00928674. doi: 10.1016/S0092-8674(00)80986-1.

Y. Okuno, G. Huang, F. Rosenbauer, E. K. Evans, H. S. Radomska, H. Iwasaki, K. Akashi, F. Moreau-Gachelin, Y. Li, P. Zhang, B. Gottgens, and D. G. Tenen. Potential Autoregulation of Transcription Factor PU.1 by an Upstream Regulatory Element. *Mol. Cell. Biol.*, 25(7):2832–2845, 2005. ISSN 0270-7306. doi: 10.1128/mcb.25.7.2832-2845.2005.

Timothy S Olson and Klaus Ley. Chemokines and chemokine receptors in leukocyte trafficking. *Am. J. Physiol. Integr. Comp. Physiol.*, 283(1):R7–R28, jul 2002. ISSN 0363-6119. doi: 10.1152/ajpregu.00738.2001.

Stuart H Orkin. Diversification of haematopoietic stem cells to specific lineages. *Nat. Rev. Genet.*, 1(1):57–64, oct 2000. ISSN 1471-0056. doi: 10.1038/35049577.

Stuart H Orkin and Leonard I Zon. Hematopoiesis: an evolving paradigm for stem cell biology. *Cell*, 132(4):631–44, feb 2008. ISSN 1097-4172. doi: 10.1016/j.cell.2008.01.025.

Marielle A. Otten, Esther Rudolph, Michael Dechant, Cornelis W. Tuk, Rogier M. Reijmers, Robert H. J. Beelen, Jan G. J. van de Winkel, and Marjolein van Egmond. Immature Neutrophils Mediate Tumor Cell Killing via IgA but Not IgG Fc Receptors. *J. Immunol.*, 174(9):5472–5480, 2005. ISSN 0022-1767. doi: 10.4049/jimmunol.174.9.5472.

James Palis. Yolk-sac hematopoiesis The first blood cells of mouse and man. *Exp. Hematol.*, 29(8):927–936, aug 2001. ISSN 0301472X. doi: 10.1016/S0301-472X(01)00669-5.

Marina Panova-Noeva, Marina Marchetti, Sabrina Buoro, Laura Russo, Annamaria Leuzzi, Guido Finazzi, Alessandro Rambaldi, Cosimo Ottomano, Hugo Ten Cate, and Anna Falanga. JAK2V617F mutation and hydroxyurea treatment as determinants of immature platelet parameters in essential thrombocythemia and polycythemia vera patients. *Blood*, 118(9):2599–2601, 2011. ISSN 00064971. doi: 10.1182/blood-2011-02-339655.

Katerina E. Panteli, Eleftheria C. Hatzimichael, Paraskevi K. Bouranta, Afroditi Katsaraki, Konstantinos Seferiadis, Justin Stebbing, and Konstantinos L. Bourantas. Serum interleukin (IL)-1, IL-2, sIL-2Ra, IL-6 and thrombopoietin levels in patients with chronic myeloproliferative diseases. *Br. J. Haematol.*, 130(5):709–715, 2005. ISSN 00071048. doi: 10.1111/j.1365-2141.2005.05674.x.

A. Pardanani, T. L. Lasho, C. M. Finke, M. Mai, R. F. McClure, and A. Tefferi. IDH1 and IDH2 mutation analysis in chronic-and blast-phase myeloproliferative neoplasms. *Leukemia*, 24(6):1146–1151, 2010. ISSN 14765551. doi: 10.1038/leu.2010.77.

Ryley Parrish, Jeremy J Day, and Farah D Lubin. Direcct bisulfite sequencing for examintation of DNA methaylation patters with gene and nucleotide resulotuion from brain tissues. *Curr Protoc Neurosci*, Chapter: U(205):1–14, 2012. doi: 10.1002/0471142301.ns0724s60.Direct.

Danilo Pellin, Mariana Loperfido, Cristina Baricordi, Samuel L Wolock, Annita Montepeloso, Olga K Weinberg, Alessandra Biffi, Allon M Klein, and Luca Biasco. A comprehensive single cell transcriptional landscape of human hematopoietic

progenitors. *Nat. Commun.*, 10(1):2395, dec 2019. ISSN 2041-1723. doi: 10.1038/s41467-019-10291-0.

Eva Petermann, Manuel Luís Orta, Natalia Issaeva, Niklas Schultz, and Thomas Helleday. Hydroxyurea-Stalled Replication Forks Become Progressively Inactivated and Require Two Different RAD51-Mediated Pathways for Restart and Repair. *Mol. Cell*, 37(4):492–502, 2010. ISSN 10972765. doi: 10.1016/j.molcel.2010.01.021.

Timothy J Peters, Michael J Buckley, Aaron L Statham, Ruth Pidsley, Katherine Samaras, Reginald V Lord, Susan J Clark, Peter L Molloy, Reginald V Lord, Susan J Clark, Peter L Molloy, Reginald V Lord, Susan J Clark, and Peter L Molloy. De novo identification of differentially methylated regions in the human genome. *Epigenetics and chromatin*, 8(1):6, 2015. ISSN 1756-8935. doi: 10.1186/1756-8935-8-6.

Larysa Pevny, M Celeste Simon, Elizabeth Robertson, William H Klein, Shih-Feng Tsai, Vivette D'Agati, Stuart H Orkin, and Frank Costantini. Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. *Nature*, 349(6306):257–260, 1991. ISSN 1476-4687. doi: 10.1038/349257a0.

Ruth Pidsley, Chloe C Y Wong, Manuela Volta, Katie Lunnon, Jonathan Mill, and Leonard C Schalkwyk. A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC Genomics*, 14:293, may 2013. ISSN 1471-2164 (Electronic). doi: 10.1186/1471-2164-14-293.

Ruth Pidsley, Elena Zotenko, Timothy J. Peters, Mitchell G. Lawrence, Gail P. Risbridger, Peter Molloy, Susan Van Djik, Beverly Muhlhausler, Clare Stirzaker, Susan J. Clark, PA Jones, SB Baylin, YA Ko, D Mohtat, M Suzuki, AS Park, MC Izquierdo, SY Han, T Dayeh, P Volkov, S Salo, E Hall, E Nilsson, AH Ols-

son, R Pidsley, J Viana, E Hannon, H Spiers, C Troakes, S Al-Saraj, C Stirzaker, PC Taberlay, AL Statham, SJ Clark, SJ Clark, J Harrison, CL Paul, M Frommer, R Lister, M Pelizzola, RH Dowen, RD Hawkins, G Hon, J Tonti-Filippini, M Bibikova, J Le, B Barnes, S Saedinia-Melnyk, L Zhou, R Shen, T Hinoue, DJ Weisenberger, CP Lange, H Shen, HM Byun, D Berg, LP Breitling, R Yang, B Korn, B Burwinkel, H Brenner, VK Rakyan, TA Down, S Maslau, T Andrew, TP Yang, H Beyan, M Bibikova, B Barnes, C Tsan, V Ho, B Klotzle, JM Le, TJ Morris, S Beck, YA Chen, S Choufani, D Grafodatskaya, DT Butcher, JC Ferreira, R Weksberg, YA Chen, M Lemire, S Choufani, DT Butcher, D Grafodatskaya, BW Zanke, H Naeem, NC Wong, Z Chatterton, MK Hong, JS Pedersen, NM Corcoran, TJ Peters, MJ Buckley, AL Statham, R Pidsley, K Samaras, R V Lord, D Wang, L Yan, Q Hu, LE Sucheston, MJ Higgins, CB Ambrosone, CD Warden, H Lee, JD Tompkins, X Li, C Wang, AD Riggs, M Lizio, J Harshbarger, H Shimoji, J Severin, T Kasukawa, S Sahin, L Siggens, K Ekwall, S Dedeurwaerder, M Defrance, E Calonne, H Denis, C Sotiriou, F Fuks, R Pidsley, Y Wong CC, M Volta, K Lunnon, J Mill, LC Schalkwyk, AE Teschendorff, F Marabita, M Lechner, T Bartlett, J Tegner, D Gomez-Cabrero, N Touleimat, J Tost, RE Thurman, E Rynes, R Humbert, J Vierstra, MT Maurano, E Haugen, R Andersson, C Gebhard, I Miguel-Escalada, I Hoof, J Bornholdt, M Boyd, A Kundaje, W Meuleman, J Ernst, M Bilenky, A Yen, ME Ritchie, B Phipson, D Wu, Y Hu, CW Law, W Shi, MB Stadler, R Murr, L Burger, R Ivanek, F Lienert, A Schöler, MJ Ziller, H Gu, F Müller, J Donaghey, LT-Y Tsai, O Kohlbacher, SP Huang, BY Bao, TC Hour, CY Huang, CC Yu, CC Liu, SL Neuhausen, ML Slattery, CP Garner, YC Ding, M Hoffman, AR Brothman, RR Reams, KR Kalari, H Wang, FT Odedina, KF Soliman, C Yates, JZ Song, C Stirzaker, J Harrison, JR Melki, SJ Clark, MW Coolen, C Stirzaker, JZ Song, AL Statham, Z Kassir, CS Moreno, M Makrides, RA Gibson, AJ McPhee, L Yelland, J Quinlivan, P Ryan, MG Lawrence, RA Taylor, R Toivanen, J Pedersen, S Norden,

DW Pook, SJ Clark, A Statham, C Stirzaker, PL Molloy, M Frommer, A Auton, LD Brooks, RM Durbin, EP Garrison, HM Kang, and WJ Kent. Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. *Genome Biol.*, 17(1):208, 2016. ISSN 1474-760X. doi: 10.1186/s13059-016-1066-1.

Eric M. Pietras, Damien Reynaud, Yoon A. Kang, Daniel Carlin, Fernando J. Calero-Nieto, Andrew D. Leavitt, Joshua A. Stuart, Berthold Göttgens, and Emmanuelle Passegué. Functionally Distinct Subsets of Lineage-Biased Multipotent Progenitors Control Blood Production in Normal and Regenerative Conditions. *Cell Stem Cell*, 17(1):35–46, jul 2015. ISSN 18759777. doi: 10.1016/j.stem.2015.05.003.

Harold Pimentel, Nicolas L. Bray, Suzette Puente, Páll Melsted, and Lior Pachter. Differential analysis of RNA-seq incorporating quantification uncertainty. *Nat. Methods*, 14(7):687–690, 2017. ISSN 15487105. doi: 10.1038/nmeth.4324.

O S Platt, S H Orkin, G Dover, G P Beardsley, B Miller, and D G Nathan. Hydroxyurea enhances fetal hemoglobin production in sickle cell anemia. *J. Clin. Invest.*, 74:652–6, 1984. ISSN 0021-9738. doi: 10.1172/JCI111464.

Orah S. Platt. Hydroxyurea for the Treatment of Sickle Cell Anemia. *N. Engl. J. Med.*, 358(13):1362–1369, mar 2008. ISSN 0028-4793. doi: 10.1056/NEJMct0708272.

Catherine Porcher, Wojciech Swat, Karen Rockwell, Yuko Fujiwara, Frederick W. Alt, and Stuart H. Orkin. The T cell leukemia oncoprotein SCL/tal-1 is essential for development of all hematopoietic lineages. *Cell*, 86(1):47–57, 1996. ISSN 00928674. doi: 10.1016/S0092-8674(00)80076-8.

Emmanuel Pourcelot, Candice Trocme, Julie Mondet, Sébastien Bailly, Bertrand Toussaint, and Pascal Mossuz. Cytokine profiles in polycythemia vera and essential

thrombocythemia patients: Clinical implications. *Exp. Hematol.*, 42(5):360–368, 2014. ISSN 18732399. doi: 10.1016/j.exphem.2014.01.006.

Ana Puda, Jelena D. Milosevic, Tiina Berg, Thorsten Klampfl, Ashot S. Harutyunyan, Bettina Gisslinger, Elisa Rumi, Daniela Pietra, Luca Malcovati, Chiara Elena, Michael Doubek, Michael Steurer, Natasa Tosic, Sonja Pavlovic, Paola Guglielmelli, Lisa Pieri, Alessandro M. Vannucchi, Heinz Gisslinger, Mario Cazzola, and Robert Kralovics. Frequent deletions of JARID2 in leukemic transformation of chronic myeloid malignancies. *Am. J. Hematol.*, 87(3):245–250, 2012. ISSN 03618609. doi: 10.1002/ajh.22257.

Qingsong Qiu, Ping Liu, Xuemei Zhao, Chun Zhang, Donghe Li, Ruihong Zhang, and Ruibao Ren. IRF8 Regulates Cell Cycle of Hematopoietic Stem Cells. *Blood*, 126(23):2353, 2015. ISSN 0006-4971.

Hanna S. Radomska, Claudia S. Huettner, Pu Zhang, Tao Cheng, David T. Scadden, and Daniel G. Tenen. CCAAT/Enhancer Binding Protein α Is a Regulatory Switch Sufficient for Induction of Granulocytic Development from Bipotential Myeloid Progenitors. *Mol. Cell. Biol.*, 18(7):4301–4314, 1998. ISSN 0270-7306. doi: 10.1128/mcb.18.7.4301.

Shaikh M. Rahman, Rachel C. Janssen, Mahua Choudhury, Karalee C. Baquero, Rebecca M. Aikens, Becky A. De La Houssaye, and Jacob E. Friedman. CCAAT/enhancer-binding protein β (C/EBPβ) expression regulates dietary-induced inflammation in macrophages and adipose tissue in mice. *J. Biol. Chem.*, 287(41):34349–34360, 2012. ISSN 00219258. doi: 10.1074/jbc.M112.410613.

Maria Luigia Randi, Elisabetta Ruzzon, Fabiana Tezza, Guido Luzzatto, and Fabrizio Fabris. Toxicity and side effects of hydroxyurea used for primary thrombocythemia. *Platelets*, 16(3-4):181–184, 2005. ISSN 09537104. doi: 10.1080/09537100400020179.

Jason S. Rawlings, Kristin M. Rosler, and Douglas A. Harrison. The JAK/STAT signaling pathway. *J. Cell Sci.*, 117(8):1281–1283, 2004. ISSN 0021-9533. doi: 10.1242/jcs.00963.

Shannon Reagan-Shaw, Minakshi Nihal, and Nihal Ahmad. Dose translation from animal to human studies revisited. *FASEB J.*, 22(3):659–661, 2008. ISSN 0892-6638. doi: 10.1096/fj.07-9574LSF.

Wolf Reik. Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature*, 447(7143):425–432, 2007. ISSN 14764687. doi: 10.1038/nature05918.

Damien Reynaud, Emmanuel Ravet, Monique Titeux, Frédéric Mazurier, Laurent Rénia, Anne Dubart-Kupperschmitt, Paul Henri Roméo, and Françoise Pflumio. SCL/TAL1 expression level regulates human hematopoietic stem cell self-renewal and engraftment. *Blood*, 106(7):2318–2328, 2005. ISSN 00064971. doi: 10.1182/blood-2005-02-0557.

Anne Ricksten, Lars Palmqvist, Peter Johansson, and Björn Andreasson. Rapid decline of JAK2V617F levels during hydroxyurea treatment in patients with polycythemia vera and essential thrombocythemia. *Haematologica*, 93(8):1260–1261, 2008. ISSN 03906078. doi: 10.3324/haematol.12801.

Michael A Rieger and Timm Schroeder. Hematopoiesis. pages 1–18, 2012.

Matthew E. Ritchie, Belinda Phipson, Di Wu, Yifang Hu, Charity W. Law, Wei Shi, and Gordon K. Smyth. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.*, 43(7):e47, 2015. ISSN 13624962. doi: 10.1093/nar/gkv007.

Lorraine Robb, Ian Lyons, Ruili Li, Lynne Hartley, Frank Köntgen, Richard P. Harvey, Donald Metcalf, and C. Glenn Begley. Absence of yolk sac hematopoiesis

from mice with a targeted disruption of the scl gene. *Proc. Natl. Acad. Sci. U. S. A.*, 92(15):7075–7079, 1995. ISSN 00278424. doi: 10.1073/pnas.92.15.7075.

A. Roller, V. Grossmann, U. Bacher, F. Poetzinger, S. Weissmann, N. Nadarajah, L. Boeck, W. Kern, C. Haferlach, S. Schnittger, T. Haferlach, and A. Kohlmann. Landmark analysis of DNMT3A mutations in hematological malignancies, 2013. ISSN 08876924.

Nathan R. Rose and Robert J. Klose. Understanding the relationship between DNA methylation and histone lysine methylation. *Biochim. Biophys. Acta - Gene Regul. Mech.*, 1839(12):1362–1372, 2014. ISSN 18764320. doi: 10.1016/j.bbagrm.2014.02.007.

Frank Rosenbauer and Daniel G. Tenen. Transcription factors in myeloid development: Balancing differentiation with transformation. *Nat. Rev. Immunol.*, 7(2):105–117, 2007. ISSN 14741733. doi: 10.1038/nri2024.

F. Rosenthal, L. Wislicki, and L. Kollek. Über die Beziehungen von Schwersten Blutgiften zu Abbauprodukten des Eiweisses. *Klin. Wochenschr.*, 7(21):972–977, may 1928. ISSN 0023-2173. doi: 10.1007/BF01716922.

Ellen V. Rothenberg, Hiroyuki Hosokawa, and Jonas Ungerbäck. Mechanisms of action of hematopoietic transcription factor PU.1 in initiation of T-cell development. *Front. Immunol.*, 10(FEB):1–23, 2019. ISSN 16643224. doi: 10.3389/fimmu.2019.00228.

Yannick Saintigny, Fabien Delaco, Fabrice Petitot, Sarah Lambert, Dietrich Averbeck, and Bernard S Lopez. Characterization of homologous recombination induced by replication inhibition in mammalian cells. *EMBO J.*, 20(14):3861–3870, 2001.

Juan Sandoval, Holger A. Heyn, Sebastian Moran, Jordi Serra-Musach, Miguel A. Pujana, Marina Bibikova, and Manel Esteller. Validation of a DNA methylation

microarray for 450,000 CpG sites in the human genome. *Epigenetics*, 6(6):692–702, 2011. ISSN 15592308. doi: 10.4161/epi.6.6.16196.

Alejandra Sanjuan-Pla, Iain C Macaulay, Christina T Jensen, Petter S Woll, Tiago C Luis, Adam Mead, Susan Moore, Cintia Carella, Sahoko Matsuoka, Tiphaine Bouriez Jones, Onima Chowdhury, Laura Stenson, Michael Lutteropp, Joanna C.A. Green, Raffaella Facchini, Hanane Boukarabila, Amit Grover, Adriana Gambardella, Supat Thongjuea, Joana Carrelha, Paul Tarrant, Deborah Atkinson, Sally Ann Clark, Claus Nerlov, and Sten Eirik W. Jacobsen. Platelet-biased stem cells reside at the apex of the haematopoietic stem-cell hierarchy. *Nature*, 502(7470):232–236, 2013. ISSN 00280836. doi: 10.1038/nature12495.

M Sasaki, C B Knobbe, J C Munger, E F Lind, D Brenner, A Brustle, I S Harris, R Holmes, A Wakeham, J Haight, A You-Ten, W Y Li, S Schalm, S M Su, C Virtanen, G Reifenberger, P S Ohashi, D L Barber, M E Figueroa, A Melnick, J C Zuniga-Pflucker, and T W Mak. IDH1(R132H) mutation increases murine haematopoietic progenitors and alters epigenetics. *Nature*, 488(7413):656–659, 2012. ISSN 1476-4687. doi: 10.1038/nature11323.

Yogen Saunthararajah, Cheryl A. Hillery, Don Lavelle, Robert Molokie, Louise Dorn, Linda Bressler, Stefana Gavazova, Yi Hsiang Chen, Ronald Hoffman, and Joseph DeSimone. Effects of 5-aza-2'-deoxycytidine on fetal hemoglobin levels, red cell adhesion, and hematopoietic differentiation in patients with sickle cell disease. *Blood*, 102(12):3865–3870, 2003. ISSN 00064971. doi: 10.1182/blood-2003-05-1738.

Guy Sauvageau, Unnur Thorsteinsdottir, Connie J. Eaves, H. Jeffrey Lawrence, Cory Largman, Peter M. Lansdorp, and R. Keith Humphries. Overexpression of HOXB4 in hematopoietic cells causes the selective expansion of more primitive populations in vitro and in vivo. *Genes Dev.*, 9(14):1753–1765, 1995. ISSN 08909369. doi: 10.1101/gad.9.14.1753.

Jochen Schmitz, Manuela Weissenbach, Serge Haan, Peter C. Heinrich, and Fred Schaper. SOCS3 exerts its inhibitory function on interleukin-6 signal transduction through the SHP2 recruitment site of gp130. *J. Biol. Chem.*, 275(17):12848–12856, 2000. ISSN 00219258. doi: 10.1074/jbc.275.17.12848.

Medical Sciences, M Bhatia, J. C. Y. Wang, U Kapp, D Bonnet, and J E Dick. Purification of primitive human hematopoietic cells capable of repopulating immune-deficient mice. *Proc. Natl. Acad. Sci.*, 94(10):5320–5325, may 1997. ISSN 0027-8424. doi: 10.1073/pnas.94.10.5320.

Edward W. Scott, M. Celeste Simon, John Anastasi, and Harinder Singh. Requirement of transcription factor PU.1 in the development of multiple hematopoietic lineages. *Science (80-.).*, 265(5178):1573–1577, 1994. ISSN 00368075. doi: 10.1126/science.8079170.

Linda M Scott, Peter J Campbell, Clare East, Nasios Fourouclas, Soheila Swanton, George S Vassiliou, Anthony J Bench, Elaine M Boyd, Natasha Curtin, Mike a Scott, Wendy N Erber, Anthony R Green, E Joanna Baxter, Linda M Scott, Peter J Campbell, Clare East, Nasios Fourouclas, Soheila Swanton, George S Vassiliou, Anthony J Bench, Elaine M Boyd, Natasha Curtin, Mike a Scott, Wendy N Erber, and Anthony R Green. Acquired mutation of the tyrosine kinase JAK2 in human myeloproliferative disorders. *Lancet*, 365(9464):1054–1061, mar 2005. ISSN 01406736. doi: 10.1016/S0140-6736(05)71142-9.

Linda M Scott, D Ph, Wei Tong, D Ph, Ross L Levine, Mike a Scott, D Ph, a Philip, M R C Path, Alan J Warren, F R C Path, F Med Sci, and D Gary Gilliland. JAK2 Exon 12 Mutations in Polycythemia Vera and Idiopathic Erythrocytosis. 356(5): 459–468, 2010. doi: 10.1056/NEJMoa065202.JAK2.

I. Screpanti, L. Romani, P. Musiani, A. Modesti, E. Fattori, D. Lazzaro, C. Sellitto, S. Scarpa, D. Bellavia, and G. Lattanzio. Lymphoproliferative disorder and

imbalanced T-helper response in C/EBP beta-deficient mice. *EMBO J.*, 14(9): 1932–1941, 1995. doi: 10.1002/j.1460-2075.1995.tb07185.x.

Takenori Seki, Takashi Kumagai, Bethel Kwansa-Bentum, Rieko Furushima-Shimogawara, William K. Anyan, Yuuki Miyazawa, Yoichiro Iwakura, and Nobuo Ohta. Interleukin-4 (IL-4) and IL-13 suppress excessive neutrophil infiltration and hepatocyte damage during acute murine schistosomiasis japonica. *Infect. Immun.*, 80(1):159–168, 2012. ISSN 00199567. doi: 10.1128/IAI.05581-11.

Teresa Sevivas, José María Bastida, David S. Paul, Eva Caparros, Verónica Palma-Barqueros, Margarida Coucelo, Dalila Marques, Francisca Ferrer-Marín, José Ramón González-Porras, Vicente Vicente, Jesús María Hernández-Rivas, Steve P. Watson, María Luisa Lozano, Wolfgang Bergmeier, and José Rivera. Identification of two novel mutations in RASGRP2 affecting platelet CalDAG-GEFI expression and function in patients with bleeding diathesis. *Platelets*, 29 (2):192–195, 2018. ISSN 13691635. doi: 10.1080/09537104.2017.1336214.

Shireen Sirhan, Terra L. Lasho, Curtis A. Hanson, Ruben A. Mesa, Animesh Pardanani, and Ayalew Tefferi. The presence of JAK2V617F in primary myelofibrosis or its allele burden in polycythemia vera predicts chemosensitivity to hydroxyurea. *Am. J. Hematol.*, 83(5):363–365, 2008. ISSN 03618609. doi: 10.1002/ajh.21149.

José M. Sogo, Massimo Lopes, and Marco Foiani. Fork reversal and ssDNA accumulation at stalled replication forks owing to checkpoint defects. *Science (80-.).*, 297(5581):599–602, 2002. ISSN 00368075. doi: 10.1126/science.1074023.

Gerald J Spangrude, Shelly Heimfeld, and Irving L Weissman. Hematopoietic Stem. *Science (80-.).*, pages 58–62, 1988.

Emmanouil Spanoudakis, Ioanna Bazdiara, Ioannis Kotsianidis, Dimitrios Margaritis, Aggelos Goutzouvelidis, Anna Christoforidou, Costas Tsatalas, and George Bourikas. Hydroxyurea (HU) is effective in reducing JAK2V617F mutated clone

size in the peripheral blood of essential thrombocythemia (ET) and polycythemia vera (PV) patients. *Ann. Hematol.*, 88(7):629–632, 2009. ISSN 09395555. doi: 10.1007/s00277-008-0650-1.

Stanley Spier. HYDROXYUREA AND MACROCYTOSIS. *Arch. Dermatol.*, 104 (5):564, nov 1971. ISSN 0003-987X. doi: 10.1001/archderm.1971.04000230106023.

Duncan Sproul, Robert R. Kitchen, Colm E. Nestor, J. Michael Dixon, Andrew H. Sims, David J. Harrison, Bernard H. Ramsahoye, and Richard R. Meehan. Tissue of origin determines cancer-associated CpG island promoter hypermethylation patterns. *Genome Biol.*, 13(10):R84, 2012. ISSN 1474760X. doi: 10.1186/gb-2012-13-10-R84.

Barbara Stearns, Kathryn A. Losee, and Jack Bernstein. Hydroxyurea. A New Type of Potential Antitumor Agent. *J. Med. Chem.*, 6(2):201, 1963. ISSN 15204804. doi: 10.1021/jm00338a026.

David P. Steensma, Rafael Bejar, Siddhartha Jaiswal, R. Coleman Lindsley, Mikkael A. Sekeres, Robert P. Hasserjian, and Benjamin L. Ebert. Clonal hematopoiesis of indeterminate potential and its distinction from myelodysplastic syndromes. *Blood*, 126(1):9–16, 2015. ISSN 15280020. doi: 10.1182/blood-2015-03-631747.

F Stegelmann, L Bullinger, R F Schlenk, P Paschka, M Griesshammer, C Blersch, S Kuhn, S Schauer, H Döhner, and K Döhner. DNMT3A mutations in myeloproliferative neoplasms. *Leuk. Off. J. Leuk. Soc. Am. Leuk. Res. Fund, U.K*, 25 (May):1217–1219, 2011. ISSN 0887-6924. doi: 10.1038/leu.2011.77.

Yvon Sterkers, Claude Preudhomme, J L Laï, Jean-Loup L Demory, M T Caulier, Eric Wattel, Dominique Bordessoule, Francis Bauters, and Pierre Fenaux. Acute myeloid leukemia and myelodysplastic syndromes following essential thrombo-

cythemia treated with hydroxyurea: high proportion of cases with 17p deletion. *Blood*, 91(2):616–22, jan 1998. ISSN 0006-4971.

Richard M. Stone, Christina Hultman, Ofir Wolach, Deya Cherpokova, Bozenna Iliadou, Benjamin L. Ebert, Dylan Adams, Martha Wadleigh, Ilene Galinsky, Kimberly Martinod, Alexander J. Silver, Rob S. Sellar, David P. Steensma, Denisa D. Wagner, Cecilia A. Castellano, Rebekka K. Schneider, Giulio Genovese, Ann Mullally, Marie McConkey, Donna Neuberg, Daniel J. DeAngelo, Steven A. McCarroll, Ryan J. Chappell, Robert F. Padera, Kimberly Martinod, Deya Cherpokova, Marie McConkey, Ryan J. Chappell, Alexander J. Silver, Dylan Adams, Cecilia A. Castellano, Rebekka K. Schneider, Robert F. Padera, Daniel J. DeAngelo, Martha Wadleigh, David P. Steensma, Ilene Galinsky, Richard M. Stone, Giulio Genovese, Steven A. McCarroll, Bozenna Iliadou, Christina Hultman, Donna Neuberg, Ann Mullally, Denisa D. Wagner, and Benjamin L. Ebert. Increased neutrophil extracellular trap formation promotes thrombosis in myeloproliferative neoplasms. *Sci. Transl. Med.*, 10(436):1–11, 2018. ISSN 19466242. doi: 10.1126/scitranslmed.aan8292.

Birgitta Swolin, Stig Rödjer, and Jan Westin. Therapy-related patterns of cytogenetic abnormalities in acute myeloid leukemia and myelodysplastic syndrome post polycythemia vera. *Ann. Hematol.*, 87(6):467–474, 2008. ISSN 09395555. doi: 10.1007/s00277-008-0461-4.

Mamta Tahiliani, Kian Peng Koh, Yinghua Shen, William A. Pastor, Hozefa Bandukwala, Yevgeny Brudno, Suneet Agarwal, Lakshminarayan M. Iyer, David R. Liu, L. Aravind, and Anjana Rao. Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. *Science (80-.).*, 324(5929):930–935, may 2009. ISSN 0036-8075. doi: 10.1126/science.1170116.

Satoru Takahashi, Kou Onodera, Hozumi Motohashi, Naruyoshi Suwabe, Norio Hayashi, Nobuaki Yanai, Yoichi Nabesima, and Masayuki Yamamoto. Arrest

in primitive erythroid cell development caused by promoter- specific disruption of the GATA-1 gene. *J. Biol. Chem.*, 272(19):12611–12615, 1997. ISSN 00219258. doi: 10.1074/jbc.272.19.12611.

Satoru Takahashi, Takuya Komeno, Naruyoshi Suwabe, Keigyo Yoh, Osamu Nakajima, Sigeko Nishimura, Takashi Kuroha, Toshiro Nagasawa, and Masayuki Yamamoto. Role of GATA-1 in proliferation and differentiation of definitive erythroid and megakaryocytic cells in vivo. *Blood*, 92(2):434–442, 1998. ISSN 00064971.

A. Tefferi, A. Pardanani, K-H H. Lim, O. Abdel-Wahab, T. L. Lasho, J. Patel, N. Gangat, C. M. Finke, S. Schwager, A. Mullally, C-Y Y. Li, C. A. Hanson, R. Mesa, O. Bernard, F. Delhommeau, W. Vainchenker, D. G. Gilliland, and R. L. Levine. TET2 mutations and their clinical correlates in polycythemia vera, essential thrombocythemia and myelofibrosis. *Leukemia*, 23(5):905–911, may 2009. ISSN 14765551. doi: 10.1038/leu.2009.47.

A Tefferi, T L Lasho, O Abdel-Wahab, P Guglielmelli, J Patel, D Caramazza, L Pieri, C M Finke, O Kilpivaara, M Wadleigh, M Mai, R F McClure, D G Gilliland, R L Levine, A Pardanani, and A M Vannucchi. IDH1 and IDH2 mutation studies in 1473 patients with chronic-, fibrotic- or blast-phase essential thrombocythemia, polycythemia vera or myelofibrosis. *Leukemia*, 24(7):1302–9, 2010. ISSN 1476-5551. doi: 10.1038/leu.2010.113.

Ayalew Tefferi, Alessandro M. Vannucchi, and Tiziano Barbui. Polycythemia vera treatment algorithm 2018. *Blood Cancer J.*, 8(1), 2018a. ISSN 20445385. doi: 10.1038/s41408-017-0042-7.

Ayalew Tefferi, Alessandro M. Vannucchi, and Tiziano Barbui. Essential thrombocythemia treatment algorithm 2018. *Blood Cancer J.*, 8(1):4–9, 2018b. ISSN 20445385. doi: 10.1038/s41408-017-0041-8.

Luciana Teofili, Maurizio Martini, Tonia Cenci, Giovanna Petrucci, Lorenza Torti,

Sergio Storti, Francesco Guidi, Giuseppe Leone, and Luigi Maria Larocca. Different STAT-3 and STAT-5 phosphorylation discriminates among Ph-negative chronic myeloproliferative diseases and is independent of the V617F JAK-2 mutation. *Blood*, 110(1):354–359, 2007. ISSN 00064971. doi: 10.1182/blood-2007-01-069237.

Andrew E. Teschendorff, Francesco Marabita, Matthias Lechner, Thomas Bartlett, Jesper Tegner, David Gomez-cabrero, and Stephan Beck. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics*, 29(2):189–196, 2013. ISSN 13674803. doi: 10.1093/bioinformatics/bts680.

Unnur Thorsteinsdottir, Guy Sauvageau, and R. K. Humphries. Enhanced in vivo regenerative potential of HOXB4-transduced hematopoietic stem cells with regulation of their pool size. *Blood*, 94(8):2605–2612, 1999. ISSN 00064971.

W G Thurman, C Bloedow, C D Howe, W C Levin, P Davis, M Lane, M P Sullivan, and K M Griffith. A phase I study of hydroxyurea. *Cancer Chemother. reports*, 29:103–107, may 1963. ISSN 0069-0112 (Print).

Ralph Tiedt, Hui Hao-Shen, Marta A. Sobas, Renate Looser, Stephan Dirnhofer, Jürg Schwaller, and Radek C. Skoda. Ratio of mutant JAK2-V617F to wild-type Jak2 determines the MPD phenotypes in transgenic mice. *Blood*, 111(8):3931–3940, 2008. ISSN 00064971. doi: 10.1182/blood-2007-08-107748.

Jennifer J. Trowbridge, Jonathan W. Snow, Jonghwan Kim, and Stuart H. Orkin. DNA Methyltransferase 1 Is Essential for and Uniquely Regulates Hematopoietic Stem and Progenitor Cells. *Cell Stem Cell*, 5(4):442–449, 2009. ISSN 19345909. doi: 10.1016/j.stem.2009.08.016.

Cristina Tufarelli, Jackie A Sloane Stanley, David Garrick, Jackie A Sharpe, Helena Ayyub, William G Wood, and Douglas R Higgs. Transcription of antisense RNA

leading to gene silencing and methylation as a novel cause of human genetic disease. *Nat. Genet.*, 34(2):157–165, jun 2003. ISSN 1061-4036. doi: 10.1038/ng1157.

IL Uchida, N, and Weissman. Searching for Hematopoietic Stem Cells: Evidence That Thy-l.1 lo Lin- Sea-1 + Cells Are the Only Stem Cells in C57BL/Ka-Thy-l.1 Bone Marrow By Nobuko Uchida and Irving L. Weissman. *J. Exp. Med*, 175 (January):175–184, 1992.

Valérie Ugo, Chloé James, and William Vainchenker. Une mutation unique de la protéine kinase JAK2 dans la polyglobulie de Vaquez et les syndromes myéloprolifératifs non-LMC. *médecine/sciences*, 21(6-7):669–670, jun 2005. ISSN 0767-0974. doi: 10.1051/medsci/2005216-7669.

Kris Vaddi, Phillip Liu, Taghi Manshouri, Jun Li, Peggy A Scherle, Eian Caulder, Xiaoming Wen, Yanlong Li, Paul Waeltz, Mark Rupar, Timothy Burn, Yvonne Lo, Jennifer Kelley, Maryanne Covington, Stacey Shepard, James D Rodgers, Patrick Haley, Hagop Kantarjian, Jordan S Fridman, and Srdan Verstovsek. Preclinical characterization of the selective JAK1 / 2 inhibitor INCB018424 : therapeutic implications for the treatment of myeloproliferative neoplasms. *Blood*, 115(15): 3109–3117, 2010. doi: 10.1182/blood-2009-04-214957.The.

Rakhee Vaidya, Naseema Gangat, Thitina Jimma, Christy M. Finke, Terra L. Lasho, Animesh Pardanani, and Ayalew Tefferi. Plasma cytokines in polycythemia vera: Phenotypic correlates, prognostic relevance, and comparison with myelofibrosis. *Am. J. Hematol.*, 87(11):1003–1005, 2012. ISSN 03618609. doi: 10.1002/ajh.23295.

Victoria Valinluck and Lawrence C. Sowers. Endogenous cytosine damage products alter the site selectivity of human DNA maintenance methyltransferase DNMT1. *Cancer Res.*, 67(3):946–950, 2007. ISSN 00085472. doi: 10.1158/0008-5472.CAN-06-3123.

Juan M. Vaquerizas, Sarah K. Kummerfeld, Sarah A. Teichmann, and Nicholas M. Luscombe. A census of human transcription factors: Function, expression and evolution. *Nat. Rev. Genet.*, 10(4):252–263, 2009. ISSN 14710056. doi: 10.1038/nrg2538.

Aisha L Walker, Shirley Steward, Thad a Howard, Nicole Mortier, Matthew Smeltzer, Yong-dong Wang, and Russell E Ware. Epigenetic and molecular profiles of erythroid cells after hydroxyurea treatment in sickle cell anemia. *Blood*, 118(20):5664–70, nov 2011. ISSN 1528-0020. doi: 10.1182/blood-2011-07-368746.

Chelsia Qiuxia Wang, Lena Motoda, Masanobu Satake, Yoshiaki Ito, Ichiro Taniuchi, Vinay Tergaonkar, and Motomi Osato. Runx3 deficiency results in myeloproliferative disorder in aged mice. *Blood*, 122(4):562–566, 2013. ISSN 15280020. doi: 10.1182/blood-2012-10-460618.

Haijun Wang, Chunhua Song, Yali Ding, Xiaokang Pan, Zheng Ge, Bi Hua Tan, Chandrika Gowda, Mansi Sachdev, Sunil Muthusami, Hongsheng Ouyang, Liangxue Lai, Olivia L. Francis, Christopher L. Morris, Hisham Abdel-Azim, Glenn Dorsam, Meixian Xiang, Kimberly J. Payne, and Sinisa Dovat. Transcriptional regulation of JARID1B/KDM5B histone demethylase by ikaros, histone deacetylase 1 (HDAC1), and casein kinase 2 (CK2) in B-cell acute lymphoblastic leukemia. *J. Biol. Chem.*, 291(8):4004–4018, 2016. ISSN 1083351X. doi: 10.1074/jbc.M115.679332.

Xuefeng Wang, Xiaoqing Yu, Wei Zhu, W. Richard Mccombie, Eric Antoniou, R. Scott Powers, Nicholas O. Davidson, Ellen Li, and Jennie Williams. A trimming-and-retrieving alignment scheme for reduced representation bisulfite sequencing. *Bioinformatics*, 31(12):2040–2042, jun 2015. ISSN 14602059. doi: 10.1093/bioinformatics/btv089.

Zhenxing Wang, Yanshuo Chu, Yongtian Wang, and Yadong Wang. A framework for

analyzing DNA methylation data from Illumina Infinium HumanMethylation450 BeadChip. *Proc. - 2017 IEEE Int. Conf. Bioinforma. Biomed. BIBM 2017*, 2017-Janua(Suppl 5):1688–1693, 2017. ISSN 1471-2105. doi: 10.1109/BIBM.2017.8217914.

Zhi Chao Wang, Myoung Kwon Choi, Tatsuma Ban, Hideyuki Yanai, Hideo Negishi, Yan Lu, Tomohiko Tamura, Akinori Takaoka, Kazuko Nishikura, and Tadatsugu Taniguchi. Regulation of innate immune responses by DAI (DLM-1/ZBP1) and other DNA-sensing molecules. *Proc. Natl. Acad. Sci. U. S. A.*, 105(14):5477–5482, 2008. ISSN 00278424. doi: 10.1073/pnas.0801295105.

Warren K. Sinclair. The combined lethal effect of hydroxyurea and x-rays on Chinese hamster cells in vitro. ANL-7409. *ANL [reports]. U. S. At. Energy Comm.*, 28 (February):3–4, 1967.

Michael Weber, Ines Hellmann, Michael B. Stadler, Liliana Ramos, Svante Pääbo, Michael Rebhan, and Dirk Schübeler. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat. Genet.*, 39(4):457–466, 2007. ISSN 10614036. doi: 10.1038/ng1990.

John J Welch, Jason A Watts, Christopher R Vakoc, Yu Yao, Hao Wang, Ross C Hardison, Gerd A Blobel, Lewis A Chodosh, and Mitchell J Weiss. Global regulation of erythroid gene expression by transcription factor GATA-1. *Blood*, 104(10): 3136–3147, nov 2004. ISSN 0006-4971 (Print). doi: 10.1182/blood-2004-04-1603.

Gerlinde Wernig. Expression of Jak2V617F causes a polycythemia vera-like disease with associated myelofibrosis in a murine bone marrow transplant model. *Blood*, 107(11):4274–4281, jun 2006. ISSN 0006-4971. doi: 10.1182/blood-2005-12-4824.

Elizabeth A. Williamson, Haixin N. Xu, Adrian F. Gombart, Walter Verbeek, Alexey M. Chumakov, Alan D. Friedman, and H. Phillip Koeffler. Identification

of transcriptional activation and repression domains in human CCAAT/enhancer-binding protein ϵ. *J. Biol. Chem.*, 273(24):14796–14804, 1998. ISSN 00219258. doi: 10.1074/jbc.273.24.14796.

Anne Wilson, Mark J. Murphy, Thordur Oskarsson, Konstantinos Kaloulis, Michael D. Bettess, Gabriela M. Oser, Anne Catherine Pasche, Christian Knabenhans, H. Robson MacDonald, and Andreas Trumpp. c-Myc controls the balance between hematopoietic stem cell self-renewal and differentiation. *Genes Dev.*, 18 (22):2747–2763, 2004. ISSN 08909369. doi: 10.1101/gad.313104.

Janine Woytschak, Nadia Keller, Carsten Krieg, Daniela Impellizzieri, Robert W. Thompson, Thomas A. Wynn, Annelies S. Zinkernagel, and Onur Boyman. Type 2 Interleukin-4 Receptor Signaling in Neutrophils Antagonizes Their Expansion and Migration during Infection and Inflammation. *Immunity*, 45(1):172–184, 2016. ISSN 10974180. doi: 10.1016/j.immuni.2016.06.025.

Wei Xu, Hui Yang, Ying Liu, Ying Yang, Ping Wang, Se Hee Kim, Shinsuke Ito, Chen Yang, Pu Wang, Meng Tao Xiao, Li Xia Liu, Wen Qing Jiang, Jing Liu, Jin Ye Zhang, Bin Wang, Stephen Frye, Yi Zhang, Yan Hui Xu, Qun Ying Lei, Kun Liang Guan, Shi Min Zhao, and Yue Xiong. Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of α-ketoglutarate-dependent dioxygenases. *Cancer Cell*, 19(1):17–30, 2011. ISSN 15356108. doi: 10.1016/j.ccr.2010.12.014.

Zongli Xu, Liang Niu, Leping Li, and Jack A. Taylor. ENmix: A novel background correction method for Illumina HumanMethylation450 BeadChip. *Nucleic Acids Res.*, 44(3):1–6, 2015. ISSN 13624962. doi: 10.1093/nar/gkv907.

Ryo Yamamoto, Yohei Morita, Jun Ooehara, Sanae Hamanaka, Masafumi Onodera, Karl Lenhard Rudolph, Hideo Ema, and Hiromitsu Nakauchi. Clonal Analysis Unveils Self-Renewing Lineage-Restricted Progenitors Generated Directly from

Hematopoietic Stem Cells. *Cell*, 154(5):1112–1126, aug 2013. ISSN 00928674. doi: 10.1016/j.cell.2013.08.007.

Ryuya Yamanaka, Carrolee Barlow, Julie Lekstrom-Himes, Lucio H. Castilla, Pu P. Liu, Michael Eckhaus, Thomas Decker, Anthony Wynshaw-Boris, and Kleanthis G. Xanthopoulos. Impaired granulopoiesis, myelodysplasia, and early lethality in CCAAT/enhancer binding protein ϵ-deficient mice. *Proc. Natl. Acad. Sci. U. S. A.*, 94(24):13187–13192, 1997. ISSN 00278424. doi: 10.1073/pnas.94.24.13187.

Hui Yang, Hui Liang, Jing Song Yan, Rong Tao, Si Guo Hao, and Li Yuan Ma. Down-regulation of hematopoiesis master regulator PU.1 via aberrant methylation in chronic myeloid leukemia. *Int. J. Hematol.*, 96(1):65–73, 2012. ISSN 09255710. doi: 10.1007/s12185-012-1106-x.

Hideo Yasukawa, Atsuo Sasaki, and Akihiko Yoshimura. Negative Regulation of Cytokine Signaling Pathways. *Annu. Rev. Immunol.*, 18(1):143–164, 2000. ISSN 0732-0582. doi: 10.1146/annurev.immunol.18.1.143.

A. Yoshimura, T. Ohkubo, T. Kiguchi, N.A. Jenkins, D.J. Gilbert, N.G. Copeland, T. Hara, and A. Miyajima. A novel cytokine-inducible gene CIS encodes an SH2-containing protein that binds to tyrosine-phosphorylated interleukin 3 and erythropoietin receptors. *EMBO J.*, 14(12):2816–2826, 1995. doi: 10.1002/j.1460-2075.1995.tb07281.x.

C. W. Young and S. Hodas. Hydroxyurea: Inhibitory Effect on DNA Metabolism. *Science (80-.).*, 146(3648):1172–1174, nov 1964. ISSN 0036-8075. doi: 10.1126/science.146.3648.1172.

Raymond Yick-Loi Yu, Xing Wang, Fiona J Pixley, J Jessica Yu, Alexander L Dent, Hal E Broxmeyer, E Richard Stanley, and B Hilda Ye. BCL-6 negatively regulates macrophage proliferation by suppressing autocrine IL-6 production. *Blood*, 105(4):1777–1784, feb 2005. ISSN 0006-4971 (Print). doi: 10.1182/blood-2004-08-3171.

Virginia M. Zaleskas, Daniela S. Krause, Katherine Lazarides, Nihal Patel, Yiguo Hu, Shaoguang Li, Richard A Van Etten, and Richard A. Van Etten. Molecular pathogenesis and therapy of polycythemia induced in mice by JAK2 V617F. *PLoS One*, 1(1), 2006. ISSN 19326203. doi: 10.1371/journal.pone.0000018.

D.-E. Zhang, Pu Zhang, N.-d. Wang, Christopher J. Hetherington, Gretchen J. Darlington, and Daniel G. Tenen. Absence of granulocyte colony-stimulating factor signaling and neutrophil development in CCAAT enhancer binding protein -deficient mice. *Proc. Natl. Acad. Sci.*, 94(2):569–574, jan 1997. ISSN 0027-8424. doi: 10.1073/pnas.94.2.569.

Pu Zhang, Junko Iwasaki-Arai, Hiromi Iwasaki, Maris L. Fenyus, Tajhal Dayaram, Bronwyn M. Owens, Hirokazu Shigematsu, Elena Levantini, Claudia S. Huettner, Julie A. Lekstrom-Himes, Koichi Akashi, and Daniel G. Tenen. Enhancement of hematopoietic stem cell repopulating capacity and self-renewal in the absence of the transcription factor C/EBPα. *Immunity*, 21(6):853–863, 2004. ISSN 10747613. doi: 10.1016/j.immuni.2004.11.006.

Yan Zhang, Huirong Yang, Xue Guo, Naiyan Rong, Yujiao Song, Youwei Xu, Wenxian Lan, Xu Zhang, Maili Liu, Yanhui Xu, and Chunyang Cao. The PHD1 finger of KDM5B recognizes unmodified H3K4 during the demethylation of histone H3K4me2/3 by KDM5B. *Protein Cell*, 5(11):837–850, 2014. ISSN 16748018. doi: 10.1007/s13238-014-0078-4.

Dewang Zhou, Kaimao Liu, Chiao Wang Sun, Kevin M. Pawlik, and Tim M. Townes. KLF1 regulates BCL11A expression and γ- to β-globin gene switching. *Nat. Genet.*, 42(9):742–744, 2010. ISSN 15461718. doi: 10.1038/ng.637.

Appendix A

Clinical parameters

Table. A.1. Clinical parameters from patients per collection time point and genome wide data included in the study.

Sample ID	Time Point	Age	Sex	Diagnosis	Molecular Diagnosis	Hct (%)	MVC (fL)	Platelets (x10^9/L)	WBC (x10^9/L)	Neutrophil (x10^9/L)	Hb (g/L)	RNA seq data		DNA meth data	
												CD34	NEU	CD34	NEU
P002	Pre-tto	50	F	PV	JAK2	0.43	92	336	5.3	3.6	143	Yes*	Yes	Yes*	Yes
P002	6 mo	51	F	PV	JAK2	0.427	94	267	5.4	3.6	141				Yes
P002	9 mo	51	F	PV	JAK2	0.389	94	219	4.3	2.6	130	Yes*	Yes	Yes	Yes
P003	Pre-tto	55	F	ET	JAK2	0.424	91	661	10.9	7.3	135	Yes*	Yes	Yes	Yes
P003	3 mo	55	F	ET	JAK2	0.399	98	369	7.6	4.7	134				Yes
P003	6 mo	55	F	ET	JAK2	0.396	103	350	8	5	131				Yes
P003	9 mo	56	F	ET	JAK2	0.387	105	259	6.8	4.4	131	Yes*	Yes	Yes	Yes
P004	Pre-tto	42	F	ET	TN	0.451	90	628	11.3	6.8	144	Yes*	Yes*	Yes	Yes
P004	3 mo	43	F	ET	TN	0.43	93	597	9.1	5.3	147				Yes
P004	9 mo	43	F	ET	TN	0.409	93	592	8.4	4.6	137	Yes*	Yes*	Yes	Yes
P006	Pre-tto	67	F	ET	JAK2	0.422	85	534	7.6	4.4	136	Yes*	Yes		Yes
P006	3 mo	67	F	ET	JAK2	0.397	88	479	5.5	3.4	133				Yes
P006	6 mo	68	F	ET	JAK2	0.435	92	439	5.4	3.1	139				Yes
P006	9 mo	68	F	ET	JAK2	0.415	91	430	6.6	4.2	144	Yes*	Yes		Yes
P007	Pre-tto	77	F	ET	JAK2	0.403	83	529	8.6	2.8	129	Yes	Yes		Yes
P007	3 mo	78	F	ET	JAK2	0.409	88	427	7.1	2.7	125				Yes*
P007	6 mo	78	F	ET	JAK2	0.389	95	487	7.7	3.4	130				Yes
P007	9 mo	78	F	ET	JAK2	0.394	94	466	6.6	2.2	134	Yes	Yes		Yes
P008	Pre-tto	88	F	ET	JAK2	0.416	82	728	10.6	7	133	Yes	Yes		Yes
P008	3 mo	88	F	ET	JAK2	0.384	92	359	6.9	4.4	127				Yes
P008	6 mo	88	F	ET	JAK2	0.38	93	351	6.3	3.5	123				Yes
P008	9 mo	89	F	ET	JAK2	0.406	93	389	6.6	4	134	Yes	Yes		Yes
P009	Pre-tto	68	M	PV	JAK2	0.455	82	635	16.6	13.6	148	Yes	Yes		Yes
P009	3 mo	69	M	PV	JAK2	0.45	83	641	13.1	10.7	144				Yes
P009	6 mo	69	M	PV	JAK2	0.453	87	423	9.8	7.8	145				Yes
P009	9 mo	69	M	PV	JAK2	0.447	98	369	8.2	6.4	153	Yes	Yes		Yes
P010	Pre-tto	77	F	ET	TN	0.386	92	787	6.4	4.1	127	Yes			Yes
P010	3 mo	77	F	ET	TN	0.389	101	563	4.5	2.3	128				Yes
P010	6 mo	78	F	ET	TN	0.388	105	642	3.5	2.2	127				Yes
P010	9 mo	78	F	ET	TN	0.376	107	522	8.3	5.9	126	Yes			Yes
P011	Pre-tto	52	M	PV	JAK2	0.427	67	699	12.6	8.7	136	Yes	Yes		Yes
P011	3 mo	52	M	PV	JAK2	0.448	70	355	9	6	144				Yes
P011	9 mo	53	M	PV	JAK2	0.486	76	209	6.6	3.8	158	Yes	Yes		Yes

*Bad quality and excluded from analysis

Appendix B

Data generation human

B.1 RNA sequencing: library preparation and sequencing results

B.2 RNA sequencing: FastQC results after sequencing in HiSeq4000 (raw reads)

B.3 RNAseqMetrics results

B.4 Removal of duplicated reads

B.5 DNA methylation normalization

Table. B.1. Details of the library preparation steps for human samples. Details of the steps modified according to the RNA input and quality (RIN). Library concentration after library preparation, total nM obtained and total nM used for sequencing.

Patient	Type	Sample	Input RNA (ng)	Frag. time (min)	PCR2 n°cycles	Library ng/µL Qubit	(Q) ng/µL Diluted	(T) Average Size [bp]	Q+T (nM)	nM (sequenced)
P002	CD34+	S01	1.1	2	15	34.8	1.19	407	4.43	4
P002	CD34+	S04	1.1	NA	15	1.37	1.37	467	4.44	4
P003	CD34+	S01	1.7	4	15	68.4	1.95	382	7.73	4
P003	CD34+	S04	1.7	NA	15	2.68	1.58	465	5.15	4
P004	CD34+	S01	0.95	NA	15	3.4	1.64	418	5.94	4
P004	CD34+	S04	0.95	NA	15	0.999	0.999	483	3.13	3.2
P006	CD34+	S01	2.7	4	15	112	1.48	387	5.79	4
P006	CD34+	S04	2.7	NA	15	1.35	1.35	446	4.59	4
P007	CD34+	S01	1.8	4	15	96.8	1.49	390	5.79	4
P007	CD34+	S04	1.8	4	15	69.2	1.88	389	7.32	4
P008	CD34+	S01	3.5	4	15	70.2	1.69	362	7.07	4
P008	CD34+	S04	3.5	4	15	106	2.48	398	9.44	4
P009	CD34+	S01	5	4	13	80.6	2.4	407	8.93	4
P009	CD34+	S04	5	4	13	83	2.48	393	9.56	4
P010	CD34+	S01	2.3	4	15	100	2.32	381	9.23	4
P010	CD34+	S04	2.3	4	15	78.8	2.46	374	9.97	4
P011	CD34+	S01	1.4	4	15	78.8	2.26	379	9.03	4
P011	CD34+	S04	1.4	4	15	59.8	2.18	353	9.36	4
P002	NEU	S01	5	3	13	27.8	1.47	510	4.37	4
P002	NEU	S04	5	4	13	91.6	1.74	416	6.34	4
P003	NEU	S01	1.9	4	15	16.2	1.64	333	7.46	4
P003	NEU	S04	1.9	4	15	64.4	1.45	416	5.28	4
P004	NEU	S01	2	4	15	1.55	1.55	366	6.42	4
P004	NEU	S04	2	4	15	85	1.33	403	5	4
P006	NEU	S01	5	4	13	5.62	1.67	370	6.84	4
P006	NEU	S04	5	4	13	55.4	1.74	432	6.1	4
P007	NEU	S01	5	4	13	11.9	1.46	377	5.87	4
P007	NEU	S04	5	4	13	67.2	1.37	423	4.91	4
P008	NEU	S01	1.8	4	15	72.2	1.58	390	6.14	4
P008	NEU	S04	1.8	4	15	44.2	1.54	423	5.52	4
P009	NEU	S01	4.8	4	13	102	2.96	445	10.08	4
P009	NEU	S04	4.8	4	13	too high	1.73	494	5.31	4
P011	NEU	S01	5	4	13	91.4	2.06	392	7.96	4
P011	NEU	S04	5	4	13	69.6	1.68	376	6.77	4

Q+T= library quantification using method described in 2.4.4
NA: no fragmentation step performed

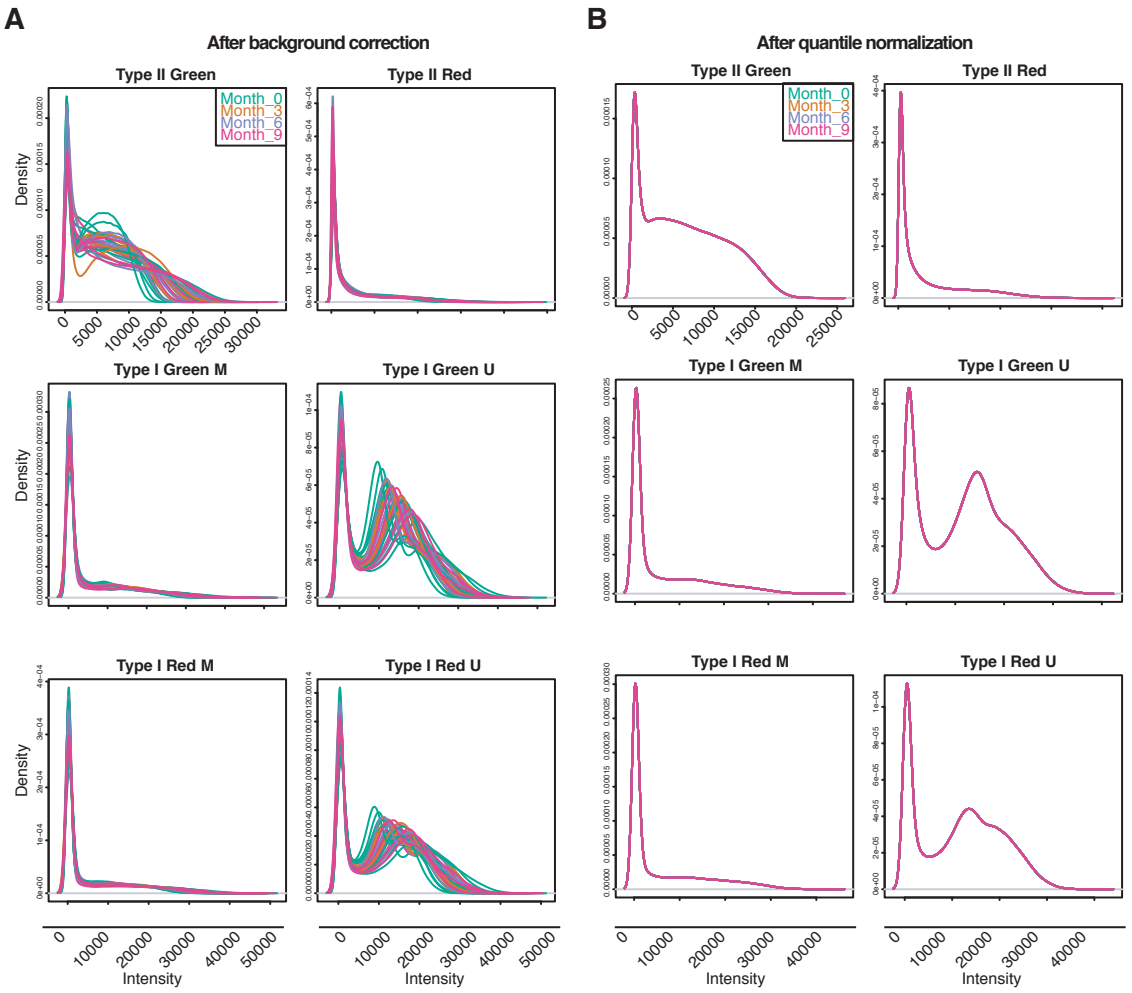

Figure. B.1. Intensity plots of neutrophils dataset before and after quantile normalization. Each graph represent a type of probe (Type I or Type II) and its channel (Red or Green). Type II probe is one probe able to emit signal in Red and Green channels. Type I probe is two probes that measure methylation (M) and unmethylation (U) and each emit signal in two channels Red or Green. A) Distribution of the signal intensity before quantile normalization and after background correction. B) Distribution of the signal intensity after quantile normalization.

Table. B.2. FastQC results: duplication (Dups), GC content and million sequences (M Seqs) from raw reads.

Sample Name	Dups	GC	M Seqs	Sample Name	Dups	GC	M Seqs
P002S01CD34_R1_001	79.8%	54%	27.4	P002S01NEU_R1_001	80.1%	47%	43.9
P002S04CD34_R1_001	93.0%	48%	32.3	P002S04NEU_R1_001	55.7%	51%	35.7
P003S01CD34_R1_001	70.8%	54%	31.4	P003S01NEU_R1_001	81.6%	49%	34
P003S04CD34_R1_001	95.2%	52%	34.4	P003S04NEU_R1_001	73.9%	53%	32.2
P004S01CD34_R1_001	92.3%	56%	42.3	P004S01NEU_R1_001	91.5%	49%	36.2
P004S04CD34_R1_001	91.4%	47%	22.4	P004S04NEU_R1_001	72.8%	52%	35.9
P006S01CD34_R1_001	51.7%	52%	33.2	P006S01NEU_R1_001	85.8%	49%	38.5
P006S04CD34_R1_001	92.5%	54%	41.6	P006S04NEU_R1_001	61.7%	52%	35.3
P007S01CD34_R1_001	51.7%	53%	31.5	P007S01NEU_R1_001	81.4%	49%	35.9
P007S04CD34_R1_001	69.4%	52%	38.7	P007S04NEU_R1_001	63.3%	52%	36.4
P008S01CD34_R1_001	51.6%	52%	36	P008S01NEU_R1_001	82.4%	48%	53.4
P008S04CD34_R1_001	48.6%	52%	31.6	P008S04NEU_R1_001	82.5%	50%	42.6
P009S01CD34_R1_001	38.2%	49%	33.7	P009S01NEU_R1_001	75.2%	50%	82.4
P009S04CD34_R1_001	41.7%	50%	34.6	P009S04NEU_R1_001	46.1%	49%	36.3
P010S01CD34_R1_001	59.1%	52%	38.8				
P010S04CD34_R1_001	64.4%	52%	35				
P011S01CD34_R1_001	56.3%	49%	34.6	P011S01NEU_R1_001	56.2%	55%	33.1
P011S04CD34_R1_001	66.5%	53%	40.9	P011S04NEU_R1_001	53.3%	56%	26

Table. B.3. Full RNAseqMetrics results after alignment to the reference genome using (Hg38)

ID	PF BASES	PF ALIGNED BASES	RIBOSOMAL BASES	CODING BASES	UTR BASES	INTRONIC BASES	INTERGENIC BASES	CORRECT STRAND READS	INCORRECT STRAND READS	PCT RIBOSOMAL BASES	PCT CODING BASES	PCT UTR BASES	PCT INTRONIC BASES	PCT INTERGENIC BASES	PCT MRNA BASES	PCT USABLE BASES	PCT CORRECT STRAND READS
P002S01CD34	3232296783	3231672858	917631	898214644	1038325407	797584115	496631076	11463197	177072	0.03%	27.79%	32.13%	24.68%	15.37%	59.92%	59.91%	98.48%
P002S04CD34	4076146344	4075628146	2879062	1088821408	757613879	1721422473	504891345	11078432	226972	0.07%	26.72%	18.59%	42.24%	12.39%	45.30%	45.30%	97.99%
P003S01CD34	4118566940	4117711366	1503965	1469798535	1235048861	1017540161	393819864	15651741	233496	0.04%	35.69%	29.99%	24.71%	9.56%	65.69%	65.67%	98.53%
P003S04CD34	1775441729	1775136297	663017	512279007	424882759	618496430	218815108	5527290	130532	0.04%	28.86%	23.94%	34.84%	12.33%	52.79%	52.78%	97.69%
P004S01CD34	4782025594	4781037075	9726645	853510911	1607013142	1587618711	723167810	14379785	280605	0.20%	17.85%	33.61%	33.21%	15.13%	51.46%	51.45%	98.09%
P004S04CD34	2895963267	2895531602	803441	549510343	384248369	1501098473	459871005	5555194	170182	0.03%	18.98%	13.27%	51.84%	15.88%	32.25%	32.24%	97.03%
P006S01CD34	4311507808	4310599496	595111	1125376080	1042721386	1712145583	429761340	12344491	241961	0.01%	26.11%	24.19%	39.72%	9.97%	50.30%	50.29%	98.08%
P006S04CD34	4809787589	4808875263	2214211	1040744102	1375667694	1443951480	946297813	14023252	365823	0.05%	21.64%	28.61%	30.03%	19.68%	50.25%	50.24%	97.46%
P007S01CD34	4089476435	4088695237	1416149	1155446046	1088084007	1423886636	419862418	12615889	279190	0.03%	28.26%	26.61%	34.83%	10.27%	54.87%	54.86%	97.83%
P007S04CD34	5246793967	5245936909	453929	1613922273	1200878482	1867674847	563007388	16241090	364916	0.01%	30.77%	22.89%	35.60%	10.73%	53.66%	53.65%	97.80%
P008S01CD34	4635413406	4634636803	655032	1512753749	1065636935	1573036718	482554383	15243786	296677	0.01%	32.64%	22.99%	33.94%	10.41%	55.63%	55.62%	98.09%
P008S04CD34	4160711085	4160015050	523474	1164540015	935647497	1544788818	514515268	12180196	277581	0.01%	27.99%	22.49%	37.13%	12.37%	50.49%	50.48%	97.77%
P009S01CD34	4595922920	4595173105	647451	1090804457	839374484	2167945108	496401618	11299858	286562	0.01%	23.74%	18.27%	47.18%	10.80%	42.00%	42.00%	97.53%
P009S04CD34	5217725988	5216718034	1232820	1361388429	1062779771	2264274085	527042942	14101296	390851	0.02%	26.10%	20.37%	43.40%	10.10%	46.47%	46.46%	97.30%
P010S01CD34	5511087193	5510016289	2518625	1636379329	1359838367	1754875484	756404515	17592602	405664	0.05%	29.70%	24.68%	31.85%	13.73%	54.38%	54.37%	97.75%
P010S04CD34	5010313600	5009378787	1504272	1524476583	1202519516	1579049522	701828926	15979465	321824	0.03%	30.43%	24.01%	31.52%	14.01%	54.44%	54.43%	98.03%
P011S01CD34	5021379939	5020306454	967383	1307992201	883970515	2026348895	801027472	13056902	408492	0.02%	26.05%	17.61%	40.36%	15.96%	43.66%	43.65%	96.97%
P011S04CD34	5610706089	5609633123	1532144	2011716914	1504041238	1540089700	552253139	21136809	356546	0.03%	35.86%	26.81%	27.45%	9.84%	62.67%	62.66%	98.34%
P002S01NEU	6413250318	6412737754	335056	2215664180	1531891990	2048605758	616240772	21629455	239254	0.01%	34.55%	23.89%	31.95%	9.61%	58.44%	58.43%	98.91%
P002S04NEU	4775878675	4775174765	1547867	1437904152	1214823157	1699683839	421215828	15248817	258659	0.03%	30.11%	25.44%	35.59%	8.82%	55.55%	55.54%	98.33%
P003S01NEU	3942263278	3941817801	203373	1320774082	1116856165	1226950585	277033596	14793184	138247	0.01%	33.51%	28.33%	31.13%	7.03%	61.84%	61.83%	99.07%
P003S04NEU	4158550416	4157892335	1703936	1546333359	1355936890	920948040	332970130	16849251	232283	0.04%	37.19%	32.61%	22.15%	8.01%	69.79%	69.79%	98.64%
P004S01NEU	4749308326	4748902462	79962	1683441039	1161107807	1591756995	312516659	16708576	144910	0.00%	35.45%	24.45%	33.52%	6.58%	59.90%	59.89%	99.14%
P004S04NEU	4612362469	4611656849	1958229	1606113171	1394457370	1142090760	467037341	17468421	267229	0.04%	34.83%	30.24%	24.77%	10.13%	65.06%	65.06%	98.49%
P006S01NEU	5047253795	5047713276	125749	1757692154	1325202256	1653883829	310349813	17593433	161386	0.00%	34.82%	26.26%	32.77%	6.15%	61.08%	61.08%	99.09%
P006S04NEU	4691850402	4691119043	1358070	1639305903	1337222256	1448163170	265069658	17187206	236519	0.03%	34.94%	28.51%	30.87%	5.65%	63.45%	63.44%	98.64%
P007S01NEU	4816837871	4816418215	115393	2098991922	1168137678	1268298859	280874379	19380116	160936	0.00%	43.58%	24.25%	26.33%	5.83%	67.83%	67.83%	99.18%
P007S04NEU	4926164855	4925434983	1339872	1745554581	1462182236	1402596052	313762320	18509322	227487	0.03%	35.44%	29.69%	28.48%	6.37%	65.13%	65.12%	98.79%
P008S01NEU	7459331169	7458576130	266788	1667118175	1514045499	2702054996	1575090674	18333718	417898	0.00%	22.35%	20.30%	36.23%	21.12%	42.65%	42.65%	97.77%
P008S04NEU	5540975744	5540270485	653756	1642593757	1353364714	1874944876	668713408	17343898	297301	0.01%	29.65%	24.43%	33.84%	12.07%	54.08%	54.07%	98.31%
P009S01NEU	11300423086	11299175798	1649815	2099928165	2240053748	3970293218	2987250897	24262304	991658	0.01%	18.58%	19.82%	35.14%	26.44%	38.41%	38.41%	96.07%
P009S04NEU	4860355265	4859656933	442063	948911024	824670633	1838329208	1247304007	10170093	460139	0.01%	19.53%	16.97%	37.83%	25.67%	36.50%	36.49%	95.67%
P011S01NEU	4084058813	4083416215	1040329	1126280298	1250663017	1201444366	503988219	12906655	297083	0.03%	27.58%	30.63%	29.42%	12.34%	58.21%	58.20%	97.75%
P011S04NEU	3108573712	3108275637	1049152	1082713316	1050884118	756146616	217482450	11503110	197480	0.03%	34.83%	33.81%	24.33%	7.00%	68.64%	68.63%	98.31%

Table. B.4. Number of sequences left after removal of duplicated reads using MarkDuplicates from Picard

ID	UNPAIRED READS EXAMINED	READ PAIRS EXAMINED	SECONDARY OR SUPPLEMENTARY RDS	UNPAIRED READ DUPLICATES	READ PAIR DUPLICATES	PERCENT DUPLICATION	ESTIMATED LIBRARY SIZE
P002S01CD34	15927212	3306421	11721069	14911884	2142241	0.851656	1253949
P002S01NEU	29732404	7015361	9771258	27567998	4262792	0.824749	3062456
P002S04CD34	19339327	4180050	4646980	19201449	4083038	0.988018	97012
P002S04NEU	22058708	5418244	11155395	15513227	1835709	0.583205	6064117
P003S01CD34	18107272	5182861	12633607	15740650	2732334	0.744752	2968436
P003S01NEU	19647526	4800401	15403375	17917122	3675418	0.863911	1142052
P003S04CD34	8957044	1660662	2432085	8808406	1619610	0.981207	41052
P003S04NEU	18277103	5281662	11005160	16368091	3130954	0.784662	2425595
P004S01CD34	21200534	5982407	25141109	20969749	5524621	0.965435	457786
P004S01NEU	23760008	4584832	10437790	23415423	4271256	0.970491	313576
P004S04CD34	14404373	2693236	2541113	14193715	2609286	0.980872	83950
P004S04NEU	21703551	5239012	12573155	18788787	2891122	0.763512	2762557
P006S01CD34	20052812	4950647	12391257	13002051	1562447	0.538388	6087459
P006S01NEU	24593273	5210626	11580177	23531515	4199366	0.911914	1017328
P006S04CD34	23251505	5072135	15852807	23014544	4809246	0.977161	262889
P006S04NEU	21814865	5331760	12119216	17149341	2077161	0.655934	4917469
P007S01CD34	19582186	4251138	11582321	12814967	1295898	0.548587	5469010
P007S01NEU	23464790	4910665	10270910	21937752	3428355	0.865059	1547011
P007S04CD34	24414042	5705283	11795978	20886882	2802264	0.739475	3688333
P007S04NEU	23224145	5470812	13439772	18229866	2141351	0.658922	5011934
P008S01CD34	22294285	4926106	12615227	14792350	1523750	0.554955	6218360
P008S01NEU	35715594	7739471	12714314	33203529	5357824	0.857888	2493554
P008S04CD34	20838879	3864501	10376927	13005086	1005516	0.525629	6072770
P008S04NEU	28317267	4983571	11510679	26550180	3320521	0.866964	1768730
P009S01CD34	22323649	4536400	9033636	11802592	883471	0.4322	10079545
P009S01NEU	60160822	8581316	22005832	52108218	4114200	0.780315	5772566
P009S04CD34	12280388	11870969	10042024	7137368	3668303	0.401806	15004401
P009S04NEU	26507028	3321699	8011299	15246100	680228	0.500945	6960467
P010S01CD34	13704365	12441405	13259654	10032050	6682102	0.606322	6893187
P010S04CD34	13151308	10847386	11004734	10341592	6377675	0.662828	5064459
P011S01CD34	12837158	10963877	8780977	9111183	5633879	0.586193	6566565
P011S01NEU	21827077	3209142	12340978	14533751	1054904	0.589249	3739710
P011S04CD34	14155838	12752558	14918212	11263004	8229364	0.698968	4881204
P011S04NEU	16292145	2652590	10910229	10589760	795838	0.564025	3483334

Appendix C

DNA methylation results

C.1 Differentially methylated CpG sites

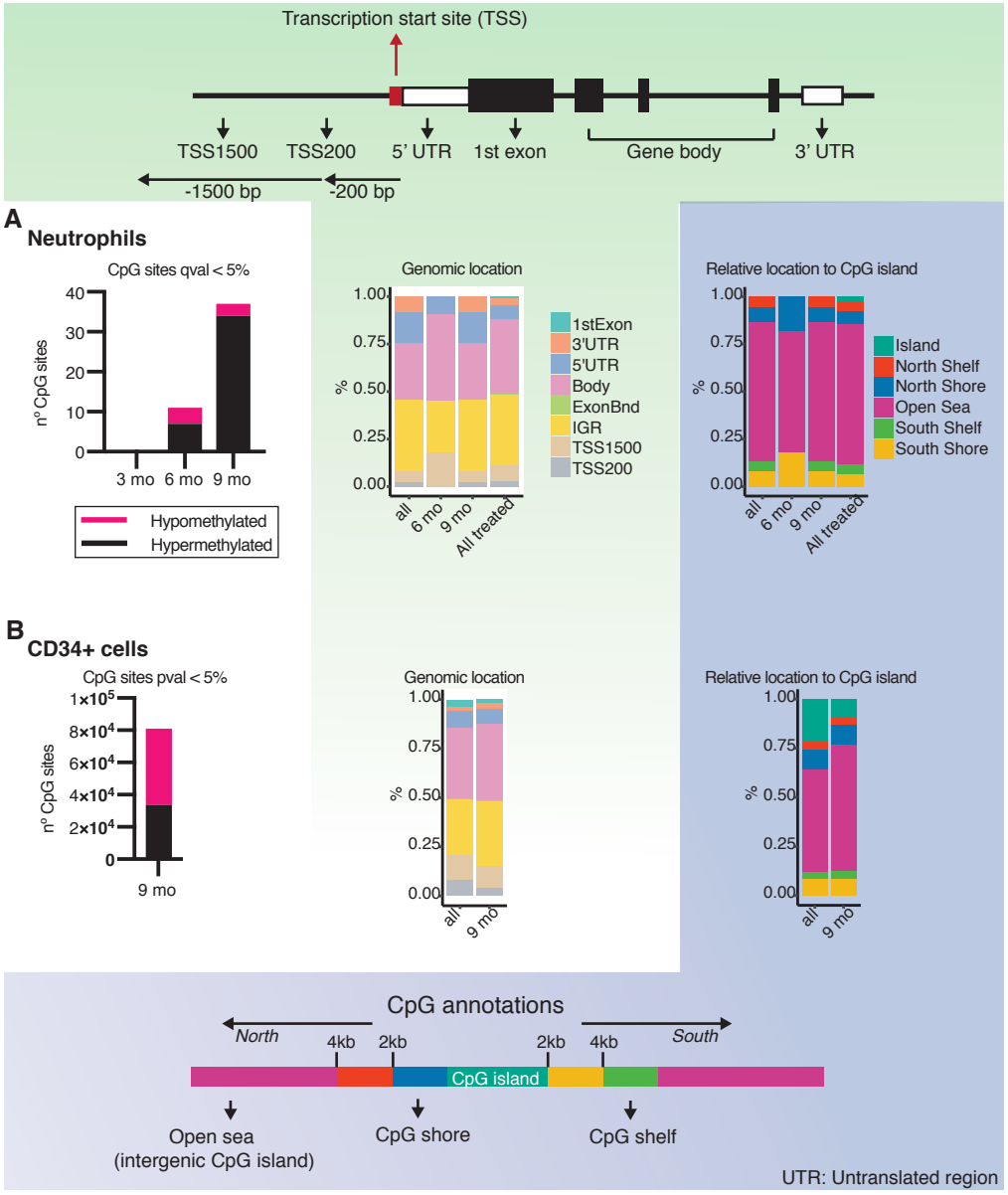

Figure. C.1. Total number of differentially methylated CpGs. Genomic locations and relative location to CpG island of the differentially methylated CpGs. A) Neutrophils; Differentially methylated CpGs per time-point in comparison to pre-treatment. No differentially methylated CpGs were identified at three months of HU treatment (3 mo). Twelve and 36 CpGs were differentially methylated at six months (6 mo) and nine months (9mo) of HU treatment, respectively. B) $CD34^{+}$ cells; differentially methylated CpG between untreated samples and nine months of HU treatment. "All" correspond to the total number of CpGs detected in each cell type and that passed the QC filters.

Appendix D

Data generation mouse

D.1 RNA sequencing: library preparation

D.2 RNA sequencing: FastQC results after sequencing in HiSeq4000 (raw reads)

D.3 RNAseqMetrics results

Table. D.1. Details of the library preparation steps for mouse samples. Details of the steps modified according to the RNA input and quality (RIN). Library concentration after library preparation, total nM obtained and total nM used for sequencing.

ID	Input RNA (ng)	Frag. time (min)	PCR2 n^{o}cycles	Library Qubit (ng/µL)	(Q) Qubit ng/µL Diluted	(T) Average Size [bp]	Q+T (nM)	nM (sequenced)
MM_4964_LSK	5	4	13	26.2	1.75	367	7.22	4
MM_5276_LSK	5	3	13	70.2	1.48	405	5.54	4
MM_5277_LSK	5	4	13	68	1.28	389	4.99	4
MM_5322_LSK	5	4	13	60.8	1.21	391	4.69	4
MM_5327_LSK	5	4	13	22.6	1.15	365	4.77	4
RB_427_LSK	5	4	13	33.8	1.95	368	8.03	4
RB_429_LSK	5	4	13	57.2	1.34	397	5.11	4
RB_430_LSK	5	4	13	13	1.44	412	5.3	4
RB_431_LSK	5	4	13	46.6	1.24	399	4.71	4
RB_438_LSK	5	4	13	27	1.21	360	5.09	4
RB_441_LSK	5	4	13	35.2	1.04	370	4.26	4
RB_444_LSK	5	4	13	62.2	1.26	410	4.66	4
RB_447_LSK	5	4	13	110	1.19	507	3.56	3.56
RB_458_LSK	5	4	13	57.2	1.34	438	4.64	4
RB_460_LSK	5	4	13	35.8	1.34	366	5.55	4
RB_461_LSK	5	3	13	14.1	1.92	565	5.15	4
RB_462_LSK	5	4	13	30.8	1.73	394	6.65	4
RB_470_LSK	5	4	13	16.2	1.1	382	4.36	4
RB_481_LSK	5	4	13	21.4	1.38	377	5.55	4
RB_489_LSK	5	4	13	30.2	1.33	410	4.92	4
RB_490_LSK	5	4	13	65	1.33	403	5	4
RB_491_LSK	5	4	13	36.2	1.48	362	6.19	4
SX_581_LSK	5	4	13	21.8	1.44	375	5.82	4
MM_4964_NEU	1	4	15	48.6	1.58	364	6.58	4
MM_5276_NEU	1	4	15	38.6	1.32	357	5.6	4
MM_5277_NEU	1	3	15	37.2	1.32	384	5.21	4
MM_5322_NEU	1	3	15	26.4	1.23	405	4.6	4
MM_5327_NEU	1	4	15	56.2	0.982	365	4.08	4
RB_427_NEU	1	3	15	55	1.43	379	5.72	4
RB_429_NEU	0.93	NA	15	3.52	1.38	382	5.47	4
RB_430_NEU	1	4	15	71.4	1.3	360	5.47	4
RB_431_NEU	0.97	NA	15	25.8	1.47	538	4.14	4
RB_438_NEU	1	NA	15	9.2	1.48	471	4.76	4
RB_441_NEU	1	4	15	58.8	1.23	368	5.06	4
RB_444_NEU	0.95	3	15	45.6	1.19	386	4.67	4
RB_447_NEU	1	4	15	50.8	1.35	387	5.29	4
RB_458_NEU	1	3	15	24.2	1.56	453	5.22	4
RB_460_NEU	1	3	15	51.2	1.69	381	6.72	4
RB_461_NEU	1	3	15	40.4	1.43	411	5.27	4
RB_462_NEU	1	3	15	65	1.7	368	7	4
RB_470_NEU	0.74	NA	15	2.06	1.36	358	5.76	4
RB_481_NEU	1	4	15	59.2	1.62	366	6.71	4
RB_489_NEU	1	3	15	49.4	1.57	416	5.72	4
RB_490_NEU	1	4	15	52.2	1.64	370	6.72	4
RB_491_NEU	1	NA	15	22.4	1.47	490	4.55	4
SX_581_NEU	1	4	15	40.2	1.43	367	5.9	4

Q+T= library quantification using method described in 2.4.4
NA: no fragmentation step performed

Table. D.2. FastQC results: duplication (Dups), GC content and million sequences (M Seqs) from raw reads.

Sample Name	Dups	GC	M Seqs	Sample Name	Dups	GC	M Seqs
MM-4964-LSK_R2_001	48.50%	56%	39.6	MM-4964-NEU_R2_001	51.70%	58%	15.4
MM-5276-LSK_R2_001	44.60%	55%	26.1	MM-5276-NEU_R2_001	51.40%	58%	10.5
MM-5277-LSK_R2_001	52.80%	59%	33.7	MM-5277-NEU_R2_001	52.40%	55%	15.6
MM-5322-LSK_R2_001	57.00%	59%	32.9	MM-5322-NEU_R2_001	61.90%	56%	22.3
MM-5327-LSK_R2_001	42.50%	55%	22.8	MM-5327-NEU_R2_001	60.40%	60%	20.8
RB-427-LSK_R2_001	48.30%	56%	30.4	RB-427-NEU_R2_001	44.50%	56%	9.5
RB-429-LSK_R2_001	52.60%	59%	38	RB-429-NEU_R2_001	76.20%	54%	18.5
RB-430-LSK_R2_001	69.50%	58%	49.6	RB-430-NEU_R2_001	65.80%	62%	27.8
RB-431-LSK_R2_001	68.50%	60%	51.5	RB-431-NEU_R2_001	64.60%	54%	18
RB-438-LSK_R2_001	47.20%	55%	33.8	RB-438-NEU_R2_001	72.90%	55%	22.2
RB-441-LSK_R2_001	52.60%	56%	49.5	RB-441-NEU_R2_001	63.10%	61%	20.8
RB-444-LSK_R2_001	46.50%	58%	21.1	RB-444-NEU_R2_001	52.90%	57%	13.7
RB-447-LSK_R2_001	28.20%	52%	21.2	RB-447-NEU_R2_001	55.70%	58%	19.1
RB-458-LSK_R2_001	46.20%	56%	27.3	RB-458-NEU_R2_001	62.00%	56%	16.6
RB-460-LSK_R2_001	40.70%	55%	20.1	RB-460-NEU_R2_001	56.10%	58%	13.6
RB-461-LSK_R2_001	48.50%	59%	23.4	RB-461-NEU_R2_001	65.70%	58%	17.9
RB-462-LSK_R2_001	50.30%	57%	25	RB-462-NEU_R2_001	60.30%	60%	18.2
RB-470-LSK_R2_001	55.40%	57%	30.9	RB-470-NEU_R2_001	83.30%	53%	24.7
RB-481-LSK_R2_001	57.80%	58%	33.8	RB-481-NEU_R2_001	55.10%	61%	13.2
RB-489-LSK_R2_001	64.00%	58%	50.5	RB-489-NEU_R2_001	59.20%	53%	26.4
RB-490-LSK_R2_001	22.60%	49%	21.9	RB-490-NEU_R2_001	46.40%	58%	10.4
RB-491-LSK_R2_001	40.20%	56%	17.8	RB-491-NEU_R2_001	69.90%	49%	27.6
SX-581-LSK_R2_001	64.30%	58%	46.9	SX-581-NEU_R2_001	57.70%	56%	24.6

Table. D.3. Full RNAseqMetrics results after alignment to the reference genome using (mm10)

ID	PF BASES	PF ALIGNED BASES	RIBOSOMAL BASES	CODING BASES	UTR BASES	INTRONIC BASES	INTERGENIC BASES	CORRECT STRAND READS	INCORRECT STRAND READS	PCT RIBOSOMAL BASES	PCT CODING BASES	PCT UTR BASES	PCT INTRONIC BASES	PCT INTERGENIC BASES	PCT MRNA BASES	PCT USABLE BASES	PCT CORRECT STRAND READS
MM-4964-LSK	4539660821	4539175026	471982	1956542701	681863414	1354168902	546129947	17641946	218352	0.000104	0.431035	0.150217	0.298329	0.120315	0.581252	0.58119	0.987774
MM-5276-LSK	3094844595	3094587251	135121	1276810814	423269783	957989226	436382533	11212638	130406	0.000044	0.412595	0.136777	0.309569	0.141015	0.549372	0.549327	0.988503
MM-5277-LSK	3579069889	3578674156	280604	1502853420	497564153	1082065949	495910214	13368917	150022	0.000078	0.419947	0.139036	0.302365	0.138574	0.558983	0.558921	0.988903
MM-5322-LSK	3458009034	3457670121	386307	1448852133	466539815	1032410052	509482255	12748504	130970	0.000112	0.419026	0.134929	0.298585	0.147348	0.553955	0.5539	0.989831
MM-5327-LSK	2626767393	2626509952	207989	1233078858	385379458	703090634	304753329	10810052	116551	0.000079	0.469474	0.146727	0.26769	0.11603	0.616201	0.616141	0.989333
RB-427-LSK	3421574626	3421204550	364019	1388850598	481654539	1076990803	473345679	12418902	158672	0.000106	0.405954	0.140785	0.314799	0.138356	0.546739	0.54668	0.987385
RB-429-LSK	3974900397	3974435291	330560	1565187344	540869973	1327624387	540423880	14178623	161782	0.000083	0.393814	0.136087	0.334041	0.135975	0.529901	0.529839	0.988718
RB-430-LSK	5630081149	5629687621	301050	2580070802	777241439	1563209852	708865114	21994060	352481	0.000053	0.458297	0.138061	0.277673	0.125916	0.596359	0.596317	0.984227
RB-431-LSK	5549702404	5549289716	434395	2560510409	784498580	1435919082	767928078	21941504	313874	0.000078	0.461412	0.141369	0.258757	0.138383	0.602781	0.602737	0.985897
RB-438-LSK	3919420809	3919048263	290213	1681915532	579907439	1182507439	474428026	15043237	183358	0.000074	0.429164	0.147971	0.301733	0.121057	0.577136	0.577081	0.987958
RB-441-LSK	5498800909	5498124904	986259	2332524208	819300057	1578067788	767250483	21054200	254849	0.000179	0.42424	0.149014	0.287019	0.139548	0.573254	0.573184	0.98804
RB-444-LSK	2245467819	2245214859	181392	843003779	270857734	742536367	388636300	7447597	120433	0.000081	0.375467	0.120638	0.33072	0.173095	0.496105	0.496049	0.984087
RB-447-LSK	2620852810	2620646654	84251	541368834	206975779	980278234	891939838	4901055	255625	0.000032	0.206578	0.078979	0.37406	0.340351	0.285557	0.285535	0.950428
RB-458-LSK	3195312039	3195004922	275137	1122310927	373352626	1109288450	589778314	9918030	178555	0.000086	0.35127	0.116855	0.347195	0.184594	0.468126	0.468081	0.982315
RB-460-LSK	2297253973	2297036828	208302	953914706	321183519	719154684	302575921	8550963	100121	0.000091	0.415281	0.139825	0.313079	0.131724	0.555106	0.555053	0.988427
RB-461-LSK	2553644223	2553382368	242234	1040196021	342198179	773318503	397427977	9131897	111327	0.000095	0.40738	0.134018	0.30286	0.155648	0.541397	0.541342	0.987956
RB-462-LSK	2779878240	2779639682	246169	1142066437	384994748	840658300	411674222	10146489	124180	0.000089	0.410869	0.138505	0.302434	0.148103	0.549374	0.549327	0.987909
RB-470-LSK	3523478445	3523201007	223283	1483904218	491369494	1111752872	435951604	13119813	143747	0.000063	0.421181	0.139467	0.315552	0.123737	0.560647	0.560603	0.989162
RB-481-LSK	3513614121	3513160078	602438	1353989457	472126118	1110743676	575700660	12159683	168124	0.000171	0.385405	0.134388	0.316167	0.16387	0.519793	0.519726	0.986362
RB-489-LSK	5393749614	5393060070	1327852	2047047273	714245993	1655342319	975102079	18284714	214581	0.000246	0.379571	0.132438	0.306939	0.180807	0.512009	0.511943	0.988401
RB-490-LSK	2834670878	2834488662	45520	540446243	186503881	996668611	1110824455	4680186	324592	0.000016	0.190668	0.065798	0.351622	0.391896	0.256466	0.25645	0.935144
RB-491-LSK	2053102623	2052909590	150366	899488902	296995869	620975491	235299254	7955351	97147	0.000073	0.438153	0.144671	0.302486	0.114617	0.582824	0.582769	0.987936
SX-581-LSK	4510264024	4509772849	1137409	1929484896	619684083	1353299553	606168511	17626078	250575	0.000252	0.427845	0.137409	0.300082	0.134412	0.565254	0.565193	0.985983
MM-4964-NEU	1790524054	1790338091	49386	582972975	253770302	710235335	237310366	5595619	145767	0.000028	0.325622	0.141744	0.400056	0.132551	0.467366	0.467318	0.974611
MM-5276-NEU	1062981494	1062797938	197173	317965938	150239188	439622663	154773545	3251529	80145	0.000186	0.299178	0.141362	0.413647	0.145628	0.44054	0.440464	0.975945
MM-5277-NEU	1892088354	1891927767	68631	656942278	287949292	717337432	229630463	6252503	146173	0.000036	0.347234	0.152199	0.379157	0.121374	0.499433	0.499391	0.977156
MM-5322-NEU	2418910467	2418574408	305769	746214676	351498924	897631125	422925061	7377528	212530	0.000126	0.308535	0.145333	0.371141	0.174865	0.453868	0.453805	0.971999
MM-5327-NEU	2357029985	2356774112	58746	826375358	371587271	847396197	311356826	7942090	163035	0.000025	0.350638	0.157668	0.359558	0.132111	0.508306	0.508251	0.979885
RB-427-NEU	1140348558	1140246725	33297	365369907	155736489	459538524	159568845	3458171	90828	0.000029	0.320431	0.136581	0.403017	0.139942	0.457012	0.456971	0.974407
RB-429-NEU	1521782841	1521535373	61060	494714833	292960969	362692692	371106172	5129072	169211	0.00004	0.325142	0.192543	0.238373	0.243902	0.517685	0.517601	0.968063
RB-430-NEU	2839446904	2839081692	208119	851969539	387894786	1231013312	367996320	8468280	159052	0.000073	0.300086	0.136627	0.433596	0.129618	0.436713	0.436657	0.981564
RB-431-NEU	2101954823	2101822266	50350	700513122	392726153	579277570	429255275	7001850	230359	0.000024	0.333288	0.18685	0.275607	0.20423	0.520139	0.520106	0.968148
RB-438-NEU	2351599332	2351416841	37369	794200683	410161674	612101474	534915733	7748760	235035	0.000016	0.337754	0.174432	0.260312	0.227487	0.512186	0.512146	0.970561
RB-441-NEU	2156657827	2156361142	125021	685744508	297304345	886667168	286520641	6627389	160552	0.000058	0.31801	0.137873	0.411187	0.132872	0.455883	0.455821	0.976347
RB-444-NEU	1515419681	1515240702	110623	504682501	219588626	545698220	245161119	4846369	129898	0.000073	0.333071	0.14492	0.36014	0.161797	0.477991	0.477934	0.973896
RB-447-NEU	2179438081	2179194610	132749	712481706	308508181	897571805	260500766	6818524	164671	0.000061	0.326947	0.14157	0.411882	0.11954	0.468517	0.468465	0.976419
RB-458-NEU	1967136411	1966952512	90783	653488467	299881726	707731915	305760030	6321857	152380	0.000046	0.332234	0.15246	0.359811	0.155449	0.484694	0.484649	0.976464
RB-460-NEU	1589107722	1588966992	30303	560693939	238803195	574931355	214508369	5313049	123916	0.000019	0.352867	0.150288	0.361827	0.134999	0.503155	0.503111	0.977209
RB-461-NEU	2066513995	2066312805	121746	923312355	436330112	816039796	390509431	8915611	120241	0.000059	0.446841	0.211164	0.152949	0.188989	0.658004	0.657939	0.986693
RB-462-NEU	2014214666	2014019498	33520	746740027	315292562	724744799	227208647	7092527	122045	0.00017	0.370771	0.156549	0.35985	0.112814	0.52732	0.527269	0.983084
RB-470-NEU	804685266	804577447	11278	173337172	89968799	293559865	247700365	1728251	111160	0.000014	0.215439	0.111821	0.364862	0.307864	0.32726	0.327216	0.939568
RB-481-NEU	1401508601	1401347831	62923	472892992	199367601	529748724	199275773	4537396	94307	0.000045	0.337456	0.142268	0.378028	0.142203	0.479724	0.479669	0.979639
RB-489-NEU	3217744262	3217503903	72802	928730771	412257757	1129103543	747339130	8753502	318144	0.000023	0.288649	0.12813	0.350925	0.232273	0.416779	0.416748	0.96493
RB-490-NEU	1113295754	1113142847	167936	347008450	156651915	454018204	155296906	3448736	88375	0.000151	0.311738	0.140729	0.407871	0.139512	0.452467	0.452405	0.975015
RB-491-NEU	3609900897	3609749994	11682	441269343	233958011	1290858804	1643652161	4130689	620661	0.000003	0.122244	0.064813	0.357603	0.455337	0.187057	0.187049	0.869372
SX-581-NEU	2875869435	2875555252	342828	907852439	409859765	1134828120	422673856	8827287	231077	0.000119	0.315714	0.142532	0.394647	0.146989	0.458246	0.458196	0.97449