EPIGENTIC LANDSCAPE OF THE PLASMINOGEN ACTIVATOR UROKINASE LOCUS IN QUEBEC PLATELET DISORDER

By ASIM SOOMRO, B.Sc.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfilment of the Requirements for the Degree Master of Science

McMaster University © Copyright by Asim Soomro, 2016

McMaster University MASTER OF SCIENCE (2016) Hamilton, Ontario (Medical Sciences)

TITLE: Epigenetic landscape of the plasminogen activator urokinase locus in Quebec platelet disorder

AUTHOR: Asim Soomro, BSc (Hon.) (McMaster University)

SUPERVISOR: Dr. Catherine P.M. Hayward

NUMBER OF PAGES: 116

ABSTRACT

Quebec platelet disorder (QPD) is a bleeding disorder characterized by a gain of function defect in fibrinolysis. The hallmark feature of QPD is the marked overexpression of urokinase plasminogen activator (uPA) in megakaryocytes (MK) and platelets. The genetic cause of QPD is a tandem duplication of a ~78 kb region that encompasses the uPA gene, *PLAU*. As the mechanism of *PLAU* overexpression is unknown, gene regulatory mechanisms specifically epigenetics were evaluated at the *PLAU* locus in QPD MK and granulocytes, a QPD unaffected lineage. The aims of the thesis were to assess if QPD is associated with 1) genome wide methylation changes of promoter CpG islands, particularly at *PLAU* and 2) genome wide changes of active histone modifications H3K27Ac, H3K36me3 and H3K4me2, particularly at the region of *PLAU* duplication. Methylation and active histone enrichment analysis revealed that in QPD and control subjects, *PLAU* promoter CpG island was characterized by unaltered hypo-methylation and changes in active histone peak enrichments that were within the realm of having one extra copy of *PLAU* in both MK and granulocytes. The findings imply that the *PLAU* CNV mutation does not induce altered promoter methylation status and/or significantly alter active histone markers as the reason for the marked *PLAU* overexpression in QPD MK. Instead, the rearrangement of an active enhancer element, particularly an H3K27Ac enhancer expressed in MK but not granulocytes, that is upstream of the second copy of *PLAU* might underlie the marked *PLAU* expression by differentiated QPD MK. The thesis provides novel insights into the epigenetic regulation of *PLAU* that will be crucial to identifying the mechanism underlying the aberrant *PLAU* expression in QPD.

ACKNOWLEDGEMENTS

I would firstly like to express my sincere gratitude to my supervisor Dr. Catherine P.M. Hayward for her support and mentorship throughout my degree. Her supervision has allowed me to improve my scientific research and critical skills. Her guidance and support is valuable beyond measure. I would also like to express my gratitude towards my committee members Dr. Michael Wilson, Dr. Bekim Sadikovic and Dr. John A. Hassell. My committee members have been a tremendous support by providing helpful advice and support throughout the course of my degree.

I had the esteemed pleasure of working with great individuals from members of both the Hayward and Wilson Labs. I would like to sincerely thank Dr. Tasneem, D'Andra, Matthew and Alex for always being there for guidance when I needed them. I would like to thank Minggao, Lina, Liis, Huayun, Azad and Xiaoli from the Wilson Lab for their help and support. I would also like to extend heartfelt gratitude towards the blood donors for their invaluable contribution towards this study.

Finally, I would like to thank my family, my parents Nabeela and Najam and brothers Aamir and Hasaan for their patience love and support. This achievement would not have been possible without them. To my better half Maria Claudia, for her unwavering support in the best and toughest of times during the course of my degree, I cannot be more grateful.

This work has been by far my greatest challenge and accomplishment and I am truly privileged to have worked on the QPD project and will always cherish the time spent at McMaster.

DATA ATTRIBUTION

I performed all of the $CD34^+$ HSC isolations, $CD41^+$ MK culturing, peripheral blood isolations of $CD66b^+$ granulocytes and verification analysis for ChIP-seq. I performed all the ChIP-seq experiments in Dr. Michael Wilson's Lab (The Hospital for Sick Kids Toronto, ON, Canada). DNA was extracted from MK and granulocyte cell pellets, bisulfite converted and assayed on the Infinium HumanMethylation 450K BeadChIP Array by the Genetic and Epidemiology Laboratory (Thrombosis & Atherosclerosis Research Institute, TaARI, Hamilton, ON, Canada)

The trimming, quality control and alignment of ChIP-seq data to the hg19 reference genome was performed by Minggao Liang in the Wilson Lab (The Hospital for Sick Kids Toronto, ON, Canada). I performed peak calling, read counting and differential analysis of ChIP-Seq histone modification profiles. Minggao Liang and I developed QPD chromosome 10 model together, where Minggao aligned QPD ChIP-Seq data sets to the model. Figure 11 was adapted from Minggao's original figure.

Methylation data analysis was conducted in collaboration with Dr. Sadikovic Lab. Laila C. Schenkel and I together performed and interpreted differential methylation analysis using standard pipeline in Dr. Sadikovic Lab. Laila C. Schenkel created the figures from the methylation data used in my thesis.

TABLE OF CONTENTS

1. INTRODUCTION 1
1.1 Hemostasis- coagulation and fibrinolysis 1
1.2 Regulation of fibrinolysis 2
1.3 Urokinase plasminogen activator 5
1.3.1 Structure and regulation 5
1.3.2 Fibrinolysis and uPA 6
1.3.3 Transcriptional regulation of *PLAU* 7
1.3.4 Murine models of uPA 10
1.4 Hematopoiesis 11
1.4.1 Megakaryopoiesis 13
1.4.1.1 Megakaryocyte development 13
1.4.1.2 Platelet production 14
1.4.1.3 Cytokine regulation of megakaryopoiesis 14
1.4.1.4 Transcriptional regulation of megakaryopoiesis 15
1.4.2 Granulopoiesis 16
1.4.2.1 Granulocyte development and transcriptional regulation 16
1.5 Quebec platelet disorder 17
1.5.1 Clinical features and symptoms 17
1.5.2 QPD and uPA fibrinolysis 19
1.5.3 QPD and uPA overexpression during megakaryopoiesis 20
1.5.4 QPD and *PLAU* linkage analysis 21
1.5.5 QPD *PLAU* mutation 21
1.5.6 QPD MK specificity and unaffected QPD cell lineages 23
1.6 Chromatin structure and regulation of gene expression 24
1.7 Epigenetics 26
1.7.1 DNA methylation 26
1.7.2 Histone modifications 28
1.8 Epigenetics and QPD *PLAU* expression 30
2. THESIS HYPOTHESIS AND SPECIFIC OBJECTIVES 32

2.1 Hypothesis 32
2.2 Objective#1 32
2.3 Objective#2 32
2.3.1 Secondary objective 32
3. MATERIALS AND METHODS 33
3.1 Blood collection 33
3.2 Isolation of CD66b$^+$ granulocytes from peripheral blood 34
3.3 Isolation of CD34$^+$ HSCs 34
3.4 CD34$^+$ HSC culturing to CD41$^+$ MK 36
3.5 Quality control analysis of CD41$^+$ MK and CD66b$^+$ granulocytes 37
3.5.1 Flow cytometry 37
3.5.2 uPA and VWF quantification in Day 14 MK cultures 38
3.6 Droplet Digital PCR for rs4065 SNP T/C allele expression 40
3.7 ChIP experimental procedure 40
3.8 Library preparation of ChIP-DNA 43
3.9 ChIP-qPCR validation and NGS of ChIP libraries 44
3.10 ChIP-seq pipeline: 46
3.11 ChIP-seq quality control 47
3.12 Peak calling and differential analysis 47
3.13 QPD breakpoint analysis 49
3.14 DNA methylation analysis using Infinium HumanMethylation450K BeadChip Array 49
4. RESULTS 51
4.1 Quality control analysis of Day 14 CD41$^+$ MK 51
4.2 Infinium HumanMethylation450K BeadChip Array analysis of the *PLAU* region 53
4.3 CD66b$^+$ granulocytes ChIP-seq for active histone modifications 58
4.4 CD41$^+$ MK ChIP-seq for active histone modifications 61
4.5 QPD ChIP-seq active histone modification enrichment at QPD breakpoint 67
5. DISCUSSION 69
5.1 Influence of CpG promoter island methylation on *PLAU* expression 69
5.2 ChIP-seq active histone modification enrichment in QPD 71
5.3 Duplication of a marked enhancer in QPD MK 73

5.4 Future directions 79
5.4.1 Expression of MKE1 during megakaryopoiesis 79
5.4.2 ChIA-PET analysis of MKE1 81
5.4.3 Potential TFs associated with *PLAU* dysregulation in QPD 82
6. CONCLUSION 84
7. REFERENCES 85
8. APPENDIX 101

LIST OF FIGURES

Figure 1. Regulation of fibrinolysis.

Figure 2. Regulatory regions upstream of *PLAU*.

Figure 3. Myeloid lineages derived from hematopoietic stem cells.

Figure 4. SNP rs4065 T/C allelic ratio analysis in myeloid lineage cells from QPD and control subjects.

Figure 5. Diagrammatic representation of the methylation status of *PLAU* CpG island and neighboring methylation sites for QPD and control MK and granulocytes, evaluated using the Infinium HumanMethylation450K BeadChip Array.

Figure 6. Diagrammatic representation of the methylation status for the *CAMK2G-PLAU-VCL* region for QPD and control MK and granulocytes, evaluated using the Infinium HumanMethylation450K BeadChip Array.

Figure 7. ChIP-seq profiles for active histone modifications in QPD and control granulocytes.

Figure 8. Normalized read counts for the differential regions identified through ChIP-seq analysis of QPD and control granulocytes.

Figure 9. ChIP-seq profiles for active histone modifications in QPD and control MK.

Figure 10. Normalized read counts for the differential regions identified through ChIP-seq analysis of QPD and control MK.

Figure 11. QPD ChIP-seq enrichment at the QPD breakpoint.

Figure 12. Model summarizing the location of the MK Enriched H3K27Ac enhancer relative to *PLAU* on the QPD disease chromosome.

LIST OF TABLES

Table 1. Differentiation data for QPD and control MK cultures, harvested on Day 14.

Table 2. Regions identified to show differences in methylation from genome wide methylation analysis of QPD and control granulocytes.

Table 3. Regions identified to show differences in methylation from genome wide methylation analysis of QPD and control MK.

Table 4. Summary of the fold enrichment data for comparisons of QPD and control samples for differential regions with active histone modifications.

Table 5. Fold enrichment analysis of differential regions on chromosome 10 in MK and granulocytes samples from QPD and control subjects.

LIST OF SUPPLEMENTARY TABLES

Supplementary Table 1. Summary of quality control data for ChIP-seq samples used for granulocyte library sequencing.

Supplementary Table 2. Summary of quality control data for ChIP-seq samples used for MK library sequencing.

LIST OF SYMBOLS AND ABBREVIATIONS

α	Alpha
β	Beta
γ	Gamma
δ	Delta
κ	Kappa
ACD	Acid citrate dextrose
ACTB	Actin, Beta
ADP	Adenosine diphosphate
AML	Acute myeloid leukemia
AP1	Activator protein 1
Arg	Arginine
ATAC	Assay for transposase-accessible chromatin
BFU	Burst forming unit
BSA	Bovine serum albumin
C/EBPα	CCAAT-enhancer-binding proteins alpha
C10orf55	C10 open reading frame 55
CAMK2G	Calcium/calmodulin-dependent protein kinase II gamma
CBFb	Core-binding factor subunit beta
CD	Cluster of differentiation
CFU	Colony-forming unit
ChIA-PET	Chromatin interaction analysis by paired-end tag sequencing
ChIP-seq	Chromatin immunoprecipitation with next generation sequencing
CLP	Common lymphoid progenitor
CMP	Common myeloid progenitor
c-MPL	Cellular myeloproliferative leukemia proto-oncogene
CNV	Copy number variation
CREB	cAMP response element-binding protein
CTCF	CCCTC-binding factor
CTF	CCAAT box-binding transcription factor
Da	Dalton
DNA	Deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetic acid
EGF	Epidermal growth factor
EGTA	Ethylene glycol tetraacetic acid
ELISA	Enzyme linked immunosorbent assay
ENCODE	Encyclopedia of DNA elements

Ets	E-twenty-six
FDP	Fibrin degradation products
FDR	False discovery rate
FITC	Fluorescein isothiocynate
FLI1	Friend leukemia integration-1
Forward	F
FSC	Forward scatter
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
GATA	Globin transcription factor
GB	Gene body
GMP	Granulocyte monocyte progenitor
HEPES	N-2-hydroxyethylpiperazine-N'-2-ethanesulfonic acid
HRP	Horseradish peroxidase
HSC	Hematopoietic stem cell
IGR	Intergenic region
LMW uPA	Low molecular weight uPA
MEP	MK/erythrocyte bipotent progenitor
MK	Megakaryocytes
MKE1	MK specific enhancer 1
MMRN1	Multimerin 1
mRNA	Messenger ribonucleic acid
NF-$_k$B	Nuclear factor kappa-light-chain-enhancer of activated B cells
NRF	Non redundant fraction of mapped reads
OPD	o-Phenylenediamine
PAI-1	Plasminogen activator inhibitor 1
PBMC	Peripheral blood mononuclear cell
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PGE_1	Prostaglandin E1
PLAU	Plasminogen activator, urokinase
PR	Promoter
PRC	Polycomb repressive complex
qPCR	Quantitative polymerase chain reaction
QPD	Quebec platelet disorder
RBC	Red blood cells
Reverse	R
RIPA	Radio immunoprecipitation assay buffer
RNaseA	Ribonuclease A
RSC	Relative strand coefficient

RT	Real time
RUNX1	Runt-related transcription factor 1
sctPA	Single chain tissue plasminogen activator
SCF	Stem cell factor
SEM	Standard error of mean
SFEM	Serum-free expansion medium
SN	Supernatants
SNP	Single nucleotide polymorphisms
SSC	Side scatter
TBP	TATA-binding protein
TBS	Tris buffered saline
tctPA	Two chain tissue plasminogen activator
tcuPA	Two chain urokinase plasminogen activator
TF	Transcription factor
TMB	Tetramethylbenzidine
tPA	Tissue plasminogen activator
TPO	Thrombopoietin
TSS	Transcription start site
UCSC	University of California, Santa Cruz
UEF	Upstream enhancer factors
uPA	Urokinase plasminogen activator
uPAR	Urokinase plasminogen activator receptor
UTR	Untranslated region
Val	Valine
VCL	Vinculin
VWF	von Willebrand factor
WCE	Whole cell extract

1. INTRODUCTION

1.1 Hemostasis- coagulation and fibrinolysis

Hemostasis can be defined as an intricate process that ensures the integrity of the vasculature and maintains a balance between bleeding and clotting [1, 2]. Hemostasis regulates physiological processes involved in maintaining the fluidity of blood and preventing excessive bleeding after vascular injury [1]. In health, the enzymatic and cellular components of blood are balanced to maintain fluidity but are also primed for a fast and effective response to vascular injury [1]. Hemostatic balance is achieved by regulating two crucial mechanisms, namely coagulation and fibrinolysis. The coagulation cascade is responsible for the formation of a hemostatic plug that limits bleeding, while fibrinolysis is the process of dissolving the fibrin plug that forms at sites of injury [1]. Hemostasis is thus a balance between pro-coagulant, anti-coagulant, pro-fibrinolytic and anti-fibrinolytic events [1]. A disturbance in hemostatic balance can result in either excessive bleeding or thrombosis [1].

In the event of vascular damage, tissue factor is exposed at the endothelial surface, triggering a series of enzymatic reactions termed the coagulation cascade that ultimately generates activated Factor II or thrombin [3]. One of the major functions of thrombin is the conversion of fibrinogen to the insoluble fibrin network, in addition to activating resting platelets [3]. The formation of a fibrin network, in combination with activated platelets that adhere and aggregate at a site of injury, form the primary hemostasis plug [1, 3, 4]. As hemostasis ensures continuous blood flow, it is imperative to dissolve the

fibrin clot. Fibrinolysis, like the coagulation cascade, consists of a series of reactions that convert zymogen (inactive form) enzymes to their active form that digest the fibrin clot in a regulated manner [1]. Fibrinolysis is of particular interest to this study and the regulation of this process is summarized below.

1.2 Regulation of fibrinolysis

Tissue plasminogen activator (tPA) and urokinase plasminogen activator (uPA) are the key activators of fibrinolysis that both convert single chain plasminogen into the active protease plasmin [1, 5]. The plasminogen activators, tPA and uPA are serine proteases, that cleave plasminogen to yield active two chain plasmin [1]. The next step of the fibrinolysis occurs when active two-chain plasmin, also a serine protease, degrades the crosslinked fibrin clot to products that are termed fibrin degradation products (FDPs) [1]. Plasmin is a broad specificity enzyme that acts not only on fibrin but many other substrates, and has additional roles in processes such as wound healing, tissue remodeling, angiogenesis and inflammation [6]. An overview of fibrinolysis regulation is presented in Figure 1.

The primary step of fibrinolysis, the conversion of plasminogen to plasmin, is positively regulated by fibrin, where it is not only a co-activator of plasminogen, but is also the target substrate of plasmin activity [7].

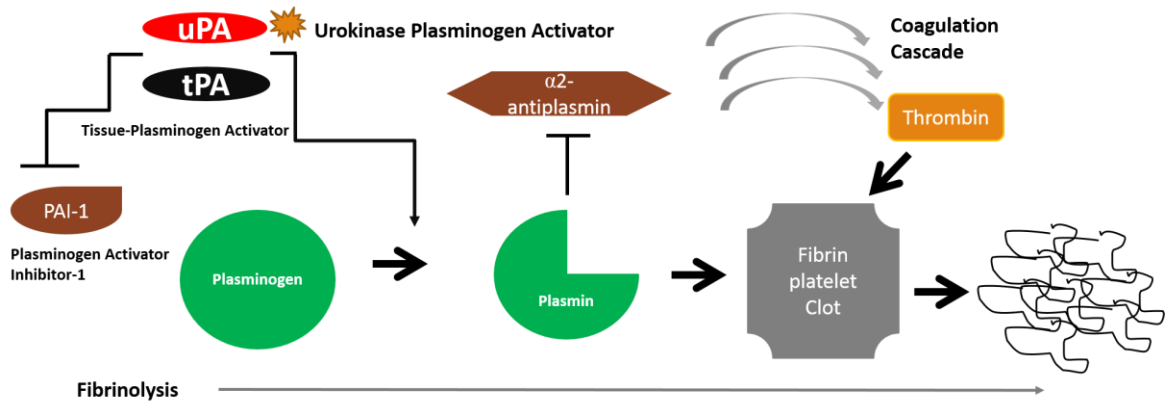

Figure 1. Regulation of fibrinolysis. A schematic representation of the major components involved in fibrinolysis are represented. Thrombin generation yields a fibrin clot (converting fibrinogen to fibrin), and activates platelets to forms a primary hemostatic plug that contains both fibrin and platelets. Fibrinolysis is then responsible for dissolving this plug and is initiated by key activators, tPA and uPA, that convert inactive plasminogen to plasmin, an active serine protease that digests the fibrin platelet clot. Plasmin is downregulated by the inhibitor α2-antiplasmin, whereas both uPA and tPA are inhibited by PAI-1.

Fibrinolysis is controlled by inhibitors that block the plasminogen activation step or block plasmin. Plasminogen activator inhibitor 1 (PAI-1) is the key inhibitor of both tPA and uPA, and α2-antiplasmin is the key inhibitor of plasmin [8, 9]. In blood, 90% of PAI-1 is found in platelets, blood cells produced by megakaryocytes (MK) [9]. PAI-1 released from platelets accounts for the antifibrinolytic effect of platelets on fibrin clots. [9, 10].

A disturbance in the fibrinolytic process can lead to states of hypo or hyper fibrinolysis resulting in thrombosis or bleeding respectively. Congenital deficiencies in α2-antiplasmin are associated with severe bleeding tendencies [11]. Similarly, PAI-1 deficiency and subsequent inefficient inhibition of plasminogen activators, is associated with delayed onset of post trauma and surgical bleeding [12-14]. Conversely, a promoter polymorphism of PAI-1 gene results in elevated levels of PAI-1, a condition that leads to reduced fibrinolysis in livedoid vasculopathy [15]. To date, there have been no reports of inherited fibrinolytic disorders that are associated with tPA or uPA deficiencies [16].

My thesis involves investigating Quebec platelet disorder (QPD), a fibrinolytic disorder that has shown to dramatically alter the regulation of uPA in MK. A literature review of uPA (Section 1.3), blood cells involved in QPD (Section 1.4) and current knowledge of this disorder and its pathogenesis (Section 1.5) are presented in the sections that follow.

1.3 Urokinase plasminogen activator

1.3.1 Structure and regulation

uPA is one of two prominent plasminogen activators involved in fibrinolysis [16-20]. The gene encoding for uPA is termed plasminogen activator urokinase (*PLAU*) and it is 6.4 kb and comprised of 11 exons, 10 introns and located on chromosome 10q24. [21, 22]. *PLAU* encodes a 2.4 kb messenger ribonucleic acid (mRNA) that subsequently produces a functional 54 kDa protein comprised of 431 amino acids [23-25]. The 2nd exon of *PLAU* contains the start codon ATG, and together with the first exon forms the 5' untranslated region (UTR) [24]. The coding region of *PLAU* ranges from the 2nd exon to a small upstream region of exon 11, whereas the remainder ~ 800 bp downstream region of exon 11 forms the 3' UTR [24]. The intron exon distribution of *PLAU* shares similar sequence homology (40%) with the exon distribution of the tPA gene [24]. Moreover, the region encoding for uPA catalytic functions bears close resemblance to the distribution of introns and exons in other serine proteases [23]. *PLAU* is flanked by two genes, *calcium/calmodulin-dependent protein kinase II γ* (*CAMK2G*) and *vinculin* (*VCL*) that are approximately 36 kb upstream and 100 kb downstream respectively. A gene termed *C10orf55* is antisense to *PLAU* and spans 12.8 kb in length [26, 27]. *C10orf55* encodes a protein whose functions are not yet characterized [26]. As *C10orf55* has a large sequence that is antisense to coding exons of *PLAU*, *C10orf55* RNA could potentially modulate *PLAU* expression

uPA is produced as a single chain glycoprotein with 3 major domains namely, epidermal growth factor (EGF), kringle and protease domains [1]. The EGF domain of uPA interacts with the membrane bound uPA receptor (uPAR), and the uPA kringle domain stabilizes the interaction [28]. The protease domain possesses a catalytic triad that cleaves plasminogen to plasmin [29]. uPA has been shown to bind uPAR, allowing uPA to localize to plasminogen and upregulate its promotion of plasmin generation by 20 fold [30-32].

The zymogen form of uPA is single chain uPA (scuPA) that in the presence of plasmin, undergoes conversion to the enzymatically active form two chain uPA (tcuPA) [28]. In plasma, scuPA concentrations are normally between 2-4 ng/mL [33-35]. Platelets contain only a small amount of uPA, up to 1.3 ng/10^9 platelets [36-38].

The hemostasis enzymes that potentially activate uPA are plasmin, factor XIIa, and kallikrein [39-41]. In addition to tcuPA, the low molecular weight uPA (LMW uPA) form is an enzymatically active protease that converts plasminogen to plasmin [42, 43]. LMW uPA is formed when the EGF and kringle like domains of tcuPA are proteolysed by plasmin [1]. PAI-2, an inhibitor that acts primarily on uPA, was found to inhibit two chain uPA five times more than two chain tPA (tctPA) and even more than single chain tPA (sctPA). [44, 45].

1.3.2 Fibrinolysis and uPA

The role of uPA in regulating fibrinolysis has been extensively evaluated [20, 32]. In the presence of fibrin, the relatively inactive scuPA, and not active tcuPA, induces fibrin specific lysis [46]. The specificity of scuPA towards fibrin is facilitated by the enhanced

binding of plasminogen to partially degraded fibrin [47]. In essence, plasminogen activation is increased when scuPA binds to uPAR, because of the co-localization of uPA and plasminogen [31, 32, 42, 48-50].

In addition to uPAR, there are other receptors that interact with uPA. A receptor on neutrophils (a granulocyte cell type), termed $\alpha_M\beta_2$ binds uPA and plasminogen to upregulate plasmin production while certain unclassified platelet receptors were shown to bind scuPA [36, 51, 52]. Moreover, uPA on monocytes or endothelial derived microparticles induces activation of platelet bound plasminogen, fibrin and extracellular matrix proteins [53]. This phenomenon of plasminogen activation on a different cell surface was shown to be mediated by uPA and not tPA [1].

1.3.3 Transcriptional regulation of *PLAU*

PLAU gene expression is low in many different cell lines [21]. The uPA-uPAR binding can activate several transcription factors (TFs) and kinases, that may subsequently regulate cell morphology, adhesion, migration and cytoskeletal structure [54-58].

The 5' flanking sequence of *PLAU* has several regulatory elements that are involved in the regulation of this gene [59, 60]. The *PLAU* promoter consists of a TATA-box and a 200 bp GC-rich region. The sequences GGGCCGG and CAAT sequences, both located upstream of *PLAU* transcription start site (TSS), are complimentary to TFs Sp1 and CCAAT box-binding transcription factor (CTF) respectively, that are responsible for the intrinsic low levels of *PLAU* expression [21]. A chloramphenicol acetyl transferase gene reporter assay in two human cell lines indicated that *PLAU* transcription was regulated by an enhancer 2 kb upstream of the TSS [61]. Further studies elucidated that the functioning

of the enhancer is dependent on two intrinsic AP-1 binding sites ($AP\text{-}1_a$ and $AP\text{-}1_b$), where $AP\text{-}1_a$ is combined with an E twenty-six (Ets) PEA3 binding site [20, 21]. The two AP-1 binding sites are separated by a cooperativity mediator (COM) region that in turn has upstream and downstream regions (uCOM and dCOM respectively) [20, 21]. The COM contains binding sites for upstream enhancer factors (UEFs) [21]. The UEFs are as follows, UEF1 (yet to be characterized), Prep-1/Pbx heterodimers and Oct-1 as UEF 2,3 and 4 respectively [20, 21]. The uCOM region has binding sites for UEF 1-4 where as dCOM has binding sites for only UEF 1-3 [59, 60]. The uPA gene can be transcriptionally activated by various stimuli such as phorbol esters and growth factors, mediated through signal transduction pathways that target the 2 kb upstream enhancer of *PLAU* [20, 59, 60, 62]. The regulatory 2 kb upstream enhancer has favorably shown to physically interact with the *PLAU* promoter through the looping of the intervening enhancer-promoter sequence [20, 63].

A nuclear factor kappa-light-chain-enhancer of activated B cells (NF_κ-B) site was found to induce *PLAU* expression independently of the regulatory enhancer AP-1 regions [20, 64]. Moreover, functional studies found a silencer in the 5' regulatory region of human *PLAU* that acted in a cell-specific manner [20, 64, 65]. The regulatory elements involved in transcriptional regulation of *PLAU* are summarized in Figure 2.

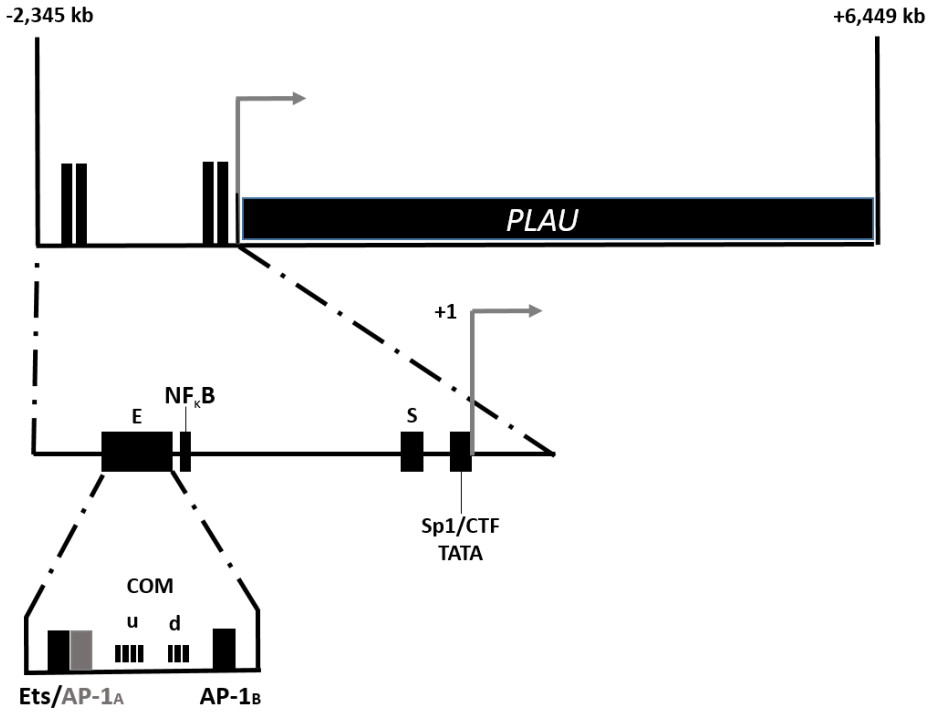

Figure 2. Regulatory regions upstream of *PLAU*. A schematic representation of the regulatory elements 5' flanking sequence of *PLAU* are shown. The *PLAU* TSS is denoted by +1, with upstream TATA box and Sp1 and CTF binding sites. The enhancer (E) located ~ 2 kb upstream of *PLAU* TSS has two AP-1 binding sites (AP-1_a and AP-1_b), that are divided by cooperativity mediator region (COM), with upstream (u) and downstream (d) regions. The nuclear factor kappa-light-chain-enhancer of activated B cells (NF_κ-B) and silencer (S) are represented (Image adapted from [20]).

1.3.4 Murine models of uPA

As neither uPA or tPA deficiencies have been reported in humans, the significance of these plasminogen activators has been studied in mouse models [66]. The uPA murine knockout model (-/-) is viable, without a reduction of lifespan [67]. The knockout mice showed increased deposition of fibrin in normal and inflamed tissues and the macrophages of these animals exhibited reduced, plasmin-induced breakdown of fibrin [68]. The uPA knockout mice have an increased risk of developing endotoxin induced thrombosis while at the same time have impaired inflammatory cell recruitment for T-cells and macrophages, thus increasing risks for contracting bacterial infections [67, 68].

As of now, QPD is the only inherited disorder with markedly increased levels of uPA and this disorder is discussed in detail in Section 1.5. A mouse model of increased uPA has been developed by expressing a *PLAU* transgene under the control of a MK specific promotor [67]. As in humans, the transgenic uPA mice, that overexpress uPA in MK exhibit increased bleeding without evidence of systemic fibrinogenolysis, and they exhibit some other features of QPD, including abnormal proteolytic degradation of α-granule proteins [69]. However, these mice experience an increase in bleeding-associated mortality during pregnancy unless treated with fibrinolytic inhibitors, which is not seen in QPD [69, 70]. The mice that overexpressed uPA in MK displayed resistance towards arterial and venous thrombosis, likely due to their increased platelet uPA [69].

1.4 Hematopoiesis

Hematopoiesis is the process of blood cell development and differentiation that gives rise to all blood cell lineages [71]. Blood cells are derived from pluripotent, self-renewing hematopoietic stem cells (HSC), which are rare in blood (< 0.01%) [72-74]. The CD34 antigen has been established as the major marker of HSC, and accordingly, monoclonal antibodies against CD34 are used to analyze and isolate HSC from blood [75]. The hematopoietic system hierarchy with the HSC at the apex has been established, where each subsequent developmental stage moves towards a specific lineage or group of cells [76]. The HSC gives rise to two progenitors, the common lymphoid progenitor (CLP) and common myeloid progenitor (CMP), that are permanently committed towards producing the lymphoid and myeloid lineages (lymphopoiesis and myelopoiesis) respectively [77, 78]. In lymphopoiesis, the CLP produces T and B cell lymphocytes [76]. Conversely in myelopoiesis, the CMP gives rise to MK, granulocytes (neutrophils, eosinophils and basophils), monocytes and erythrocytes (Figure 3) [76, 79].

MK are the key affected lineage in QPD that produce abnormal levels of uPA and studies have been conducted to assess QPD effects on levels of uPA in granulocytes [80, 81].

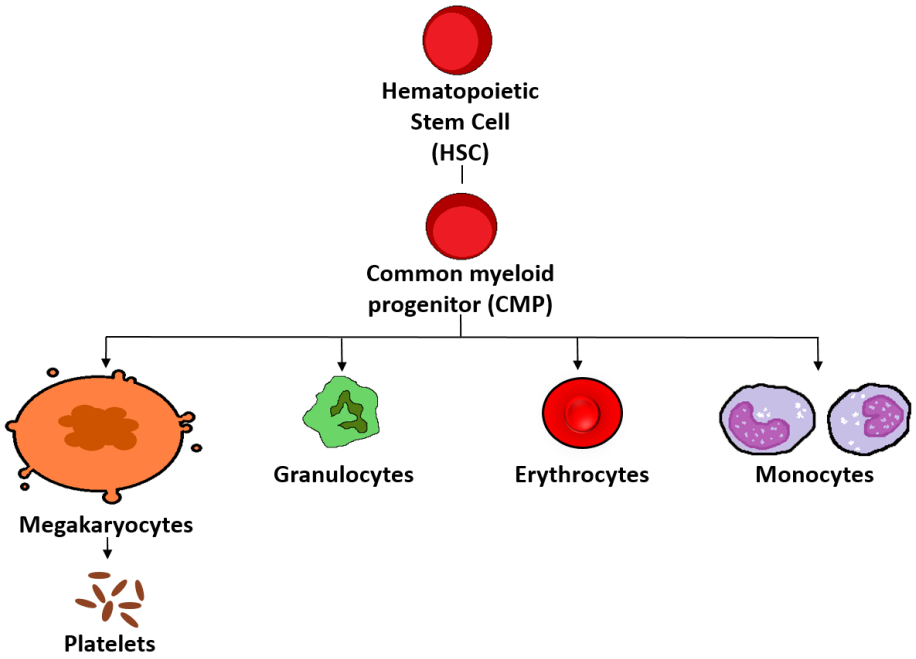

Figure 3. Myeloid lineages derived from hematopoietic stem cells. A multipotent hematopoietic stem cell (HSC), gives rise to a common myeloid progenitor (CMP) that undergoes further differentiation to produce cells of all myeloid lineages; including MK (which generate platelets), granulocytes, erythrocytes and monocytes (Image adapted from [82]).

1.4.1 Megakaryopoiesis

1.4.1.1 Megakaryocyte development

MK are myeloid lineage cells that are derived from pluripotent hematopoietic stem cells (HSCs) that differentiate to generate CMP [83, 84]. CMPs then give rise to a MK/erythrocyte bipotent progenitor (MEP) that may become solely committed towards MK differentiation [1]. The MK differentiation lineage is characterized by three important developmental stages namely proliferation, endoreduplication and maturation. The bipotential MEP has the capacity to differentiate into premature MK burst forming unit (BFU-MK) or the mature colony-forming unit MK (CFU-MK), and both of these MK progenitors display the CD34 marker, which is not found on mature MK or platelets [85].

The defining characteristics of MK maturation are endoreduplication and cytoplasmic expansions [84]. Mature MK, which are responsible for the production of circulating platelets, have a modified cytoplasm [86, 87] and have cell diameters of < 50-100 microns, with a ploidy ranging between 2 N to 64 N, with a maximum of 128 N [88-90]. As MK undergo maturation and cytoplasmic expansion, there are increases in platelet organelles and demarcation membrane system [84]. The MK form proplatelet projections that ultimately produce the de novo circulating platelets [91, 92].

The α and dense granules that are found in circulating platelets appear in the early and later stages of MK maturation respectively [93, 94]. MK synthesize cargo proteins, such as the α-granule protein von Willebrand factor (VWF), along with the α- and δ granule membrane protein P-selectin [95, 96]. MK also synthesize proteins found on their external

and granule membranes, including CD41 or $\alpha_{IIb}\beta_3$, which is often used as a specific marker for MK differentiation [1, 84].

1.4.1.2 Platelet production

Megakaryopoiesis and platelet production (thrombopoiesis) take place within the bone marrow in the presence of chemokines and cytokines [97]. The process of megakaryopoiesis produces 10^{11} platelets per day, as platelets have a turnover of about 8-10 days, and the platelet counts in blood of healthy individuals ranges from 150-400 $x10^9$ platelets/L [1, 84]. MK are capable of increasing platelet production by 10 fold under conditions that demand increasing circulating platelets (e.g., acute blood loss) [98, 99].

As of now, the widely accepted mechanism of thrombopoiesis claims that platelets are synthesized in an assembly line at the end of proplatelets, by budding of the terminal end [86, 100, 101]. The proplatelets extend to sinusoidal areas, where they release and fragment into about 2,000-5,000 new platelets [102, 103].

1.4.1.3 Cytokine regulation of megakaryopoiesis

Thrombopoietin (TPO), alternatively known as c-MPL ligand, is the fundamental growth factor for MK differentiation that also ensures survival and production of HSCs. [99-101, 104]. The cytokine TPO stimulates MK progenitor stimulation and proliferation, where MK increase in size and ploidy [99-101]. TPO significantly enhances the development of proplatelet processes that divide into single platelets for release in blood circulation [98, 99]. Moreover, TPO has been shown to act synergistically with other cytokine factors to promote MK proliferation and differentiation [100]. Stem cell factor

(SCF) is a cytokine that in vitro has shown to mimic the effects of TPO on the growth of MK progenitors, by increasing size and proliferation of MK colonies [105-107]. At the same time, interleukins (IL), such as IL-6 and IL-9, work synergistically with SCF to increase proliferation of MK [108, 109].

1.4.1.4 Transcriptional regulation of megakaryopoiesis

The development of MK and platelet production are processes regulated by a host of TF that regulate the chromatin organization of genes, allowing transcriptional upregulation or downregulation of specific genes crucial to maintain the MK differentiation lineage [84]. A host of MK lineage genes are regulated by globin transcription factor (GATA) family, FOG, along with runt-related TF 1 (RUNX1) and ETS proteins [84]. RUNX1 is a TF that is crucial to the regulation of all hematopoietic lineages [84]. The core binding factor beta (CBFb) complexes with RUNX1, to bind GATA-1 and ensure commitment to the MK lineage [110]. GATA-1 is also crucial for ensuring MK differentiate, and it works in conjunction with other TFs [111]. At one of the primary stages, when commitment towards the MK/Erythroid lineages becomes evident, there is the downregulation of PU.1, a major TF pertaining to myeloid lineage commitment, at the same time that GATA1 expression is upregulated [112]. The antagonistic regulation patterns of GATA and PU.1 TFs dictate lineage commitment decisions [113]. The loss of GATA-1 leads to halt in differentiation, cell death of erythroid progenitors and an increase in the number of immature MK [114]. On the other hand, forced GATA-1 expression ensures the CLP and CMP to differentiate towards the MK/Erythroid lineage [115]. Moreover, GATA-1 is implicated in regulating all specific developmental stages of MK

[98].

MK specific promoters contain sequences for binding members of the Ets family of TF, that are located proximal to GATA binding sequences [100, 116]. The Ets TF family member, Friend leukemia integration-1 (FLI1), is expressed in heamangioblasts and early MK [100]. Mutations in FLI1 and RUNX1 have been reported to cause congenital platelet function disorders associated with thrombocytopenia [117].

1.4.2 Granulopoiesis

1.4.2.1 Granulocyte development and transcriptional regulation

In addition to MK, Granulocytes are also members of the myeloid lineage that are derived from HSCs. Granulocytes have several defining characteristics; they are the most abundant myeloid cells in blood, and have a multilobulated nucleus and also granules (storage vesicles) in their cytoplasm [118]. Granulocytes can be divided into three categories: neutrophils, basophils and eosinophils [118]. The most abundant form of granulocytes are neutrophils, which are produced in large amounts from the bone marrow and released into the blood circulation [118].

Similar to MK, granulocytes arise from the CMP, but then develop into granulocyte-monocyte progenitor (GMP) cells [118]. Granulopoiesis is regulated by GATA-1 and Ets family derived PU.1 TFs and the latter has been implicated as a crucial regulator for myeloid lineage commitment [118]. The TF CCAAT-enhancer-binding proteins alpha (C/EBPα) has been implicated as a major deterministic factor for granulopoiesis [119]. A deletion of C/EBPα in bone marrow cells of knockout mice models

resulted in the absence of mature granulocytes at the CMP to GMP transitional stage [119].

C/EBPα is a crucial TF for the CMP-GMP step. The decision for the GMP to divide either into granulocytes or monocytes is influenced by the levels of PU.1 [118]. High levels of PU.1 expression are associated with monopoiesis (formation of monocytes), whereas suppression of PU.1 allows C/EBPα to induce granulopoiesis and not monopoiesis [118].

1.5 Quebec platelet disorder

1.5.1 Clinical features and symptoms

Quebec platelet disorder (QPD) is an inherited bleeding disorder, characterized by a unique cellular defect that accelerates clot lysis [120]. Inherited bleeding disorders with a gain of function defect are rare and include QPD, as this disorder is associated with a tremendous increase in uPA expression [121-123]. In the province of Quebec in Canada, the prevalence of QPD was found to be 1 in 220,000 people [70]. QPD cases have been reported in Canada and the USA, where a majority of QPD affected individuals have been traced back to single family of French ancestry from Quebec [70].

The prominent symptoms associated with QPD individuals are delayed onset bleeding following trauma or invasive surgery that can only be treated or prevented with fibrinolytic inhibitor therapy by prescribing tranexamic acid or amino caproic acid [37]. Individuals with QPD, after an exposure to trauma, can develop extensive bruising and hematomas [70]. Some experience joint bleeds that can lead to arthropathy [70, 124]. Some individuals with QPD (~ 50%) experience mild hematuria and ~ 26% experience issues with wound healing, which is associated with having lower platelet counts [70]. Bleeding

scores in persons with QPD can vary, and tend to be low in those without prior and/or untreated hemostatic challenges [70].

Only some women with QPD experience menorrhagia and most women with the disorder have favorable outcomes with uncomplicated pregnancies, which do not require therapy [70]. There is no record of a person with QPD having suffered a myocardial infarction or a spontaneous thrombotic stroke, however some have experienced intracranial hemorrhage and hemorrhagic stroke [70]. Some individuals with QPD, during a period of high risk for thrombotic events and treatment with antifibrinolytic drugs, have experienced venous thrombotic events [1]. Nonetheless, the predominant clinical features of QPD are restricted to bleeding symptoms [122].

Some individuals with QPD have mild thrombocytopenia, as their platelet counts are reduced by about 50% compared to unaffected family members [70]. The cause for decreased platelet numbers remains to be resolved, as the number and morphology of MK, and platelets, appear to be normal [125].

Persons with QPD have normal plasma levels of coagulation and fibrinolytic proteins. In some cases, a mild plasma factor V deficiency has been noted, but this is not evident when samples are rapidly prepared, with inhibitors of platelet activation and ex-vivo plasmin generation [125]. Platelet factor V levels are abnormal in QPD, due to factor V degradation. [120, 124-127].

QPD is associated with some abnormalities in platelet aggregation tests, the cause of which is not fully understood. The characteristic finding is reduced primary and or absent secondary aggregation with epinephrine, although QPD platelets have normal

numbers of epinephrine receptors [122, 124, 126]. Some individuals with QPD also have reduced platelet aggregation responses to ADP and collagen [120, 124, 128].

1.5.2 QPD and uPA fibrinolysis

The paramount feature of QPD is the marked overexpression of uPA in MK and platelets [37, 129, 130]. The plasma levels of uPA in this disorder is normal or only modestly increased, whereas the urinary uPA levels are normal; the discrepancy between platelet compared to urinary and plasma uPA levels in QPD, established that the defect in uPA overexpression is a cell-type specific defect [120, 129, 130]. D-dimer levels were found to be significantly higher in the serum of clotted QPD whole blood compared to QPD platelet poor plasma, as the former group had platelet uPA that increased clot lysis [130].

In QPD, plasma levels of D-dimers, α2-antiplasimn-plasmin and uPA-PAI-1 complexes are normal [37, 129, 130]. In addition, plasminogen levels in QPD plasma are also normal [37, 129, 130]. On the other hand, platelets from persons with QPD have high levels of uPA-PAI-1 complexes, and α2-antiplasmin-plasmin complexes, which are consistent with intraplatelet consumption of PAI-1 and intraplatelet plasmin generation [37, 129, 130]. The excess uPA in QPD platelets fully consumes the active forms of PAI-1 found in platelets, which normally makes platelets anti-fibrinolytic [37, 81, 122]. The most abundant forms of uPA in QPD platelets are active tcuPA and LMW uPA (resulting from proteolytic cleavage of scuPA), although QPD platelets also contain some further degraded forms of uPA and uPA complexed to PAI-1 [37, 81].

1.5.3 QPD and uPA overexpression during megakaryopoiesis

The cause of the marked overproduction of uPA in QPD platelets, leading to excess plasmin production within these cells, is not fully understood [122]. The markedly increased uPA levels in QPD platelets triggers plasmin mediated degradation of diverse α granule proteins including factor V, VWF, multimerin 1 (MMRN1), thrombospondin 1, fibrinogen, fibronectin, osteonectin and P-selectin [37, 123, 124, 126, 127, 129, 131]. QPD platelets also accelerate clot lysis when platelets are activated and release uPA [130].

Insights on the cause of QPD have come from the analysis of *PLAU* mRNA and uPA protein levels during megakaryopoiesis in QPD and control subjects [81]. Firstly, QPD and control $CD34^+$ HSCs have similar *PLAU* mRNA levels. A significant finding of QPD was observing an unexpected abnormal increase in *PLAU* transcription and uPA protein production as $CD34^+$ HSC differentiated into MK in culture [81]. During normal megakaryopoiesis, uPA continues to be expressed and produced at low levels [81]. In QPD however, *PLAU* mRNA and uPA expression is increased by about 4 fold by Day 7 and >100 fold by Day 13 of MK cultures [81]. In addition, platelets derived from peripheral blood of QPD individuals contain a >100 fold increase in *PLAU* transcripts and uPA protein compared to controls [81]. However, these abnormalities were restricted to *PLAU*, as the expression of *VCL* (the gene downstream of *PLAU*) and VWF were not significantly different between QPD and control samples of Day 13 cultured MK or peripheral blood platelets [81]. Moreover, in Day13 MK cultures, there were no significant differences in the expression of *CAMK2G* (the gene upstream of *PLAU*) between QPD and control subjects [81].

In analyzing the protein levels, >100 fold increase in uPA protein was also observed in QPD Day 13 MK cultures compared to controls [81]. The levels of uPA between pooled sets of QPD and control peripheral blood isolated $CD34^+$ HSC were within the range of one another [81]. The analysis of α granule proteins, platelet factor 4 (PF4), VWF, thrombospondin 1 and uPA inhibitor PAI-1 were found to be normal in both QPD and control MK cultures harvested on Days 7 and 13 [81]. The timing of the increased production of uPA in QPD MK cultures was similar to the timing of the increased production of thrombospondin 1, PF4 and VWF [81]. tPA is not detected in QPD and control and MK cultures, indicating that uPA is the sole plasminogen activator produced during MK differentiation in culture [81]. An analysis of uPAR in QPD and control platelets revealed that it was expressed in similar low amounts [81].

1.5.4 QPD and *PLAU* linkage analysis

A genetic linkage analysis determined that QPD is linked to the *PLAU* gene on chromosome 10q [121]. DNA sequencing was then used to exclude mutations within the *PLAU* gene [121]. Moreover, no mutations were found in the 5' and 3' regulatory regions of *PLAU* or 24 kb upstream and 2 kb downstream of the TSS and 3' UTR respectively [121]. A comprehensive sequencing for the aforementioned region in both *PLAU* alleles was confirmed [121].

1.5.5 QPD *PLAU* mutation

As sequences did not reveal abnormal changes in DNA sequence of the *PLAU* gene in QPD, the possibility of a copy number variation (CNV) was investigated [132]. Analysis

of the *PLAU* region in QPD individuals, in fact, revealed the presence of one extra copy of *PLAU* [132]. The duplication boundaries in the *PLAU* region were estimated using real time dosage analyses and estimated to lie between 75 658 940 and 75 659 061 bp at the centromeric end and between 75 736 665 and 75 737 145 bp at the telomeric end (hg19 genome assembly) [132]. The duplication was concluded to be local as fluorescence in situ hybridization analysis showed QPD chromosomes having normal hybridization patterns with *PLAU* and chromosome 10 probes [132].

In order to assess the orientation of the duplication segment, PCR assays were employed using primers specifically targeting direct tandem *PLAU* duplication [132]. The 930 bp fragment was amplified in QPD subjects and not controls thus confirming the tandem duplication of *PLAU* in QPD [132]. Bidirectional sequencing of the fragment determined the boundaries of the QPD breakpoint sequence, ~11.87 kb upstream of telomeric *PLAU* TSS and ~59.69 kb downstream of the centromeric copy of *PLAU* [132]. Alignment of QPD sequences prodded the discovery of a direct tandem duplication of a ~78 kb segment of the *PLAU* region [132].

Analysis on more QPD subjects for the breakpoint region confirmed its presence and thus established the *PLAU* CNV as the causative mutation for QPD [132]. The QPD overexpression abnormality has been linked to a tandem, duplication mutation of ~78 kb region on chromosome 10 that includes the *PLAU* and *C10orf55* on the antisense strand [132]. The two *PLAU* genes on the diseased chromosome each contain the normal regulatory elements, in addition to large upstream and downstream regions [132]. As a result of the *PLAU* CNV mutation, a new region termed the QPD breakpoint is introduced

into the *PLAU* gene landscape [132]. Interestingly, the QPD breakpoint has motifs for megakaryopoiesis related Forkhead domain factors, suggesting that these factors could potentially regulate *PLAU* in QPD [132-134].

Overall, in light of the *PLAU* CNV mutation discovery, QPD is the only inherited bleeding disorder known to be caused by gene duplication event and it is also the only bleeding disorder to be associated with a *PLAU* mutation [132].

1.5.6 QPD MK specificity and unaffected QPD cell lineages

While a duplication mutation would be expected to increase uPA expression and protein levels by 50% (1.5 times normal), QPD MK and platelets express uPA mRNA at >100 fold increased levels, with an accompanying >100-fold increases in uPA protein levels [81]. As of now, MK and platelets are the only cells known to show the dramatic uPA overexpression in QPD [81]. Analysis of uPA expression in other cell types is provided below.

The increased *PLAU* expression for QPD $CD34^+$ HSCs and saliva cells is within the range of having one extra copy of *PLAU* [81, 121]. The SNP rs4065 (T/C) analysis in heterozygous QPD subjects was conducted previously to determine if QPD had any cis or trans effect on *PLAU* expression [121]. Interestingly, the T/C allelic ratio was >150 in QPD platelets in contrast to the T/C ratio of 1.5 for healthy control platelets [121]. The T/C allele ratios at rs4065 for QPD $CD34^+$ HSC and saliva cells (unaffected cells with normal levels of *PLAU* mRNA) were found to be more similar to corresponding control cells, with modest 4 fold and 2 fold increases for QPD cells respectively [121]. As mutations in *PLAU* mRNA stability elements were excluded by Sanger sequencing, it was established that

during MK differentiation, QPD results in a cis regulatory defect that enhances the rate of transcription to the linked allele of *PLAU* [121].

Recently, studies have been carried out to investigate uPA protein levels in myeloid lineages such as granulocytes [80]. A 2.4 fold increase in uPA protein levels between QPD and control granulocytes has been observed [80]. The modest 2.4 fold increase observed is significantly lower than the dramatic increase in uPA levels observed in QPD versus control MK [81]. The protein levels make a strong case for granulocytes as a QPD unaffected myeloid lineage. Moreover, it strengthens the notion of the QPD defect being specific towards MK.

Based on the modestly increased uPA protein levels in QPD granulocytes compared to controls, granulocytes do not show the tremendous increases in uPA found in QPD MK and platelets, even though they are also derived by myeloid differentiation. I propose that granulocytes could thus be useful as a reference for an unaffected myeloid lineage for further studies of *PLAU* regulation in QPD.

1.6 Chromatin structure and regulation of gene expression

In order to investigate alterations in *PLAU* gene expression, it is necessary to have a basic understanding of gene regulation.

In eukaryotes, DNA is found in compact domains termed chromatin [135]. A nucleosome, the basic structural unit of chromatin, is found to have ~1.65 turns of DNA around an octamer of histones that include two of each H2A, H2B, H3 and H4 histones [135]. The core histone proteins are composed of a tightly structured globular domain along

with a more flexible N-terminus tail that protrudes out of the nucleosome [135]. The N-terminus tails have large numbers of positively charged lysine residues that attract the negatively charged DNA, forming a compact structure [135].

In essence, the first step of gene expression is transcription of the DNA from a gene into mRNA by the action of RNA polymerase II [136]. Gene promoter, a region located within the vicinity of the TSS, comprises of regulatory elements involved in gene expression [137-139]. A locus within the gene promoter termed core promoter, serves as a platform for assembling the transcriptional machinery complex, that includes RNA polymerase II for transcribing RNA and general TFs that act as cofactors for RNA synthesis [136]. The basal rate of transcription is low in the absence of enhancers also known as cis-regulatory modules that are located at a distance from the TSS [136]. The sequences of enhancers contain sequence binding motifs for specific TFs [136]. The TFs, in turn, can recruit cooperative proteins such as co-activators and co-repressors, which form a complex that determines the activity of the enhancer and level of target gene expression [136]. The chromatin structure correlates with enhancer activity, as at sites where chromatin is in an open conformation, there is increased accessibility to TF binding [136].

A characteristic feature of enhancers is the ability to influence transcription of a gene, regardless of distance and strand orientation. Indeed, enhancers can function on genes at seemingly large distances away (e.g., several 100,000 bp or even Mbp) through the action of looping that brings them physically closer to gene promoters [140]. Moreover, enhancers have modular properties, where they contribute additionally to the expression trend of their target genes [136]. This trend of enhancer influence on gene expression can

be observed in reporter assays, where the combination of multiple enhancer sequences results in gene expression trends that are reflective of their combined activity [136, 141].

1.7 Epigenetics

QPD presents a unique abnormality, where the presence of a 78 kb duplicated region containing *PLAU* results in a dramatic >100-fold increase of *PLAU* expression in MK. In order to address this issue, it is crucial to understand regulatory mechanisms associated with gene expression, primarily epigenetics.

Epigenetics can be defined as stable, inheritable changes that modify chromatin structure without changes in the DNA sequence [142, 143]. Epigenetic modifications include promoter CpG island DNA methylation, histone modifications, histone variants and nucleosome positioning [142]. As epigenetic modifications do not associate with specific DNA sequences, it gives them the ability to maintain a chromatin state in any chromosomal location in the genome [144]. Moreover, epigenetic modifications are found to sustain epigenetic domains in terminally differentiated cell types [144]. The two major forms of epigenetic modifications that will be focused on are promoter CpG island DNA methylation and histone modifications.

1.7.1 DNA methylation

DNA methylation is an inheritable, epigenetic modification which plays a significant role in gene expression [145, 146] and it has been implicated in maintaining cellular fate and differentiation [147, 148]. DNA methylation is a process characterized by methylation of cytosine residues that mostly occurs in CpG dinucleotide dense regions

termed CpG islands [126]. It is estimated that almost half of all gene promoter regions have CpG islands [149, 150]. A large degree of methylation in the CpG islands of promoter regions has been associated with transcriptional silencing [151]. DNA hyper-methylation of CpG promoter islands has been associated with reduced chromatin accessibility and therefore low gene expression [152]. Conversely, hypo-methylation of gene promoter regions results in an open chromatin conformation, thus potentially increasing TF accessibility and increased gene expression [152]. In vitro studies have shown DNA methylation can directly interfere with TF binding [153]. Moreover, once CpG islands are methylated, methyl-CpG binding domain proteins (MBDs) bind to those regions causing potential interference with TF binding and altering chromatin structure [151]. MBDs have been shown to recruit histone modifying proteins that alter chromatin density resulting in decreased accessibility and transcriptional repression [153]. One of the most prominent commercial assays, the Infinium HumanMethylation450K BeadChip Array is a widely used approach to determine the methylation status of CpG islands in the genome [154].

MethBase is a central reference methylome database derived from bi-sulphite sequencing projects [155]. As of now, MethBase does not contain data on genome wide promoter CpG island methylation status from MK and granulocytes, however it does have data from other blood cells, particularly PBMCs, $CD4^+$ T cells, B cells and CD133 HSC [155]. MethBase is a central reference methylome database derived from bisulphite sequencing projects [155]. In analyzing methylation status on blood cells using MethBase on the UCSC genome browser, the promoter CpG island of *PLAU* was found to be hypo-

methylated in all blood cells namely neutrophils, PBMCs, $CD4^+$ T cells, B cells and CD133 HSC [155].

1.7.2 Histone modifications

Post translational modification of N-terminus tails on histones can have a profound effect on modifying chromatin structure and the binding of DNA regulatory binding proteins [156]. As a result, histone modification can exhibit a profound influence on gene transcription [154]. Examples of some common histone modifications include acetylation, methylation, phosphorylation, ubiquitination and sumoylation [154]. The two forms of histone modifications that have been widely characterized in relation to the regulation of gene expression are acetylation and methylation.

Acetylation of lysine residues on N-terminus histone tails reduces the positive charge of histones and as a result it reduces the electrostatic interactions between negatively charged DNA and positively charged histones [156]. The result is that the chromatin region is less compact and therefore its DNA is more accessible for binding protein complexes that are involved in gene transcription [156]. Histone modifications by acetylation were found to be enriched in gene regulatory regions such as enhancers, where they are predicted to allow the binding of TFs [157].

Histone methylation, unlike acetylation, does not influence the charge of histones and can be found in mono, di and tri methylated forms [143, 158]. Chromatin associated factors have been shown to specifically interact with methylated histone modifications via several unique protein domains. For example, PHD fingers, chromodomain, tudor and MBT domains are found in chromatin binding protein complexes that bind methylated

lysine residues of histones [156, 159, 160]. Histone methylation enrichment at gene promoters, gene bodies and enhancers are associated with altered chromatin accessibility for TF and gene expression [149, 158, 161, 162].

As QPD is a gain of function defect associated with a dramatic increase in *PLAU* expression, histone modifications associated with increased chromatin accessibility were of particular interest. Prominent examples of such active histone modifications include H3K27Ac (Histone 3, Lysine residue 27, Acetylation), H3K36me3 (Histone 3, Lysine residue 36, tri-methylation) and H3K4me2 (Histone 3, Lysine residue 4, di-methylation). H3K27Ac as is a characterized enhancer mark, that is able to distinguish active from poised enhancer elements [163]. H3K27Ac modification is deposited by CREB binding protein and p300 [163]. The H3K36me3 histone marker is found in regions of chromatin that bind RNA POL II and are associated with active transcription [164]. H3K36me3 is enriched mostly in the bodies of active genes [164]. H3K36me3 has been found to peak and decline intermittently within active gene bodies [164]. Finally, the histone mark H3K4me2 is primarily found in promoters of transcriptionally active genes, along with being found in genes that are primed for future expression during cell development [165, 166]. In addition to being marked at promoters, H3K4me2 also strongly associates with active enhancers [167, 168]. Chromatin immunoprecipitation with next generation sequencing (ChIP-seq) is a well-established procedure that allows the genome wide characterization of histone modifications of interest [169].

There has been little work done characterizing the histone marker enrichment for $CD41^+$ MK. As extracting MK from bone marrow is a tedious process, MK from peripheral

or cord blood derived $CD34^+$ HSC cultures are preferred sources [170]. All ChIP-Seq histone modification studies to date have been conducted using cord blood derived $CD34^+$ HSC that are cultured to $CD41^+$ MK [171, 172]. ChIP-seq of prominent TF involved in megakaryopoiesis, GATA-1/2, RUNX1, FLI1 and SCL have been characterized [171]. Moreover, the same group analyzed the histone modifications H3K4me1 and H3K4me3, prominent markers for promoters and enhancers respectively [168, 171, 173-175]. BLUEPRINT has recently emerged as a new epigenome project, focused on characterizing selective histone modifications in cells from the hematopoietic system [172]. In addition to the aforementioned studies, BLUEPRINT has profiled H3K27Ac and H3K36me3 in cord blood derived $CD41^+$ MK (accession ID: EGAD00001000916) [172].

1.8 Epigenetics and QPD *PLAU* expression

The most important unanswered question about QPD still needs to be solved: How does the presence of a 78 kb duplication of the *PLAU* region result in >100 fold levels in *PLAU* and uPA expression? This question led me to investigate the regulatory mechanism at *PLAU* that causes such a dramatic imbalance in the gene's expression in QPD MK. The study of epigenetics can provide great insight into the regulatory mechanisms of *PLAU* in QPD MK, which have not been previously established. I postulated that the increase in *PLAU* expression and uPA content in QPD MK was linked to altered epigenetic regulation. I also postulated that studying granulocytes, as a non-affected QPD myeloid lineage, would provide useful information for comparison with MK. Plausible epigenetic mechanisms that could be linked to *PLAU* dysregulation in QPD MK are DNA methylation of promoter CpG islands and enrichment of active histone modifications.

The hypo-methylation of gene promoter CpG islands are associated with active gene expression [152]. As QPD is a gain of function defect, the *PLAU* promoter could be hypo-methylated in QPD MK. Moreover, there may be QPD cell specific changes to the methylation status of *PLAU* in MK and granulocytes.

As previously described, active histone modifications at promoters and enhancers are crucial regulators of gene expression [136]. The introduction of active histone enrichments, particularly for active enhancers in QPD due to the *PLAU* duplication could have a profound effect in regulating *PLAU* expression, and it might result in changes in expression patterns that are cell-type specific. The *PLAU* duplication could result in the rearrangement of active histone enrichments, such as active enhancers, being positioned upstream to the second copy of *PLAU* in MK but not granulocytes. Moreover, the *PLAU* mutation can introduce aberrant enrichment in QPD MK with the possibility of a functional enhancer or promoter, especially at the novel QPD breakpoint. One might also expect differences of active histone enrichments in QPD MK versus QPD granulocytes. Overall, knowledge of the epigenetic landscape of *PLAU* region will be critical to assessing the regulation of *PLAU* expression in QPD.

2. THESIS HYPOTHESIS AND SPECIFIC OBJECTIVES

2.1 Hypothesis

QPD is associated with *PLAU* promoter hypo-methylation in MK and/or changes and rearrangements of MK specific active histone modifications, particularly active enhancers in the region of *PLAU* duplication.

2.2 Objective#1

Evaluate promoter CpG island methylation status using genome wide Infinium HumanMethylation450K BeadChip Array at *PLAU* in MK and granulocytes prepared from QPD and control subjects.

2.3 Objective#2

Evaluate active histone enrichments (H3K27Ac-active enhancers, H3K36me3-active transcription and H3K4me2-active promoters and enhancers) using genome wide ChIP-seq, at the *PLAU* locus in MK and granulocytes prepared from QPD and control subjects.

2.3.1 Secondary objective

Evaluate active histone enrichment (H3K27Ac, H3K36me3 and H3K4me2) at the QPD breakpoint in QPD MK and granulocytes.

3. MATERIALS AND METHODS

The research study was conducted with institutional ethics board approval from the Hamilton Integrated Research Ethics Board and the research ethics board of Centre Hospitalier Universitaire Sainte Justine. Blood samples were collected with the informed consent of blood donors in Hamilton and Montreal. Donors included six healthy controls (designated C51, C102, C104, C108, C112 and C117) and six persons with QPD (designated P10, P126, P132, P182, P183 and P186).

3.1 Blood collection

The volume of blood collected for each sample donation ranged from 10-180 mL. For $CD66b^+$ isolations, blood was collected using unfractionated heparin anticoagulant (final concentration: 10 units/mL; Sandoz Canada Inc., Boucherville, QC, Canada). For isolation of $CD34^+$ cells, peripheral blood was collected into sterile acid citrate dextrose anticoagulant (ACD: 6 g citric acid, 12 g sodium citrate, 11.5 g dextrose in 500 mL of distilled water), at 1:6 vol/vol (anticoagulant/blood) and supplemented with 1 mM theophylline, 3 μM aprotinin and 3 μM prostaglandin E_1 (PGE_1) to limit platelet activation and plasmin generation during processing. The reagents, theophylline, aprotinin and PGE_1 were purchased from Sigma-Aldrich, St. Louis, Missouri, USA with Catalogue #s T-1633, A-1153 and P5515 respectively.

3.2 Isolation of CD66b$^+$ granulocytes from peripheral blood

CD66b$^+$ granulocytes were isolated by immuno-magnetic techniques from either 10 or 180 mL peripheral blood samples using EasySep Human Whole Blood CD66b$^+$ Selection Kit (Catalogue # 18682, Stem Cell Technologies, Vancouver, BC, Canada), as per the manufacturer's protocol. Briefly, samples were divided into 5 mL aliquots and diluted with an equal volume (5 mL) of 1x EasySep red blood cells (RBC) lysis buffer. Samples were then incubated with 250 µL (25µL/mL) EasySep human whole blood CD66b positive selection cocktail for 15 minutes at room temperature (RT). Next, samples were incubated for 10 minutes, RT, with 250 µL (25µL/mL) of EasySep magnetic particles before isolating the cells using the Easy Sep magnet (Stem Cell Technologies, Vancouver, Canada). Isolated cells were washed twice with 10 mL of recommended medium (RM) [composition: phosphate buffered saline (PBS) with 2% fetal bovine serum (FBS) and 1mM ethylenediaminetetraacetic acid (EDTA)]. Cells were then resuspended in 2 mL PBS. Harvested cells were counted after a 1:20 dilution in PBS using a C-Chip (Catalogue # DHC-N01, INCYTO, Covington, GA, USA) hemocytometer to determine the cells/mL.

3.3 Isolation of CD34$^+$ HSCs

CD34$^+$ HSCs were isolated from 180 mL blood samples under sterile conditions. Samples were centrifuged to obtain platelet rich plasma (PRP) at 150g for 20 minutes. After removing approximately 2/3 of the PRP, samples were aliquoted into 12.5 mL volumes, then diluted with 17.5 mL of CD34$^+$ buffer, with the following ingredients: PBS

containing: 0.8% bovine serum albumin (BSA, Catalogue # A9418, Sigma-Aldrich, St. Louis, Missouri, USA), 2 mM EDTA, 1 mM theophylline, 1 µM PGE_1 and 1 µM aprotinin. Diluted blood suspensions were then layered onto 15 mL of Ficoll Paque Plus (Catalogue # 17-1440-03, GE Healthcare, Little Chalfont, Buckinghamshire, United Kingdom) using SepMate gradient tubes (Catalogue #15450, Stem Cell Technologies, Vancouver, BC, Canada). Samples were then centrifuged at 1200g for 10 minutes and the peripheral blood mononuclear cell (PBMC) layer was collected. After pelleting the cells (300g, 10 minutes), PBMCs were combined, resuspended into 25 mL of $CD34^+$ buffer and layered onto 25 mL of an Opti-Prep plus density gradient (Catalogue # 1114542, Axis-Shield Dundee, Scotland) to further deplete any contaminating platelets. After centrifuging the samples (350 g, 15 minutes), the supernatant was discarded, and the cell pellet was washed with 50 mL of $CD34^+$ buffer, pelleted (200 g, 12 minutes) and resuspended in 2 mL of $CD34^+$ buffer. Next, $CD34^+$ cells were affinity isolated from the harvested PBMC by incubation with 200 µL (100 µL/mL) StemSep human CD34 selection cocktail (15 minutes at 4°C), followed by incubation with 120 µL (60 µL/mL) StemSep magnetic colloid (10 minutes at 4°C) according to manufacturer's protocol (Catalogue # 14756, Stem Cell Technologies Vancouver, BC, Canada). The cell suspension was washed with 50 mL of $CD34^+$ buffer to remove unbound antibody and subsequently loaded onto an LS column with filter on a magnetic QuadroMACS separator (Catalogue # 130-042-401, Miltenyi Biotech Inc., San Diego, CA, USA). The column was washed three times with 3 mL $CD34^+$ buffer before detaching the column from the magnetic stand and plunging out the cells with 5 mL of $CD34^+$ buffer. Eluted cells were passed through a fresh LS column, washed three times

with 3 mL $CD34^+$ wash buffer before their final elution with 5 mL $CD34^+$ buffer. Next, the final 5 mL $CD34^+$ HSC suspension was pelleted (350g, 12 minutes) and cells were suspended in 1 mL of StemSpan Serum-Free Expansion Medium (SFEM) II media (Catalogue # 09605, Stem Cell Technologies Vancouver, BC, Canada). $CD34^+$ HSCs were counted after a 1:5 dilution in PBS, using a C-Chip hemocytometer (Catalogue # DHC-N01, INCYTO Covington, GA, USA) to determine the yield of $CD34^+$ HSC in cells/mL.

3.4 $CD34^+$ HSC culturing to $CD41^+$ MK

Isolated $CD34^+$ cells (with confirmed viability >90%) were cultured at cell density $2x10^5$ cells/mL in six well culture plates (Catalogue # 657185, Greiner Bio-One North America Inc., Monroe, North Carolina, USA) using SFEM II media with 1:100 v/v, MK expansion supplement comprising of TPO, SCF, IL-6 and IL-9 (Catalogue #s 09605, 02696, Stem Cell Technologies Vancouver, BC, Canada). $CD34^+$ cells were cultured for 14 days at 37^0C, with 5% CO_2. $CD34^+$ cultured cells were replenished with fresh SFEMII and MK expansion supplement on Day 7. MK were harvested on Day 14 and assessed for viability using trypan blue exclusion (viability >90% considered acceptable). Briefly, 10 µL of trypan blue (Catalogue # 15250-061, Thermo Fisher Scientific Waltham, MA, USA) was added to 100 µl cell suspensions to determine the viability of the cell culture populations.

3.5 Quality control analysis of $CD41^+$ MK and $CD66b^+$ granulocytes

3.5.1 Flow cytometry

Day 14 harvested MK were assessed for $CD41^+$ expression, a specific surface marker associated with MK differentiation, [101] using flow cytometry. A 4 channel flow cytometer (Beckman Coulter Epics XL-MCL, Brea, CA, USA) was used to assess the harvested cells. The visualization and statistical analysis of cell purity was conducted using FlowJo software (Tree Star, Inc., Ashland, Oregon, USA). Control isotype matched antibodies were run at equal concentrations to the primary target antibody of interest. The control isotype antibodies were used to verify the specific binding of the primary target antibody. The monoclonal antibodies used for flow analysis were obtained from Becton Dickinson (BD) Biosciences, Mississauga, ON, Canada.

CD41 expression was analyzed by labelling 2 sets of $1x10^5$ MK cell populations (diluted in 50 µL of PBS), one with 7 µl of FITC Mouse Anti-Human CD41 antibody (TEST sample) (Catalogue # 555466) and the other with 7 µl of ISO-FITC IgG1 K Mouse antibody (ISO sample) (Catalogue # 555748). The samples were incubated for 20 mins at 4°C in the dark. Samples were then diluted with 400 µL of Flow buffer (PBS, 2mM EDTA) before analysis on the flow cytometer (Beckman Coulter Epics XL-MCL, Brea, CA, USA). A minimum of 5,000 events were collected for both TEST and ISO samples. Using FlowJo software (Tree Star, Inc., Ashland, Oregon, USA), a gate using forward scatter (FSC) and side scatter (SSC) linear scales was created around the viable cell population in order to exclude cellular debris and cells that had undergone apoptosis. An additional gate was

created around $CD41^+$ cells using FSC linear scale vs flow channel 1 log scale. Purity was determined by the percentage of cells positive for CD41 staining compared to nonspecific staining using isotype antibody. The purity of $CD66b^+$ granulocytes was assessed similarly with the exception that 7 µl of FITC Mouse Anti-Human CD66b (Catalogue # 555724) was used for the TEST sample.

3.5.2 uPA and VWF quantification in Day 14 MK cultures

uPA enzyme-linked immunosorbent assay (ELISA) (Catalogue # 06489892, Nuclea Biotechnologies Inc., Pittsfield, MA, USA) and VWF ELISA (DAKO Canada Inc., Mississauga, ON, Canada) were used as directed by the assay manufacturer to quantify uPA and VWF, in culture media of Day 14 harvested MK cultures to confirm that these cultures were producing the expected amounts of α-granule proteins.

For the uPA ELISA, 100 µl of undiluted control Day 14 MK culture media and 100 µl of 1:20 diluted QPD Day 14 MK culture media were loaded onto uPA ELISA plates. Briefly, samples and standards (100 µL final volume) were incubated (2 hours, 37ºC) in duplicates on the microtiter plates that were pre-coated with the capture antibody, as per manufacturer's recommendations. Plates were then washed 6 times with 300 µL wash buffer. All subsequent wash steps were performed in the same manner. The samples were then incubated with 100 µL of detector anti-uPA rabbit antiserum for 1 hour at RT, followed by wash steps. The detector antibody-antigen complexes were then measured by adding 100 µL of goat-anti-rabbit-IgG-horseradish peroxidase (HRP) conjugate, for 30 minutes at RT. Samples were then incubated with 100 µL of o-phenylenediamine (OPD)

substrate for 45 minutes at RT in the dark, and the colorimetric reaction was stopped by adding 100 µL of 2.5 M H_2SO_4 before reading the plate at 495 nm.

For the VWF ELISA, NUNC Maxisorp plates (Catalog # 442404, eBioscience Inc., San Diego, CA, USA) were coated overnight at 4°C with polyclonal rabbit anti-human VWF antibodies (A0082, DAKO Canada Inc., Mississauga, ON, Canada) at a dilution of 1:600 in coat buffer (14.2 mM Na_2CO_3, 35mM $NaHCO_3$ pH 9.6). Plates were then washed three times with 200 µL wash buffer (PBS-Tween 20 (PBS-T) 0.1%, 0.15 M NaCl). Plates were then blocked in 100 µL blocking buffer (PBS-T, 0.5 M NaCl, 2% BSA) at RT for 2 hours followed by three washes. Samples (100 µL, diluted in blocking buffer), were loaded onto the plate in triplicates and incubated at RT for 2 hours. For detection, peroxidase conjugated rabbit anti-human VWF (P0226, DAKO Canada Inc., Mississauga, ON, Canada) was used at 1:8000 dilution at 100 µL final volume in blocking buffer, incubated at RT for 1 hour followed by washes. Plates were developed using 100 µL of 3,3',5,5'-tetramethylbenzidine (TMB) substrate (Catalog# 90101, AlerCHEK Inc., Springvale, Maine, USA) by incubation in the dark at RT for 30 minutes. The reaction was stopped with 100 µL of 1M H_2SO_4 and read at wavelength of 450 nm.

KC4 analysis software (Bio-Tek Instruments, Inc. Winooski, Vermont, USA) generated a four parameter logistic curve from the standards that was used to calculate the concentrations of uPA and VWF from each ELISA. In some cases, for controls, uPA concentrations below the lowest standard were reported as below the lowest standard concentration on the standard curve. The absolute amounts of uPA and VWF for each culture were calculated by accounting for the total Day 14 culture media volume and

normalizing per $1x10^6$ Day 14 harvested MK cells. The absolute amount of uPA and VWF/$1x10^6$ Day 14 MK cells were compared to previously published data [81] for verification of acceptable quality cell cultures.

3.6 Droplet Digital PCR for rs4065 SNP T/C allele expression

TaqMan SNP Genotyping Assays (with primers included) were used for allelic quantification of rs4065 (*PLAU*, Chr10 (GRCh37) 75,676,464 C>T) (Catalogue # 4351379, Thermo Fisher Scientific Waltham, MA, USA) on QX200 Droplet Digital PCR system (Bio-Rad Laboratories, Inc., Hercules, CA, USA) at The Centre for Applied Genomics (TCAG, Sick Kids, Toronto, ON, Canada). Cycling conditions for the reaction were 95°C for 10 min, followed by 45 cycles of 94°C for 30 seconds and 57°C for 1 min, 98°C for 10 minutes and finally a 4°C hold on a Veriti thermal cycler (Life Technologies, Waltham, MA, USA). Data was analyzed using QuantaSoft v1.4 (Bio-Rad Laboratories, Inc., Hercules, CA, USA).

3.7 ChIP experimental procedure

Freshly harvested, Day 14 cultured MK and PB $CD66b^+$ granulocyte cell suspensions were fixed with 1 % formaldehyde (final) (Catalogue# F-5900, ACP, Montreal, QC, Canada) in Solution A [50 mM N-2-hydroxyethylpiperazine-N'-2-Ethanesulfonic Acid (HEPES)–KOH pH 7.4, 100 mM NaCl, 1 mM EDTA, 0.5 mM ethylene glycol tetraacetic acid (EGTA)] for 10 minutes at room temperature. The formaldehyde was quenched with 1/20 volume 2.5 M glycine before cell suspensions were centrifuged at 2,000 g for 5 mins, followed by 2x washes with PBS. MK and granulocyte

cell pellets were aliquoted accordingly at 2-3 x 10^6 and 1x10^7 cell numbers and stored at -80 °C.

ChIP experiments were conducted as previously described [169]. Antibodies used for ChIP experiments were purchased from EMD Millipore Darmstadt, Hesse, Germany and included: mouse anti-H3K27ac monoclonal (Catalogue # 05-1334), rabbit anti-H3K4me2 polyclonal (Catalogue # 07-030) and rabbit anti-H3K36me3 polyclonal (Catalogue # ab9050). Immunoprecipitations were performed with 100 µL of magnetic Protein-G dynabeads (Catalogue #10004D, LifeTechnologies, Waltham, MA, USA), which were washed and pre-blocked 3x with 0.5% BSA w/v in PBS (Catalogue #A8412, Sigma-Aldrich, St. Louis, Louisiana, USA) before incubating with 5 µg of antibody on a rotating apparatus for 4-18 hours at 4°C. Beads were washed 3 times with 0.5 % BSA in PBS to remove residual traces of unbound antibody before adding samples.

MK or granulocyte cell pellets were thawed and lysed on ice, and were serially treated with lysis buffers LB1 (50 mM of (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid HEPES)-KOH, pH 7.5, 140 mM NaCl, 1 mM EDTA, 10% Glycerol, 0.5% Igepal CA-630, 0.25% Triton X-100), LB2 (10 mM Tris–HCL pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA) and LB3 (10 mM Tris– HCl, pH 8, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% sodium deoxycholate, 0.5% N-lauroylsarcosine) as described below. The reagents sodium deoxycholate, N-lauroylsarcosine and Igepal CA-630, catalogue # D6750, L5125 and I8896 were purchased from Sigma Aldrich St. Louis, Missouri, USA. All lysis buffers (LB1, LB2 and LB3) were supplemented with (1:100 vol/vol) protease inhibitor cocktail mix (Catalogue #05056489001, Roche, Basel, Switzerland). Cells were lysed in

10 mL of LB1, incubated for 10 minutes on a rotor at 4° C followed by centrifugation at 2000g for 5 minutes at 4° C. The pelleted nuclei were then washed with 10 mL of LB2 for 5 minutes rotating at 4° C, followed by centrifugation at 2000g for 5 minutes. The nuclei pellet was then resuspended in 3 mL of LB3 and stored on ice.

A microtip Misonix Sonicator (Farmingdale, NY, USA) was used for sonication. The conditions for fixed $CD66b^+$ granulocytes were sonication for 3 minutes and 20s (10 seconds ON, 1 minute OFF for 20 cycles) whereas fixed MK were sonicated for 7 minutes (30S ON and 1 minute OFF for 14 cycles). DNA was fragmented to obtain a range between 200-400 bp. The total amount of energy output for sonication for fixed granulocytes and MK were between 6000-6500 J and 12000 to 12500 J respectively. After adding 300 µL of 10 % Triton X-100 to the sonicated LB3 lysate, cellular debris was removed by centrifuging at 20,000 g for 10 minutes at 4° C. 50 µL of the sonicate or whole cell extract (WCE) was aliquoted and stored at -20°C to be used as the input sample for ChIP analysis. The sonicated DNA supernatants were divided according to the number of ChIP experiments to be performed, where a minimum of 2 million cells were needed for each histone modification ChIP.

Sonicates were then incubated with antibody bound protein G dynabeads overnight on a rotor at 4° C. DNA-antibody bead complexes were captured using a magnetic stand, washed with 1 mL radio immunoprecipitation assay (RIPA) buffer (50 mM HEPES–KOH, pH 7.5; 500 mM LiCl; 1 mM EDTA; 1% NP-40 or Igepal CA-630; 0.7% Na–Deoxycholate) seven times and then once with 1 mL Tris buffered saline (TBS) (20 mM Tris-HCl pH 7.6; 150 mM NaCl). After centrifugation (950 g, 3 minutes, 4° C) to remove

residual TBS buffer, ChIP complexes were eluted with 200 µL ChIP elution buffer [(50 mM Tris–HCl pH 8, 10 mM EDTA, 1% sodium dodecyl sulfate (SDS)] and reverse crosslinked by incubation at 65°C overnight. 50 µl of WCE were also reverse crosslinked after adding 150 µL of ChIP elution buffer. After the incubation, supernatants from ChIP were collected using a magnetic stand and diluted with 200 µL of Tris EDTA pH 8 (TE). Next, all samples were incubated with 8 µL of 1 mg/mL ribonuclease A (RNaseA) (Catalogue #AM2271, Ambion Life Technologies, Waltham, MA, USA) for 30 minutes at 37° C followed by treatment with 4 µL of 15mg/mL Proteinase K (Catalogue #100005393, Invitrogen, Life Technologies, Waltham, MA, USA) for 2 hours at 55° C. DNA was enriched using phenol-chloroform-isoamyl alcohol extraction with the aid of 2 mL heavy phase lock gel tubes (PLG). The aqueous layer (composed of ChIP-DNA) was isolated during the extraction. Next, 400 µL of the aqueous layer was ethanol precipitated using 800 µL 100% ethanol, 200 mM NaCl and 20 µg of Glycoblue (Catalogue #AM9516, Ambion, Life Technologies, Waltham, MA, USA) to obtain ChIP and WCE DNA. The precipitated DNA was then dissolved in 44 µL of elution buffer (10 mM Tris-HCL, pH 8.5), incubated at RT for 5 minutes and stored at -20°C. ChIP DNA yields ranged between 200-500 ng and WCE DNA was aliquoted into 220 ng quantities until further use.

3.8 Library preparation of ChIP-DNA

The NEBNext DNA library preparation kit (Catalogue # E6040L, New England Biolabs, Ipswich, MA, USA) was used to prepare libraries of ChIP purified DNA for next generation sequencing (NGS) according to manufacturer's protocol. Briefly, DNA was blunt end repaired, dA-tailed and ligated with Illumina adapters using the NEBNext kit

reagents as follows: blunt end repair of DNA was performed for 30 minutes at RT followed by Zymo DNA purification using Clean and Concentrator DNA kit (Catalogue #D4014, Zymo Research Corporation, Irvine, CA, USA). ChIP and WCE DNA then underwent dA tailing for 30 minutes at 37°C, followed by another round of Zymo DNA purification. Illumina adapters with uracil hairpins were added at a final concentration of 6.67 nM and incubated at RT for 15 minutes, followed by treatment with USER enzyme to cleave the hairpins at RT for 15 minutes. DNA then underwent a final round of Zymo DNA purification, before being ready for ChIP-PCR.

The ChIP and WCE DNA libraries were PCR amplified for 16 cycles with universal and barcoded primers that allowed DNA libraries to be pooled prior to next generation sequencing (Catalogue # E7335L, NEB, Ipswich, MA, USA). PCR amplified ChIP and WCE DNA libraries were purified using QIAQuick PCR purification kit (Catalogue # 20810, QIAGEN, Hilden, North Rhine-Westphalia, Germany), and eluted in 30 µL of Tris EDTA pH 8.0 buffer. DNA libraries were then size selected between 200-350 bp after separation on 2% dye free Agarose Pippin Prep Gel cassettes (Catalogue # CDF2010, Sage Sciences, Beverly, MA, USA) according to manufacturer's protocol.

3.9 ChIP-qPCR validation and NGS of ChIP libraries

ChIP libraries were PCR re-amplified for 25 cycles according to the following conditions (98°C for 2 minutes, cycling 98°C for 30 seconds, 65°C for 30 s, 72°C for 1 minute followed by extension 72°C for 5 mins). Universal primers designed against the Illumina adapters used were Primer 1.3 (AATGATACGGCGACCACCGAGA) and Primer 2.4 (CAAGCAGAAGACGGCATACGAGAT). Samples were then purified using

MinElute PCR purification Kit (Catalogue # 28006, QIAGEN, Hilden, North Rhine-Westphalia, Germany), quantified using Nanodrop 2000 spectrophotometer (Thermo Fisher Scientific Waltham, MA, USA) and stored at -20°C.

A preliminary check of histone marker enrichment in ChIP DNA libraries was conducted using quantitative PCR (qPCR). The reaction volumes of 11 µL were comprised of 5 µL of 0.5 ng/µL ChIP-DNA, 5.5 µL SYBR Select Green Master Mix (Catalogue # 4367659, Life Technologies, Waltham, MA, USA), 0.2 µL of 25 µM for each primer pair and 0.3 µL of nuclease free H_2O. The qPCR reactions were run in triplicates in both ChIP and WCE DNA using forward (F) and reverse (R) primers against regions of a positive control gene *GAPDH* for each active histone enrichment as follows: H3K27Ac (F: GCAAGGAGAGCTCAAGGTCA, R: TGGGGACTAGGGGAAGGAG), H3K36me3 (F: TGCTTGCTTCGAGAACCATT, R: GAGCGAGGAGAGGAAAAGG), H3K4me2 (F: GAACCAGCACCGATCACC, R: CAGCCGCCTGGTTCAACT). A region with no enrichment termed "gene desert", was used as a negative control with primers as follows F: TTCTCCCTCTGCTTTGCACT, R: ACTGCACCATGACCTGGATT. The reactions were performed in Via7 Real Time qPCR System (LifeTechnologies, Waltham, MA, USA), according to default conditions that amplified 80-120 bp amplicons: 95°C for 10 minutes, followed by 40 cycles of 95°C for 15 seconds, 60°C for 1 minute. Fold enrichment was calculated using the ΔΔCt method, where Ct values of ChIP DNA were normalized to negative control region (gene desert) Ct, followed by the corresponding WCE DNA Ct [176]. A minimum threshold of 20-fold enrichment was used to determine an acceptable quality of ChIP samples for NGS.

ChIP and WCE DNA underwent further quality control checks at the Donnelley Sequencing Centre (University of Toronto, Toronto, ON, Canada) using a Bioanalyzer to visualize the DNA and KAPA Library Quantification Kit (KAPA Biosystems, Wilmington, MA, USA) to quantify the libraries. Libraries were sequenced on an ultra high throughput sequencing system, Illumina HiSeq 2500 (Illumina, San Diego, CA, USA) according to the sequencing facility protocol for 75 cycles to obtain approximately 20 million 100 bp single end reads per sample.

3.10 ChIP-seq pipeline:

NGS results were obtained in FASTQ file format that undergo a fast quality control to negate any technical errors during sequencing or library preparation. The Wilson lab trimmed 100 bp read files according to the base calling quality [177] which were then aligned to the hg19 (GRCh37) reference genome using Burrows Wheel Aligner (BWA) [178]. The output BED files were then normalized for read depth and converted to bedGraph using Homer [179], which takes each sequence and extends it by the empirically determined fragment length (approximately 200 bp). Homer normalizes the hg19 aligned BED files by ensuring the total number of mapped sequences is normalized to 10 million reads. The normalized bedGraph profiles were then uploaded and visualized on the UCSC genome browser. Homer was also used to pool replicates from within the same group for QPD or control $CD41^+$ MK and $CD66b^+$ granulocyte ChIP-seq profiles H3K27Ac (n=3 of each type of donor), H3K36me3 (n=3 of each type of donor) and H3K4me2 (QPD n=3, CTL n=2 of each type of donor).

3.11 ChIP-seq quality control

ChIP-seq experiments were evaluated by quality control guidelines set by the Encyclopedia of DNA Elements (ENCODE) consortia [180]. These metrics were used to describe the quality of MK and granulocyte ChIP-seq data from QPD and control subjects. PBC (PCR bottleneck coefficient) and NRF (non redundant fraction of mapped reads) are measures that estimate library complexity, where ChIP-seq libraries with PBC < 0.5 and NRF < 0.8 may indicate libraries with poor quality [180]. In addition, cross correlation analysis by ENCODE estimates ChIP-seq quality by measuring the signal to noise ratio independent of peak calling [180]. The analysis suggests that ChIP-seq libraries with normalized and relative strand coefficients (NSC and RSC) of < 1.05 and < 0.8 respectively may be of low quality [180]. All ChIP-seq experiments for QPD and control granulocytes and MK passed the aforementioned ENCODE quality control metrics summarized in Supplementary Tables 1 and 2 respectively. After the quality control checks were met, biological replicates of QPD or control ChIP-seq libraries for each cell type were merged and used for downstream analysis.

3.12 Peak calling and differential analysis

MACS2 [181] was used to call peaks in QPD and control ChIP-seq hg19 aligned sequences (option broad enabled for H3K36me3 and genome size defined at 3.1e9 and q-value 0.05). Peaks were called in each set of QPD and control samples of each histone mark evaluated. The peak summit bed files produced from MACS2 were then uploaded onto the UCSC Genome Browser for visualization [182]. Differential analysis on QPD and control

MK and granulocytes for H3K27Ac, H3K36me3 and H3K4me2 was conducted using diffReps [183]. The diffReps tool uses a sliding window approach to identify all genomic regions that have differential binding for a histone modification between ChIP-seq profiles. A negative binomial distribution is used to determine variation in and between groups for discrete counting data. diffReps initially filtered out regions with read counts lower than the geometric mean +2 standard deviations followed by selecting differential sites that have a p-value < 0.0001 and false discovery rate (FDR) <10%. A sliding window of 1000 bp was used with a moving step size of 100 bp. The –nsd=20 (BROAD) option was used for H3K36me3 and –nsd=2 (SHARP) option was used for H3K27Ac and H3K4me2 differential analysis.

The read counting tool, featureCounts [184] was used to count reads across Diffreps differential regions for H3K27Ac, H3K36me3 and H3K4me2 of MK and granulocytes. In order to account for the basal 1.5-fold difference between QPD and control subjects, read count matrices for control subject data were modified by a factor of 1.5 for regions found only in the region of *PLAU* duplication for all histone modifications evaluated. The modified read count matrices for controls together with QPD read count matrices were used for differential analysis of histone markers, H3K27Ac, H3K36me3 and H3K4me2 using DESeq2 [185]. Differential analysis using DESeq2 scales all reads by the mean normalized counts across all replicates. Differential regions with an adjusted p-value or FDR < 0.1 are considered to be statistically significant [185].

3.13 QPD breakpoint analysis

An in silico QPD chromosome 10 (QPD chr10) model containing the duplicated *PLAU* locus was created based on previous structural mapping studies and cloned break point region [132]. The ~78 kb region of *PLAU* that is duplicated and the QPD breakpoint were inserted into FASTA sequence of chr10. The uniquely mapped reads from all QPD active histone modification profiles were aligned to a 10 kb region with QPD breakpoint at the midpoint in the QPD chr10 model using BWA default settings [178].

3.14 DNA methylation analysis using Infinium HumanMethylation450K BeadChip Array

DNA for methylation analysis was extracted from cells using AllPrep DNA/RNA Mini Kit, eluting with AllPrep DNA spin columns (Catalogue# 80204, QIAGEN, Hilden, North Rhine-Westphalia, Germany). DNA from QPD and control samples were bisulfite converted using EZ DNA Methylation Kit (Catalogue # D5001, Zymo Research Corporation, Irvine, CA, USA) and hybridized to an Infinium Human Methylation450K BeadChip with scanning on Illumina's iScan microarray scanner (Illumina, San Diego, CA, USA). The bisulfite conversion and hybridization steps were carried out the Genetic and Epidemiology Laboratory at Thrombosis & Atherosclerosis Research Institute (TaARI) in Hamilton, ON, Canada. The hybridized DNA underwent nucleotide extension while conjugated to a fluorophore that has the capability to differentiate between CpG island cytosine residues that are methylated and un-methylated. The microarray scanner was used to detect the methylation signal, determined by the ratio of fluorescent signals from the

methylated versus un-methylated sites. The Infinium HumanMethylation450K assay was developed using a combination of Infinium I and II assays, with specific probes to target either methylated or un-methylated cytosine residues [154]. The incorporation of the Infinium II assay technology gauges the methylation status of a CpG locus independently of adjacent CpG sites as the Infinium I technology was based on the assumption that CpG methylation status within 50 bp remained the same [154]. The microarray coverage includes 96% of CpG islands and >485,000 individual methylation sites of the human genome [154]. The differential methylation analysis was performed between QPD and control MK, granulocytes and QPD MK vs QPD granulocytes (n=3 for each group) by Dr. Sadikovic's lab (University of Western Ontario, London, ON, Canada) using Partek Genomics Suite (St. Louis, Missouri, USA). QPD and control individuals from each group were matched for age and sex in order to negate any possible biases in the methylation analysis. The differential beta values (methylation level) between the groups was filtered using p-values < 0.01, F-value (signal to noise ratio) >10 and estimates (difference in beta methylation level between two groups) > 15%. *PLAU*, *VCL* and *CAMK2G* genes were visualized using Genomic Browser Wizard (Partek) and methylation levels compared between groups. Furthermore, only differential regions with the aforementioned statistical cut offs and >4 methylation probes were considered to be statistically significant.

4. RESULTS

4.1 Quality control analysis of Day 14 CD41$^+$ MK

Quality control analysis was conducted on Day 14 harvested MK cultured from CD34$^+$ HSCs to ensure that appropriately differentiated cells had been obtained for downstream analysis. Quality was assessed based on the % Day 14 MK expressing CD41 and Day 14 MK culture media levels of the α granule proteins uPA and VWF (Table 1).

CD41 expression of Day 14 MK cell cultures ranged between 80-90% for controls and 80-91% for QPD cultures, and these values fell within the range of previously reported data for MK grown in culture [81]. QPD Day 14 MK culture media samples composed of much higher levels of uPA than control samples (ranges, QPD: 12,000-16,000 pg/10^6 cells; controls: 50-87 pg/10^6 cells). The VWF levels however, were similar between QPD and control Day 14 MK culture media samples (ranges, QPD: 35-116 ng/10^6 cells; controls: 35-112 pg/10^6 cells). The uPA and VWF levels were within the range of previously published data [81].

	Day 14 Megakaryocytes	Day 14 Culture Media	
Sample	**CD41 %**	**uPA pg/10^6 cells**	**VWF ng/10^6 cells**
QPD Day 14 (n=4)	80-91	12000-16000	35-116
Control Day 14 (n=4)	80-90	50-87	35-112
Previously Published Ranges	QPD: 45-76 Control: 61-84	QPD: 7000-26000 Control: 40-440	QPD/Control: 21-145

Table 1. Differentiation data for QPD and control MK cultures, harvested on Day 14.

The % of cells that were positive for CD41 is shown, along with the levels of uPA and VWF in Day 14 culture media, evaluated by ELISA and expressed as the amount/10^6 Day 14 cultured MK. Previously published ranges were obtained from [123].

4.2 Infinium HumanMethylation450K BeadChip Array analysis of the *PLAU* region

The methylation status of *PLAU* promoter CpG island has not been previously evaluated in QPD. *PLAU* promoter methylation status was determined using Infinium HumanMethylation450K BeadChip Array in QPD and control MK and granulocytes. The methylation profile at the *PLAU* promoter CpG island was found to be hypo-methylated in both QPD and control MK and granulocytes (Figure 5). At the same time, the promoter region at *C10orf55* was also found to be hypo-methylated. Moreover, differential analysis of the promoter CpG islands of the flanking genes *CAMK2G* and *VCL* revealed a similar pattern of hypo-methylation in both cell types for QPD and control samples (Figure 6). Taken together, there was similar hypo-methylation detected for QPD and control subjects, for both MK and granulocytes, at all promoter CpG islands in the *CAMK2G-PLAU-VCL* region. In addition, whole genome differential methylation status analysis revealed significant differential regions according to the statistical filters described (Methods, 3.16) between QPD and control subjects in both granulocytes and MK (Tables 2 and 3).

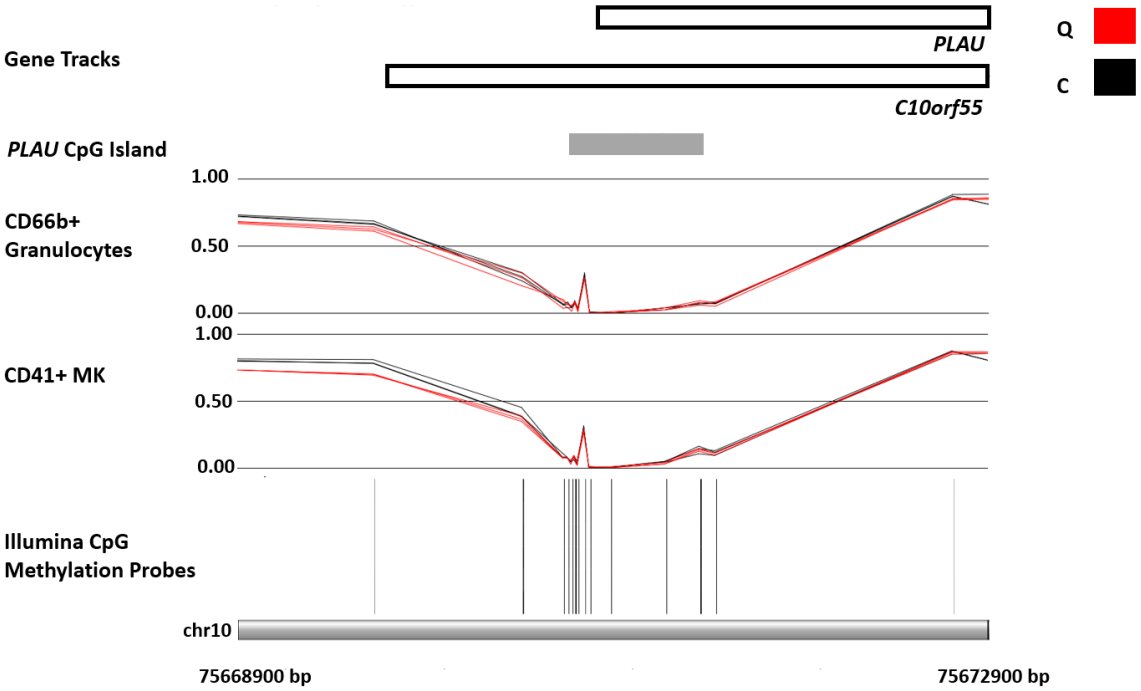

Figure 5. Diagrammatic representation of the methylation status of *PLAU* CpG island and neighboring methylation sites for QPD and control MK and granulocytes, evaluated using the Infinium HumanMethylation450K BeadChip Array. Granulocytes and MK from subjects (n=3 in each group; individual profiles for each cell type are shown) were assessed as outlined in Methods. No significant differences detected in the methylation status between QPD and control groups for either cell types at the *PLAU* CpG promoter island. The y-axis shows the β methylation level (0= un-methylated and 1=100% hyper-methylated), and the x-axis shows a 4000 bp genomic region that includes the *PLAU* CpG island (grey bar).

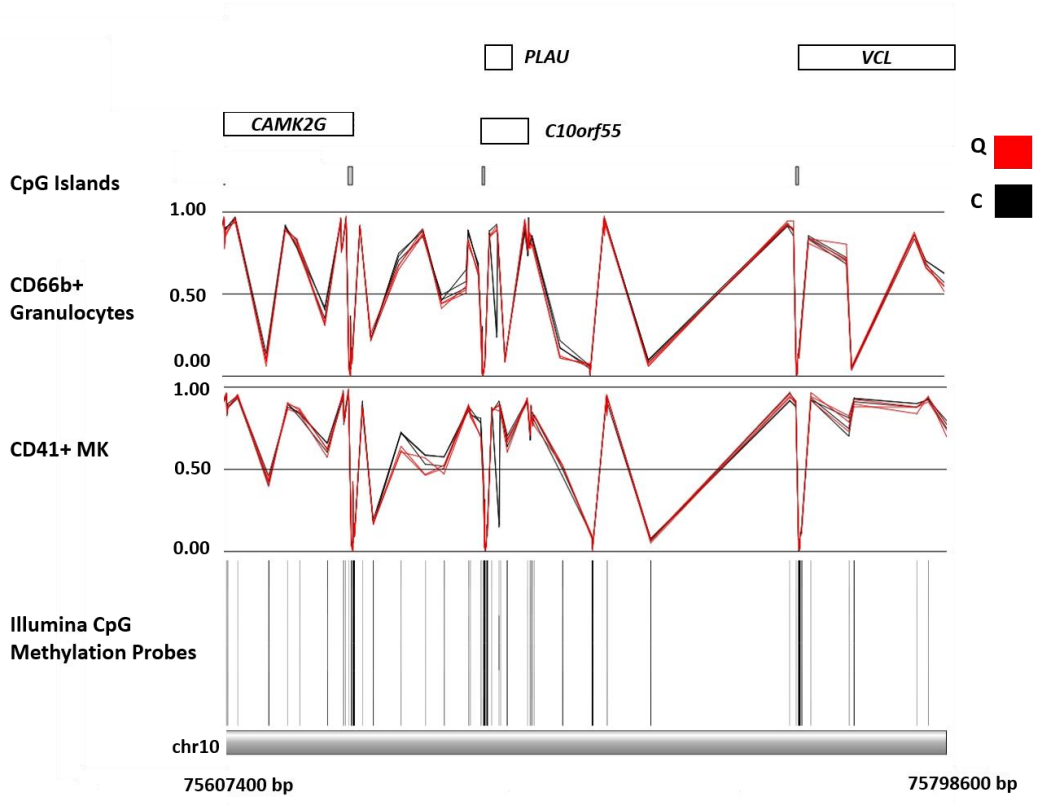

Figure 6. Diagrammatic representation of the methylation status for the *CAMK2G-PLAU-VCL* region for QPD and control MK and granulocytes, evaluated using the Infinium HumanMethylation450K BeadChip Array. Methylation profiles compare QPD (Q) and control (C) data for granulocytes and MK (individual profiles for 3 samples of each QPD and control cell type are shown). There were no significant differences detected in the methylation status between QPD and control groups for either cell types at the promoter CpG islands of *PLAU* or the flanking genes *CAMK2G* and *VCL*. The y-axis shows the β methylation level (0= un-methylated, 1=100% hyper-methylated), and the x-axis shows a 191,200 bp region with CpG islands (grey boxes).

Chr	Start	End	F value	Nearest gene	Estimate
chr11	289764	290457	579.007	ATHL1 (+)	-0.222022
chr10	76996270	76996543	258.806	COMTD1 (-)	-0.355734
chr1	2.07E+08	2.07E+08	86.4489	EIF2D (-)	-0.168677
chr12	1.33E+08	1.33E+08	82.9983	LRCOL1 (-)	-0.174614
chr1	26527860	26527938	78.8479	CATSPER4 (+)	-0.312172
chr11	18477143	18477763	73.636	LDHAL6A (+)	-0.190641
chr2	8596898	8597933	50.2424	LOC101929567 (-)	-0.164379
chr11	11642001	11642370	46.8553	GALNT18 (-)	-0.161868
chr16	85362953	85363219	43.4342	MIR5093 (-)	-0.174626
chr2	1.71E+08	1.71E+08	34.5128	KLHL23 (+)	-0.186079
chr6	887694	888046	32.8198	LOC101927691 (+)	-0.170728
chr2	20870802	20871411	31.0244	GDF7 (+)	-0.30325
chr16	1583800	1584526	25.481	TMEM204 (+)	-0.206201
chr12	1.32E+08	1.32E+08	25.3674	SFSWAP (+)	-0.174614
chr1	17026248	17026827	24.391	ESPNP (-)	-0.155155
chr17	7253330	7253730	22.9546	ACAP1 (+)	-0.267799
chr15	1.01E+08	1.01E+08	22.009	PRKXP1 (-)	-0.44891
chr8	6418968	6419580	21.0364	ANGPT2 (-)	-0.17767
chr2	20870077	20870372	20.7539	GDF7 (+)	-0.2738
chr2	2.19E+08	2.19E+08	20.4022	RUFY4 (+)	-0.177087
chr5	1594011	1594873	20.125	SDHAP3 (-)	-0.233079
chr10	99734071	99735212	20.0875	CRTAC1 (-)	-0.191739

Table 2. Regions identified to show differences in methylation from genome wide methylation analysis of QPD and control granulocytes. The table summarizes the chromosome positions, F-values (signal to noise ratio) >10, the nearest gene for the region showing differential methylation for QPD and the estimates (difference in beta methylation level between QPD and control groups) >15% for granulocytes. The (+) indicates greater methylation level in QPD compared to the control group whereas (-) indicates the reverse.

Chr	Start	End	F value	Gene	estimate
chr10	93736	94326	205.367	TUBB8 (-)	-0.202402
chr7	62859483	62859567	114.975	LOC100287834 (-)	0.254321
chrX	117628661	117629173	77.148	DOCK11 (+)	-0.20515
chr1	26527860	26527938	76.1445	CATSPER4 (+)	-0.293331
chr16	85362953	85363219	38.3965	MIR5093 (-)	-0.16313
chr6	887694	888046	32.248	LOC101927691 (+)	-0.165237
chr5	1594272	1594873	29.7677	SDHAP3 (-)	-0.291022
chr10	99734071	99735212	19.2447	CRTAC1 (-)	-0.183632
chr7	5270918	5272074	18.5614	WIPI2 (+)	0.155472
chr2	20870802	20871411	16.8086	GDF7 (+)	-0.222888
chr6	31803196	31805339	16.549	SNORD52 (+)	-0.156069
chr19	52954452	52954732	13.5838	ZNF534 (+)	-0.158534
chr6	29911285	29911846	12.3639	HLA-A ($^{+}$)	0.174343
chr8	21916246	21917025	12.3057	DMTN (+)	-0.151386
chr11	3534627	3535109	11.9355	TRPC2 (+)	-0.178359
chr16	8995886	8995994	11.8992	USP7 (-)	0.232909
chr21	47581032	47581568	11.5785	SPATC1L (-)	-0.197149
chr2	233272517	233273005	11.4222	ALPPL2 (+)	-0.179735
chr2	20870077	20870372	11.2513	GDF7 (+)	-0.197587
chr2	173539081	173539631	10.6696	RAPGEF4-AS1 (-)	-0.15895

Table 3. Regions identified to show differences in methylation from genome wide methylation analysis of QPD and control MK. The table summarizes the chromosome positions, F-values (signal to noise ratio) >10, the nearest gene for the region showing differential methylation for QPD and the estimates (difference in beta methylation level between QPD and control groups) >15% for MK. The (+) indicates greater methylation level in QPD compared to the control group whereas (-) indicates the reverse.

4.3 CD66b^{+} granulocytes ChIP-seq for active histone modifications

We first established histone modification profiles using ChIP-seq in QPD unaffected lineage granulocytes, where samples from QPD and control subjects were evaluated for the active histone modifications H3K27Ac (active enhancers), H3K36me3 (active transcription) and H3K4me2 (active promoters and enhancers). Figure 7 shows the histone modification enrichment profiles merged for each group (QPD and control), normalized by a factor of $10x10^6$ reads and visualized for the *CAMK2G-PLAU-VCL* landscape on the UCSC genome browser (Figure 7). The figure shows the UCSC genes, annotated with open reading frames, TSS, CpG islands and the probes used in the Infinium HumanMethylation450K BeadChip Array, in addition to DESeq2 differential regions, defined by either promoter (PR), intergenic region (IGR) or gene body (GB) for the proximal gene of interest. The figure shows reads at the region of duplication for control profiles after adjusting for the extra copy of *PLAU* in QPD. The DESeq2 analysis for QPD and control granulocytes indicated that at *PLAU* there were a few differential regions (FDR < 0.1) detected for H3K27Ac *PLAU* (designated IGR1 and IGR2) and H3K36me3 *PLAU* (designed GB) (Figure 7). Boxplots of the normalized read counts for the QPD and control granulocyte ChIP-seq regions (Figure 8) indicated that the *PLAU* IGR1 had a -1.6 fold change (greater enrichment in control vs QPD), whereas *PLAU* IGR2 and *PLAU* GB had +1.5 and +1.8 fold change (greater enrichment in QPD versus control) respectively, which were within the realm of having one extra copy of the *PLAU* gene. There were no significant QPD and control differential regions for any active histone modifications evaluated at *CAMK2G* and *VCL*.

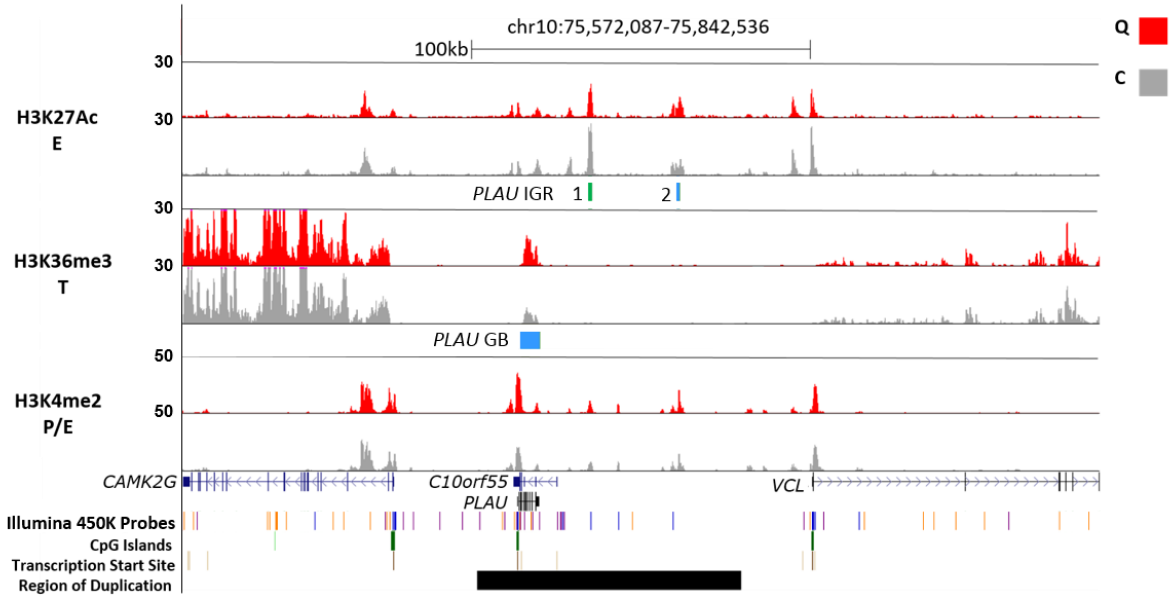

Figure 7. ChIP-seq profiles for active histone modifications in QPD and control granulocytes. Genome browser snapshot of the *CAMK2G-PLAU-VCL* locus shows ChIP-seq profiles from merged biological replicates obtained for QPD (Q) and control (C) granulocyte H3K27Ac (E denotes active enhancers; n=3), H3K36me3 (T denotes active transcription; n=3) and H3K4me2 (P/E denotes active promoters and enhancers; QPD n=3, CTL n=2). The differential regions (FDR <0.1) for H3K27Ac and H3K36me3 are shown in aqua blue (greater enrichment in QPD vs. control), green (greater enrichment in control vs. QPD) bars and denoted intergenic region (IGR) and gene body (GB). The UCSC browser vertical viewing range (y-axis) and the genomic region of 270,450 bp (x-axis) are represented.

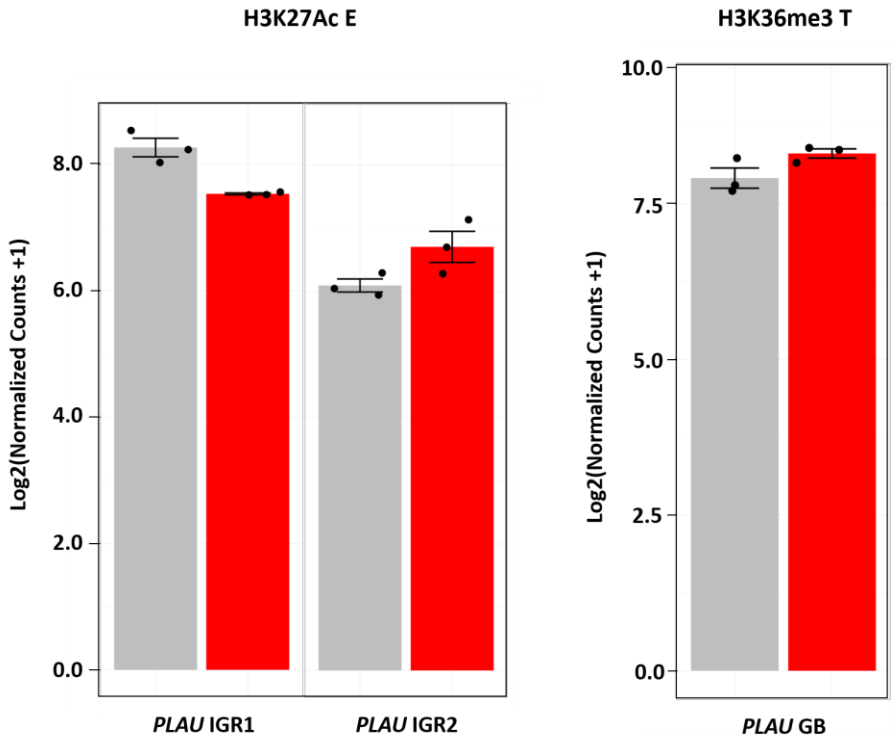

Figure 8. Normalized read counts for the differential regions identified through ChIP-seq analysis of QPD and control granulocytes. Bars compare average normalized read counts for the statistically significant differential regions (FDR < 0.1) between QPD (Q) and control (C) samples at the *CAMK2G-PLAU-VCL* locus, evaluated for H3K27Ac (E denotes active enhancers, n=3) and H3K36me3 (T denotes active transcription, n=3) in granulocytes. The differential region read counts were normalized to the mean read count across all differential regions in the genome as outlined in Methods, with control read counts modified to account for copy number variation. The y-axis represents log2 normalized counts with SEM error bars. The differential regions at *PLAU* are denoted by either intergenic region (IGR) or gene body (GB).

4.4 CD41^{+} MK ChIP-seq for active histone modifications

As of now, active histone modification enrichment at the *PLAU* locus in QPD MK have not been evaluated. The same histone modification profiles as granulocytes were evaluated for Day 14 cultured MK. QPD and control ChIP-seq profiles for H3K27Ac, H3K36me3 and H3K4me2 were processed as described above (with control MK reads also adjusted for *PLAU* copy number in QPD) and the DESeq2 differential regions visualized on the UCSC genome browser (Figure 9). The location of differential regions (FDR < 0.1) for the cultured MK were as follows; H3K27Ac: *PLAU* IGR1, H3K36me3: *PLAU* GB, *VCL* GB1, *VCL* GB2, H3K4me2: *PLAU* PR, *PLAU* IGR1, *PLAU* IGR2 and *PLAU* IGR3. The boxplots of normalized read counts for these regions are shown in Figure 10. H3K27Ac *PLAU* IGR1 was found to have a fold enrichment of +1.8 (Figure 10). H3K36me3 GB region for *PLAU* had a fold enrichment of +3, whereas the two GB regions of *VCL*, GB1 and GB2 had +3 and +2.6 fold changes respectively (Figure 10). H3K4me2 differential enrichment was detected at the promoter region, *PLAU* PR with a +1.7-fold enrichment whereas the H3K4me2 IGRs *PLAU* IGR1, *PLAU* IGR2 and *PLAU* IGR3 were found to have fold enrichment differences of +1.9, +1.3 and +1.4 respectively (Figure 10). The aforementioned fold differences could be due to the extra copy of *PLAU* in QPD. There were no significant QPD and control differential regions detected in *CAMK2G* for any of the active histone modifications assessed in MK.

A summary of the fold change and genomic regions of histone modification differential sites between QPD and control subjects for H3K27Ac, H3K36me3 and H3K4me2 in both MK and granulocytes are summarized in Table 4.

In order to assess cell specific changes, particularly in QPD, further differential analysis on within subject groups, by cell types to compare findings for QPD cultured MK and QPD granulocytes, and for control MK and control granulocytes, for all active histone modifications was conducted. Table 5 summarizes the fold enrichments observed for within each subject group for the region of *PLAU* duplication. For QPD subjects, a MK specific enhancer was found to be 9 fold enriched compared to $CD66b^{+}$ granulocytes (Table 5) and a granulocyte specific enhancer, upstream of *C10orf55* was found to be 12 fold enriched compared to MK (Table 5). Similar differences in enrichments were observed in approximately the same peak region for within control group by cell type analysis (Table 5).

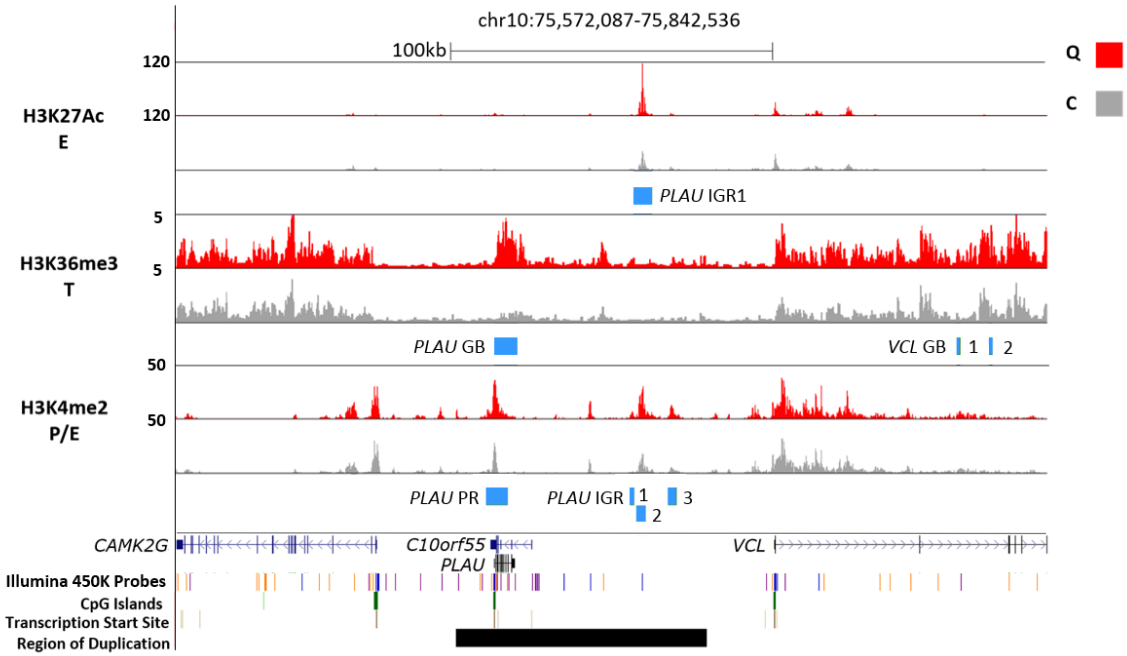

Figure 9. ChIP-seq profiles for active histone modifications in QPD and control MK. Genome Browser snapshot of the *CAMK2G-PLAU-VCL* locus shows ChIP-seq profiles from merged biological replicates obtained for QPD (Q) and control (C) MK H3K27Ac (E denotes active enhancers; n=3), H3K36me3 (T denotes active transcription; n=3) and H3K4me2 (P/E denotes active promoters and enhancers; QPD n=3, CTL n=2). The differential regions (FDR < 0.1) for the active histone modifications are shown in aqua blue (greater enrichment in QPD vs. control) bars denoted by intergenic region (IGR), gene body (GB) and promoter (PR). The UCSC browser vertical viewing range (y-axis) and the genomic region of 270,450 bp (x-axis) are represented.

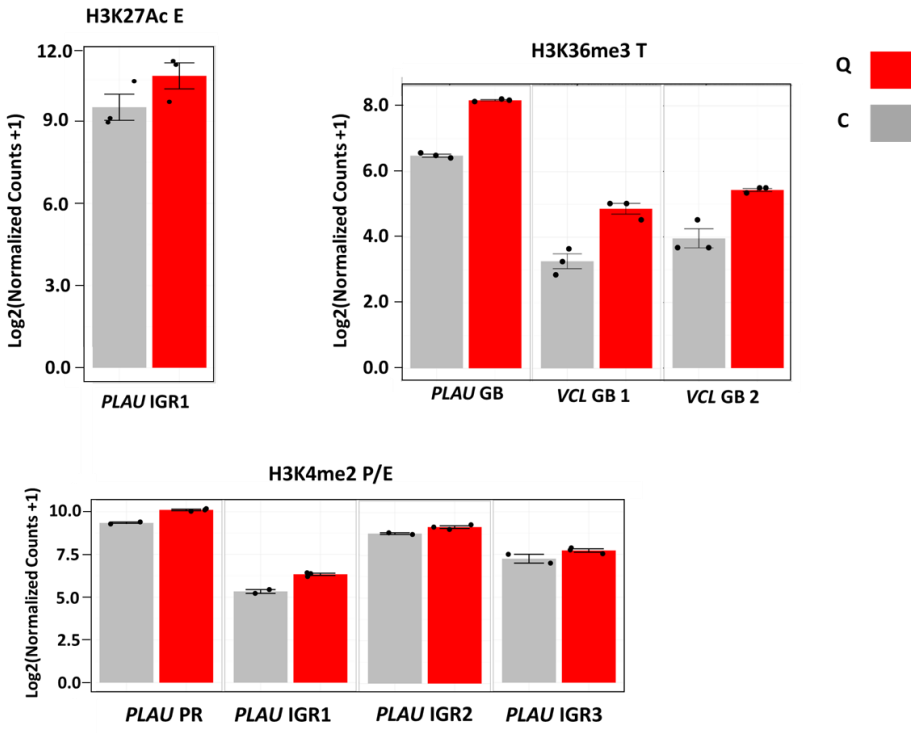

Figure 10. Normalized read counts for ChIP-seq differential regions in QPD and control MK. Bars compare average normalized read counts for the statistically significant (FDR < 0.1) differential regions between QPD (Q) and control (C) samples at the *CAMK2G-PLAU-VCL* locus, evaluated for H3K27Ac (E denotes active enhancers, n=3), H3K36me3 (T denotes active transcription, n=3) and H3K4me2 (P/E denotes active promoters and enhancers, QPD n=3, control n=2) in MK. The differential region read counts were normalized to the mean read count of all differential regions in the genome, as outlined in Methods, with control read counts modified for copy number variation. The y-axis represents log2 normalized counts with SEM error bars. The differential regions at *PLAU* and *VCL* are denoted by intergenic region (IGR), promoter (PR) and gene body (GB).

Histone Modifications/Cell Types	Region	Chr	Start	End	Fold Change
H3K27Ac E					
Granulocytes	PLAU IGR1(G)	chr10	75691901	75692900	1.6(-)
	PLAU IGR2(G)	chr10	75717801	75718900	1.5(+)
MK	PLAU IGR1(M)	chr10	75715101	75718700	1.8(+)
H3K36me3 T					
Granulocytes	PLAU GB(G)	chr10	75672001	75677500	1.3(+)
MK	PLAUGB (M)	chr10	75671001	75678100	3(+)
	VCL GB 1 (M)	chr10	75814501	75815600	3(+)
	VCL GB 2 (M)	chr10	75824501	75825500	2.6(+)
H3K4me2 P/E					
MK	PLAU PR (M)	chr10	75668401	75675200	1.7(+)
	PLAU IGR1(M)	chr10	75712901	75714400	1.9(+)
	PLAU IGR2(M)	chr10	75714901	75718000	1.3(+)
	PLAU IGR3(M)	chr10	75724901	75727400	1.4(+)

Table 4. Summary of the fold enrichment data for comparisons of QPD and control samples for differential regions with active histone modifications. QPD and control differential regions in granulocytes (G) and MK are shown, along with their respective genomic locations and fold enrichment values for regions in the *CAMK2G-PLAU-VCL* locus analyzed for H3K27Ac (E denotes active enhancers), H3K36me3 (T denotes active transcription) and H3K4me2 (P/E denotes active promoters and enhancers). IGR, GB and PR represent intergenic region, gene body and promoter respectively. The (+) indicates greater enrichment in QPD vs control and (-) represents the reverse direction of enrichment.

Histone Modifications /Sample	Start	End	Fold Enrichment
H3K27Ac E	QPD MK vs QPD Gran		
Q	75691301	75693500	12(-)
	75715101	**75718700**	**9(+)**
	CTL MK vs CTL Gran		
C	75691401	75693500	25(-)
	75715801	75718100	4(+)
H3K36me3 T	QPD MK vs QPD Gran		
Q	75673101	75674400	2(-)
	75676501	75681000	4(+)
	CTL MK vs CTL Gran		
C	75673001	75675400	3(-)
	75676501	75677800	4(+)
H3K4me2 P/E	QPD MK vs QPD Gran		
Q	75667001	75672600	2(-)
	75672401	75675200	3(+)
	75675301	75678000	4(-)
	75685301	75687200	6(-)
	75690901	75693900	12(-)
	75714901	75718300	3(+)
	75717601	75720000	2(-)
	75724901	75727400	14(+)
	CTL MK vs Gran		
C	75667201	75672900	2(-)
	75675501	75677900	8(-)
	75685701	75687000	7(-)
	75691101	75693700	17(-)
	75715101	75718100	2(+)
	75717701	75720100	3(-)
	75725001	75727000	14(+)

Table 5. Fold enrichment analysis of differential regions on chromosome 10 in MK and granulocytes samples from QPD and control subjects. Differential regions between QPD (Q) and control (C) samples in MK and granulocytes are shown along with their respective genomic locations at the region of *PLAU* duplication and fold enrichment values for H3K27Ac (E denotes active enhancers), H3K36me3 (T denotes active transcription) and H3K4me2 (P/E denotes active promoters and enhancers). The (+) indicates greater enrichment in QPD vs control and (-) represents the reverse direction of enrichment. The MK specific enhancer is highlighted in blue.

4.5 QPD ChIP-seq active histone modification enrichment at QPD breakpoint

Figure 11 summarizes the ChIP-seq data at the QPD breakpoint on the QPD chr10 in silico model as described in methods. QPD ChIP-seq active histone modification profiles for H3K27Ac, H3K36me3 and H3K4me2 from all replicates of both MK and granulocytes were aligned at the QPD breakpoint region. There were no reads detected at the region for all QPD ChIP-seq profiles evaluated (Figure 11).

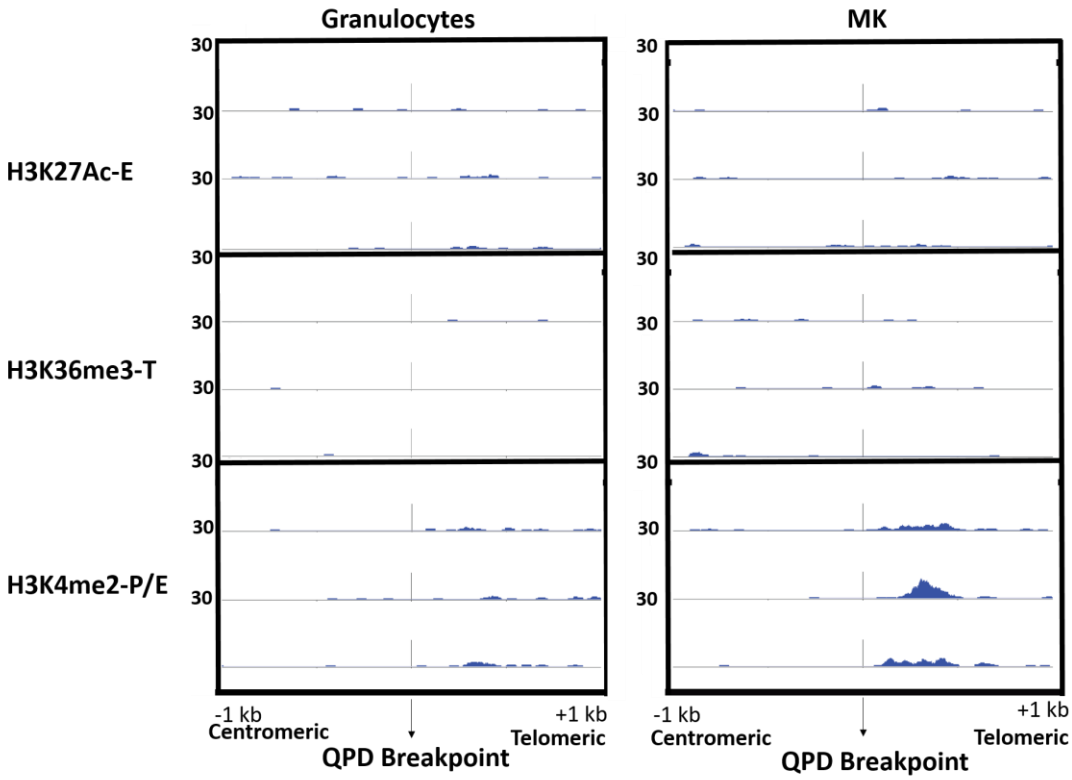

Figure 11. QPD ChIP-seq enrichment at the QPD breakpoint. QPD ChIP-seq profiles of all biological replicates from MK and granulocytes for histone modifications H3K27Ac (E denotes active enhancers, n=3), H3K36me3 (T denotes active transcription, n=3) and H3K4me2 (P/E denotes active promoters and enhancers, n=3) at the QPD breakpoint region. There were no QPD ChIP-seq active histone modification reads that aligned to the QPD breakpoint for either cell type evaluated. The UCSC browser vertical viewing range (y-axis) along with 1 kb upstream (Centromeric) and downstream (Telomeric) of the QPD breakpoint region (x-axis) are represented.

5. DISCUSSION

QPD is a unique autosomal dominant disorder that leads to delayed onset bleeding with hemostatic challenges in affected individuals [122]. A CNV mutation of the *PLAU* gene results in QPD affected individuals having two 78 kb *PLAU* regions, with a unique QPD breakpoint sequence [132]. QPD presents an inexplicable anomaly, where the presence of a gene duplication results in an aberrant increase of *PLAU* expression by MK and concurrent increases in uPA levels [81]. QPD affected individuals have a > 100 fold increase of *PLAU* mRNA and uPA protein levels in terminally differentiated cultured MK and peripheral blood platelets compared to healthy controls [81]. In order to understand the QPD over-expression anomaly, we proposed to evaluate epigenetic modifications at the *PLAU* gene locus in QPD individuals and unaffected controls for $CD41^+$ MK. At the same time, we evaluated granulocytes as an unaffected myeloid lineage based on observations that QPD granulocytes contain only modestly increased uPA, with the increase in the range expected for one extra copy of *PLAU* [80].

5.1 Influence of CpG promoter island methylation on *PLAU* expression

Methylation of CpG promoter islands is a crucial aspect of epigenetic regulation. DNA hyper methylation of CpG promoter islands is associated with transcriptional repression whereas hypo-methylation is associated with transcriptional activation [163]. I conducted my study using QPD and control samples that were age and gender matched, as these factors have been shown to heavily influence the status of methylation [186]. Based on the direct analysis of the methylation level of the *PLAU* CpG promoter island in

QPD and control MK and granulocytes, I concluded that *PLAU* is primarily hypo-methylated in both MK and granulocytes of QPD and control subjects (Figure 5). The hypo-methylation of the *PLAU* CpG promoter islands is in agreement with my initial hypothesis and implies that this gene is similarly accessible for TF binding in control and QPD subjects. Moreover, QPD does not bring about any aberrant changes in DNA methylation status of the nearby *CAMK2G* and *VCL* CpG promoter islands (Figure 6). My findings are in keeping with other recent observations. For example, visual inspection of methylation tracks from MethBase revealed a conserved state of hypo-methylation in *PLAU* CpG promoter island for the evaluated CD133 HSC and white blood cells [155]. Similarly, my findings are also in agreement with the trend of hypo-methylation observed in MethBase for the neighboring genes of *PLAU*: *CAMK2G* and *VCL* [155]. Moreover, our findings are similar to Bocker et al DNA methylation studies indicating that as HPSCs differentiate into granulocytes and monocytes, a large cohort of genes and a subset involved in myeloid differentiation are hypo-methylated [186]. Interestingly, the promoter regions of the genes that were identified to be hypo-methylated in the Bocker et al study, were also highly enriched with motifs for TFs involved in hematopoiesis including GATA-1,2 CEBPA and RUNX-1 [186].

My data indicates that changes to the promoter CpG island methylation do not influence the changes in *PLAU* expression in QPD individuals, and are in keeping with the finding and that the *PLAU* promotor is hypo-methylated across most published blood cell lines. The observation that QPD does not affect the *PLAU* promotor methylation level in either MK or granulocytes, implies that at the epigenetic level, altered *PLAU*

accessibility to TF binding in QPD could be associated with another epigenetic modulation, particularly changes in active histone modifications.

5.2 ChIP-seq active histone modification enrichment in QPD

Once I excluded aberrant differential methylation changes as a possible mechanism influencing *PLAU* expression in QPD, I turned my attention towards assessing active histone modification enrichment, namely alterations in H3K27Ac (a mark that separates active from poised enhancers), H3K36me3 (which is associated with active gene bodies) and H3K4me2 (which marks both active promoters and enhancers) [163, 164]. In my analyses, I considered that as a result of the CNV mutation, QPD affected individuals have a total of 3 *PLAU* copies in comparison to QPD unaffected individuals (controls) who have only 2 copies of *PLAU*. One would thus expect a 1.5 increase in *PLAU* up-regulation due to the presence of one extra *PLAU* copy on the QPD diseased chromosome. In order to account for copy number variation in QPD subjects, control reads were multiplied by 1.5 prior to differential analysis.

The differential analysis of QPD and control granulocytes for the active histone modifications indicated modest differential binding for H3K27Ac and H3K36me3 but not for H3K4me2. Furthermore, the changes in H3K27Ac and H3K36me3 marks indicate that QPD CD66b^{+} granulocytes have only modestly increased active enhancer and transcription activity. The differential enrichments in histone markers observed for granulocytes were within the range of QPD affected individuals having one extra copy of *PLAU* and in agreement of the modest increases in uPA levels observed previously [80]. The findings further reinforce the notion that granulocytes represent an unaffected QPD myeloid lineage

that would be an ideal reference lineage to gauge the epigenetic changes in QPD MK. There was an H3K27Ac enhancer peak (*PLAU* IGR1) was modestly less enriched (1.6 fold) in QPD compared to control $CD66b^+$ granulocytes (Figures 7,8), suggesting that the *PLAU* duplication causes a structural rearrangement that potentially disrupts the enrichment of that enhancer.

I then explored the active histone enrichment in MK, the key QPD affected lineage using similar methods. I observed increased, albeit moderate, fold enrichment differences in QPD compared to control MK for all three histone modifications tested, H3K27Ac, H3k36me3 and H3K4me2 (Figures 9,10). The slight increases observed for the fold enrichment differences across active histone modifications were within the range expected for having one extra copy of *PLAU*. We thus conclude that QPD *PLAU* duplication induces minimal changes to the chromatin accessibility through active histone enrichment to *PLAU* in QPD MK.

An analysis of QPD and control cell specific active histone enrichment differences between MK and granulocytes revealed several differential regions at the *PLAU* duplication region (Table 5). It is interesting to note that there are more H3K4me2 differential enrichments in granulocytes compared to MK. As the intrinsic uPA protein expression is higher in granulocytes compared to MK, the presence of a larger number of enhancers in the *PLAU* duplication region may be responsible for this phenomenon [80]. Moreover, the observance of granulocyte and MK specific H3K27Ac enhancers in QPD subjects (Table 5) could also explain the differences in uPA expression observed in these cell types.

QPD ChIP-seq profiles for all active histone modifications in both cell types were aligned to the QPD genome model, where little or no reads spanned the QPD breakpoint (Figure 11). This is a particularly important finding, as it indicates that the novel breakpoint sequence caused by the *PLAU* duplication in QPD does not introduce abnormal active histone enrichment. We can thus rule out the possibility that the unique QPD breakpoint creates regions with increased active histone enrichment.

Overall in accordance with our hypothesis, the QPD compared to control differential enrichment of active histone modifications, unaltered *PLAU* promoter CpG island methylation status and minimal enrichment over the QPD breakpoint in both MK and granulocytes, indicates a relatively stable epigenetic landscape at *PLAU*. The *PLAU* CNV mutation does not induce aberrant >100 fold changes in active histone modification enrichment as were observed with *PLAU* mRNA and uPA protein levels. We find that instead of aberrant enrichment, the rearrangement of active enhancer elements, particularly a MK specific H3K27Ac enhancer might underlie the marked *PLAU* overexpression by differentiated QPD MK as discussed in the next section (5.3).

5.3 Duplication of a marked enhancer in QPD MK

ChIP-seq analysis of H3K27Ac (active enhancers) reveals a highly expressed enhancer (9-fold difference) in QPD MK compared to QPD granulocytes (Figure 12, Table 5) that for the following discussion will be termed MK enhancer 1 (MKE1). A visual QPD duplication model, with two copies of ~78 kb *PLAU* region was created to better understand the spatial enrichment of H3K27Ac enhancers in QPD. In the visual QPD duplication model, MKE1 is shown to be placed 34 kb upstream of the *PLAU* copy that is proximal to

VCL (*PLAU*2). MKE1 could potentially upregulate the expression of the second copy of *PLAU*. The enhancer would be present in normal MK but due to the rearrangement caused by the CNV *PLAU* mutation, a copy of the enhancer is now proximal to the second copy of *PLAU* and thus has the potential to drive *PLAU* expression.

A deeper analysis on MKE1 and the regions it interacts with were studied from ENCODE published data on the UCSC Genome Browser (accession ID: GSE39495) [187]. MKE1 was found to be highly enriched in K562 and GM12878, lymphoblastoid and leukemia cell lines respectively [187].

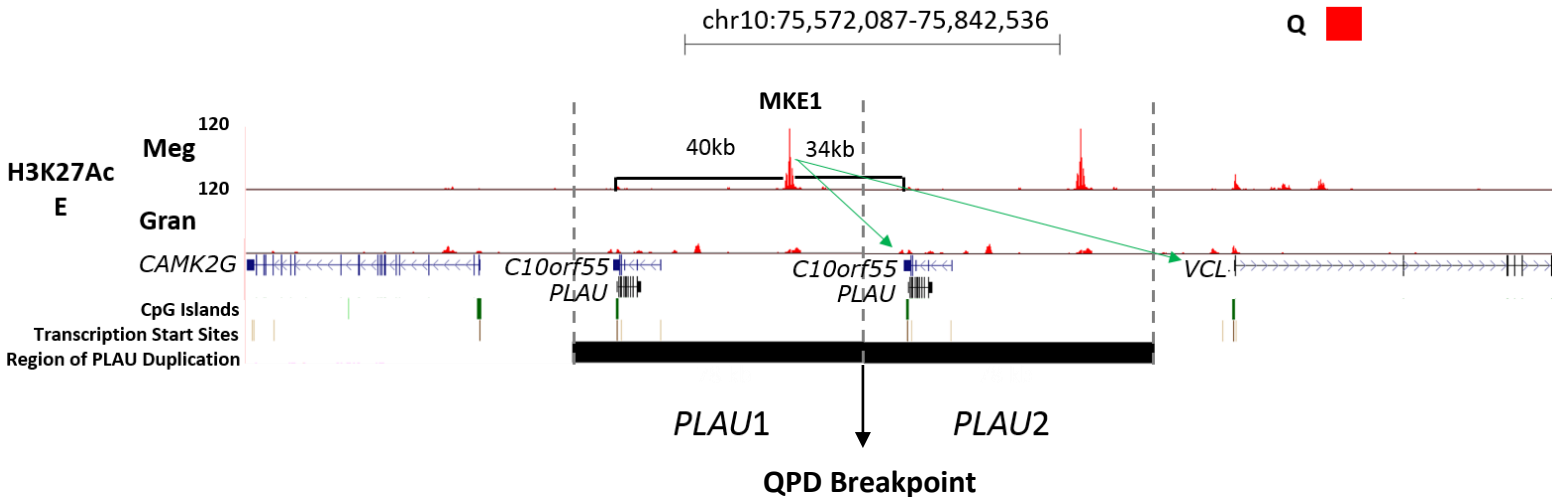

Figure 12. Model summarizing the location of H3K27Ac MKE1 relative to *PLAU* on the QPD disease chromosome. A QPD duplication model was constructed by replicating the ~78 kb *PLAU* region, resulting in two regions, *PLAU*1 and *PLAU*2. The position of the H3K27Ac (E denotes active enhancers) element, termed MKE1 that was enriched 9 fold in QPD compared to control MK, and was less evident in granulocytes is shown in the *CAMK2G-PLAU-VCL* landscape. The H3K27Ac enhancer that is highly enriched in QPD MK but not in granulocytes (9-fold) is located ~34 kb upstream of the second copy of *PLAU*. The UCSC browser vertical viewing range (y-axis) and genomic region of 270,450 bp (x-axis) are represented.

Chromatin Interaction Analysis by Paired-End Tag Sequencing (ChIA-PET) is a method for characterizing long range chromatin interactions for specific proteins of interest, e.g. TFs and RNA polymerase II [188]. ChIA-PET analysis generates two types of genome-wide datasets: 1) the binding sites defined by ChIP enrichment and 2) the interactions between two binding loci revealed by proximity ligation events [136, 188-190]. ChIA-PET data from ENCODE for POLR2A indicates that MKE1 binds RNA polymerase II and that this enhancer element also interacts with the *VCL* promoter in one replicate [187]. The findings suggest that MKE1 could serve as a potential enhancer that regulates *VCL* expression. In light of this finding, it is important to understand the significance of *VCL*, the neighboring gene downstream of *PLAU*. *VCL* encodes vinculin, a protein involved in linking cell-substrate/cell contacts to the actin cytoskeleton [191]. Murine models elucidated the significance of the *VCL* gene, where global knockout models (*VCL* -/-) resulted in embryonic lethality and heterozygous models (*VCL* $^{+}$/-) indicated changes to myocyte structure [192, 193]. Moreover, a MK/platelet lineage specific *VCL* knockout murine model (non fatal) indicated that *VCL* was not essential for maintaining platelet actin cytoskeleton [194].

As the QPD duplication moves one copy of MKE1 to upstream of the 2nd copy of *PLAU* on the disease chromosome, it may be that this element (that normally regulates the expression of *VCL*) becomes a positive up regulator of the expression of *PLAU* in QPD.

In order to understand the potential role of this putative MKE1 in upregulating the expression of *PLAU* in QPD it is crucial to understand the role of the interacting TFs. A visual inspection of ENCODE published data revealed the presence of CCCTC-binding factor (CTCF) and POLR2A (the largest subunit of RNA Polymerase II) binding sites at this MKE1 in both K562 and GM12878 cell lines [187, 195].

CTCF has been defined as a multi diverse TF with roles in transcriptional activation and repression [195]. An example of CTCF associated with active gene expression is observed when CTCF was shown to upregulate the activity of APP gene by binding to the GC promoter rich region [195]. CTCF has also been implicated in creating geometrical loops causing activation by bringing promoter and enhancer elements in close proximity [195]. Moreover, CTCF has been shown to associate with POLR2A in upregulating gene expression [195, 196]. The repositioning of MKE1 that has binding sites for both CTCF and POLR2A could play an important role in the pathogenesis of QPD. The CTCF could act as a chromatin remodeling agent that brings together the enhancer and the promoter region of the second *PLAU* copy in QPD affected individuals. Moreover, the presence of both CTCF and POLR2A binding sites could potentially result in CTCF-RNA polymerase II assembly that upregulates *PLAU* transcription.

A visual inspection of TF ChIP-seq profiles at the MKE1 region using the BloodChIP database (accession ID: GSE24674) was conducted [171, 197]. TF- binding peaks were observed for FLI1 and SCL, two TFs that are involved in regulating genes during MK lineage development [198, 199]. It is possible that the expression of these TF during MK differentiation, coupled with relocation of MKE1 to upstream of the 2nd copy

of *PLAU*, are the reasons that *PLAU* expression is markedly upregulated during the later stages of megakaryopoiesis in QPD.

To further explore this possibility, I evaluated current information on the H3K27Ac enhancer enrichment in the BLUEPRINT and Epigenome roadmap data on blood cells [172, 200]. Interestingly, *PLAU* is devoid of H3K27Ac enrichment near the promoter region in all blood cells compared to all non-blood cells in the Epigenome Roadmap [200] This lack of enhancer enrichment might account for the low expression of *PLAU* in normal blood cells. Interestingly, an H3K27Ac enhancer is also enriched at the same locus as MKE1 in other blood cells, particularly in mobilized $CD34^+$ HSC and cord blood derived cultured MK [171, 172]. This H3K27Ac enrichment is also complemented with a peak from H3K4me1, a histone mark that is also strongly associated with active enhancers [136, 172]. It is encouraging to see that these findings are in agreement with our characterization of MKE1. Moreover, the MKE1 enrichment in mobilized $CD34^+$ HSC implies that the enhancer is initially upregulated and this MKE1 expression level is maintained upon terminal myeloid differentiation in MK.

An alternative way that *PLAU* could be dysregulated in QPD, that I did not have the opportunity to explore, is through loss of an important repressor element. I evaluated the H3K27me3 ChIP-seq data for control cultured MK, generated by Tijssen et al that are found in the BLUEPRINT database [171, 172] and found that in control cells, there is a modest enrichment for H3K27me3 at the *PLAU* promoter region. H3K27me3 is deposited by the polycomb repressive complex 2 (PRC2) and is associated with repressed transcription [162]. In light of this evidence, it may be worthwhile in future studies to

determine if the QPD duplication disrupts repressor activity at the *PLAU* promoter region. The presence of a CTCF binding site proximal to *PLAU* could imply that normally CTCF has insulator activity towards *PLAU* but due to spatial rearrangement of the QPD duplication there is a disruption in the insulator effects of CTCF.

In consideration of the H3K27me3 enrichment at *PLAU* in control MK, it seems necessary to reevaluate this marker through comparisons of QPD and control cultured MK. More importantly, based on the aforementioned analysis of MKE1, further characterization of this enhancer could provide tremendous insights into the pathogenesis of QPD. A summary of future experiments to explore the role of MKE1 are discussed in more detail below.

5.4 Future directions

In the current study, the major finding is that of the MK specific enhancer termed MKE1. To further characterize the role of MKE1 in QPD, I discuss experimental approaches that assess MKE1 expression, linkage to *PLAU* promoter and TF binding enrichment in the subsequent sections.

5.4.1 Expression of MKE1 during megakaryopoiesis

A detailed analysis of the H3K27Ac MKE1 during the time course of megakaryopoiesis is warranted, given the evidence that the overexpression of *PLAU* during megakaryopoiesis in QPD is tied to MK differentiation [81]. It was previously shown that uPA mRNA and protein levels show little alteration during normal megakaryopoiesis in cultures whereas in QPD, *PLAU* mRNA and uPA protein levels are

fairly normal in progenitors (Day 0), then gradually increase (Day 7) and show the most drastically rise towards the terminal stages of megakaryopoiesis, when the levels are >100 fold higher than in control MK [81]. It would be interesting to explore potential correlations between the enrichment of MKE1 and increased *PLAU* expression during the time course of normal and QPD MK differentiation. Additionally, comparison of ChIP-seq data for H3K27Ac, using QPD and control Day 7 and Day 14 MK cultures, might be helpful to further evaluate temporal changes in MKE1. Significantly increased enrichment of MKE1 over this time period would provide evidence that the expression of this enhancer element is tied to MK differentiation.

In order to fully understand the dynamics of MKE1 enrichment during megakaryopoiesis, it would be ideal to study samples harvested at a larger number of time points. This could be challenging given the small number of cells obtained in MK cultures derived from PB $CD34^+$ HSCs (~ $1x10^6$ cells per isolation from 200 ml PB) and that the largest increase in MK expansion occurs during the last stages of culturing, a method ChIP-seq becomes unfeasible. However, some new methods have been developed to evaluate genes that use significantly lower cell numbers (5,000-50,000) such as assay for transposase-accessible chromatin (ATAC)-seq. ATAC-seq is emerging to be a useful technique that complements DNASE-seq, by employing a Tn5 transposase to cleave at regions with open chromatin conformation. ATAC-seq helps to identify active regions of chromatin such as enhancer regions identified by ChIP-seq [201].

The qPCR method used to verify ChIP enrichment of libraries can also be used on ATAC samples, a method called ATAC-qPCR. Fold enrichment is calculated by

normalizing to a region devoid of active gene expression (gene desert). It would be interesting to perform ATAC-qPCR to detect the enrichment of MKE1 during differentiation in MK cultures and to track changes between Days 0, 4, 7, 11 and 14. ATAC-qPCR analysis of MKE1 region, to determine the fold enrichment over a negative control region of DNA, would be helpful to quantify the enrichment of MKE1 over the progressive stages of normal and QPD MK differentiation.

At the same time, it is crucial to assess the mRNA expression of *PLAU* and *VCL* during the aforementioned time points. We would expect to find upward increases in enrichment of MKE1 that mirror the increases in *VCL* and *PLAU* mRNA expression during the QPD MK differentiation time course. If we do observe a correlation in expression trends between *PLAU* and MKE1 in QPD MK only, this would further implicate MKE1 as having a potent role in regulating the expression of *PLAU* in QPD.

5.4.2 ChIA-PET analysis of MKE1

As transcription is dependent on promoter bound RNA polymerase II, one favorable application of ChIA-PET is the immunoprecipitation of RNA polymerase II to classify physical interactions between transcriptional enhancers and target genes [188]. It would be interesting to perform ChIA-PET for RNA polymerase II using control and QPD cultured MK, to determine if MKE1 physically binds to the either *PLAU* or *VCL* promoter or in fact both in QPD; so far, there is no evidence that this element regulates *PLAU* in control samples. As enhancers are known to potentially function at large distances, in some instances Mbp from their target site [136], ChIA-PET genome wide findings can identify other potential enhancers that may interact with *PLAU* in QPD.

ChIA-PET analysis with CTCF could indicate the interactions of the TF binding to the enhancer with either promoter region of *VCL* or *PLAU*. If a linkage between TF bound MKE1 region and *PLAU* promoter is established in QPD and is not evident in control cultured MK, the findings would provide direct evidence that MKE1, along with the bound TF(s) of interest, are strongly associated with *PLAU* overexpression in QPD.

5.4.3 Potential TFs associated with *PLAU* dysregulation in QPD

My evaluation of the Tijssen et al dataset (via Blood ChIP database) for TFs involved in megakaryopoiesis revealed binding peaks at MKE1 for SCL and FLI1 [171, 197]. If future ChIA-PET experiments establish a physical interaction occurs between the MKE1 and the *PLAU* promoter in QPD MK, it would be worthwhile to explore the candidate TFs that interact with the MKE1 site. The top candidate TFs to evaluate, based on published data sets, would be: SCL, FLI1 and CTCF using TF-ChIP-seq. TF-ChIP-seq, is similar to histone modification ChIP-seq with the exception that it requires a larger number of cells and antibodies specific to TF of interest [169]. Once TFs that exhibit binding MKE1 are identified in control and QPD MK, their significance to the regulation of *PLAU* expression in QPD MK needs to be validated. Genomic editing of regions, particularly alterations or deletions can be induced by using CRISPR/Cas9 [202, 203]. The type II CRISPR system was shown to induce double-stranded breaks of DNA in vitro studies by the use of artificial guide RNA sequences [203]. The custom RNA sequences can easily be designed against any particular region of interest in the genome [204]. The type II CRISPR system has recently been adapted to human cells, in one instance being used to modify human-induced pluripotent stem cells in culture [202-206]. Based on these

published protocols, it would be interesting to use type II CRISPR/Cas9 to delete the MKE1 region in $CD34^+$ HSC and then allow these cells to differentiate in culture to $CD41^+$ MK. This might be feasible given that *VCL* is not essential for producing platelets, given the findings for the MK/platelet lineage specific *VCL* knockout mice [194]. We could then directly assess the effect of MKE1 by determining *PLAU* expression across the different time points of QPD and control MK cultures, with and without deletion of the MKE1 element. A reduction of *PLAU* expression in QPD but not in control cells lacking the MKE1 element would indicate MKE1 is involved in the upregulation of *PLAU* mRNA levels in QPD MK.

6. CONCLUSION

My thesis work findings indicate that QPD is not associated with aberrant changes in the methylation status of the *PLAU* promoter CpG island in MK or granulocytes. The hypo-methylated state of the *PLAU* CpG promoter island in these cells indicates that *PLAU* is ready to be actively transcribed at the same level in both QPD and control cell types. In analyzing active histone enrichment at and near *PLAU*, we found that there is a modest increase in enrichment of H3K27Ac (indicative of active enhancers), H3K36me3 (active transcription) and H3K4me2 (associated with active promoters and enhancers) in both evaluated cell types. However, the fold enrichment differences observed are within the realm of what is expected for having one extra copy of *PLAU* in QPD. We thus conclude that the overall epigenetic landscape of *PLAU* remains relatively stable in both QPD MK and granulocytes.

The most interesting finding of my thesis work is the identification of a H3K27Ac MKE1 element that is active in MK but not granulocytes, and that the *PLAU* duplication repositions one copy of this enhancer so that it is proximal to the second copy of *PLAU* on the QPD disease chromosome. The new position of this MK specific highly expressed MKE1, in combination with the production of TF that upregulate genes during megakaryopoiesis, may underlie the >100 fold aberrant expression of *PLAU* in QPD.

7. REFERENCES

1. Marder, V.J., et al., *Hemostasis and thrombosis: basic principles and clinical practice*. 2012: Lippincott Williams & Wilkins.

2. Versteeg, H.H., et al., *New fundamentals in hemostasis.* Physiological reviews, 2013. **93**(1): p. 327-358.

3. Smith, S.A., R.J. Travers, and J.H. Morrissey, *How it all starts: Initiation of the clotting cascade.* Critical reviews in biochemistry and molecular biology, 2015(ahead-of-print): p. 1-11.

4. Davie, E.W., K. Fujikawa, and W. Kisiel, *The coagulation cascade: initiation, maintenance, and regulation.* Biochemistry, 1991. **30**(43): p. 10363-10370.

5. Rijken, D.C. and H.R. Lijnen, *New insights into the molecular mechanisms of the fibrinolytic system.* J Thromb Haemost, 2009. **7**(1): p. 4-13.

6. Law, R.H., D. Abu-Ssaydeh, and J.C. Whisstock, *New insights into the structure and function of the plasminogen/plasmin system.* Current opinion in structural biology, 2013. **23**(6): p. 836-841.

7. Hoffman, R., et al., *Hematology: basic principles and practice*. 2009: Churchill Livingstone.

8. Czekay, R.P., et al., *Plasminogen activator inhibitor-1 detaches cells from extracellular matrices by inactivating integrins.* J Cell Biol, 2003. **160**(5): p. 781-91.

9. Collen, D. and B. Wiman, *Fast-acting plasmin inhibitor in human plasma.* Blood, 1978. **51**(4): p. 563-9.

10. Konkle, B.A., et al., *Plasminogen activator inhibitor-1 mRNA is expressed in platelets and megakaryocytes and the megakaryoblastic cell line CHRF-288.* Arterioscler Thromb, 1993. **13**(5): p. 669-74.

11. Carpenter, S. and P. Mathew, *α2-Antiplasmin and its deficiency: fibrinolysis out of balance.* Haemophilia, 2008. **14**(6): p. 1250-1254.

12. Iwaki, T., T. Urano, and K. Umemura, *PAI-1, progress in understanding the clinical problem and its aetiology.* Br J Haematol, 2012. **157**(3): p. 291-8.

13. Iwaki, T., et al., *The first report of uncontrollable subchorionic and retroplacental haemorrhage inducing preterm labour in complete PAI-1 deficiency in a human.* Thromb Res, 2012. **129**(4): p. e161-3.

14. Mehta, R. and A.D. Shapiro, *Plasminogen activator inhibitor type 1 deficiency.* Haemophilia, 2008. **14**(6): p. 1255-60.

15. Deng, A., et al., *Livedoid vasculopathy associated with plasminogen activator inhibitor-1 promoter homozygosity (4G/4G) treated successfully with tissue plasminogen activator.* Arch Dermatol, 2006. **142**(11): p. 1466-9.

16. Chapin, J.C. and K.A. Hajjar, *Fibrinolysis and the control of blood coagulation.* Blood reviews, 2015. **29**(1): p. 17-24.

17. Husain, S.S., V. Gurewich, and B. Lipinski, *Purification and partial characterization of a single-chain high-molecular-weight form of urokinase from human urine.* Arch Biochem Biophys, 1983. **220**(1): p. 31-8.

18. Larsson, L.I., et al., *Distribution of urokinase-type plasminogen activator immunoreactivity in the mouse.* J Cell Biol, 1984. **98**(3): p. 894-903.

19. Grau, E. and L.A. Moroz, *Fibrinolytic activity of normal human blood monocytes.* Thromb Res, 1989. **53**(2): p. 145-62.

20. Crippa, M.P., *Urokinase-type plasminogen activator.* Int J Biochem Cell Biol, 2007. **39**(4): p. 690-4.

21. Nagamine, Y., R.L. Medcalf, and P. Munoz-Canoves, *Transcriptional and posttranscriptional regulation of the plasminogen activator system.* Thromb Haemost, 2005. **93**(4): p. 661-75.

22. Tripputi, P., et al., *Human urokinase gene is located on the long arm of chromosome 10.* Proc Natl Acad Sci U S A, 1985. **82**(13): p. 4448-52.

23. Riccio, A., et al., *The human urokinase-plasminogen activator gene and its promoter.* Nucleic Acids Research, 1985. **13**(8): p. 2759-2771.

24. Verde, P., et al., *Identification and primary sequence of an unspliced human urokinase poly (A)+ RNA.* Proceedings of the National Academy of Sciences, 1984. **81**(15): p. 4727-4731.

25. Masanori, N., et al., *Molecular cloning of cDNA coding for human preprourokinase.* Gene, 1985. **36**(1): p. 183-188.

26. Deloukas, P., et al., *The DNA sequence and comparative analysis of human chromosome 10.* Nature, 2004. **429**(6990): p. 375-381.

27. Gyetko, M., S.B. Shollenberger, and R.G. Sitrin, *Urokinase expression in mononuclear phagocytes: cytokine-specific modulation by interferon-gamma and*

tumor necrosis factor-alpha. Journal of leukocyte biology, 1992. **51**(3): p. 256-263.

28. Lijnen, H.R., B. Van Hoef, and D. Collen, *Activation with plasmin of tow-chain urokinase-type plasminogen activator derived from single-chain urokinase-type plasminogen activator by treatment with thrombin.* Eur J Biochem, 1987. **169**(2): p. 359-64.

29. Blasi, F. and N. Sidenius, *The urokinase receptor: focused cell surface proteolysis, cell adhesion and signaling.* FEBS Lett, 2010. **584**(9): p. 1923-30.

30. Ploug, M., *Identification of specific sites involved in ligand binding by photoaffinity labeling of the receptor for the urokinase-type plasminogen activator. Residues located at equivalent positions in uPAR domains I and III participate in the assembly of a composite ligand-binding site.* Biochemistry, 1998. **37**(47): p. 16494-505.

31. Ellis, V., *Plasminogen activation at the cell surface.* Curr Top Dev Biol, 2003. **54**: p. 263-312.

32. Ellis, V., *Functional analysis of the cellular receptor for urokinase in plasminogen activation. Receptor binding has no influence on the zymogenic nature of pro-urokinase.* J Biol Chem, 1996. **271**(25): p. 14779-84.

33. Wun, T.C., W.D. Schleuning, and E. Reich, *Isolation and characterization of urokinase from human plasma.* J Biol Chem, 1982. **257**(6): p. 3276-83.

34. Darras, V., et al., *Measurement of urokinase-type plasminogen activator (u-PA) with an enzyme-linked immunosorbent assay (ELISA) based on three murine monoclonal antibodies.* Thromb Haemost, 1986. **56**(3): p. 411-4.

35. Andreotti, F. and C. Kluft, *Circadian variation of fibrinolytic activity in blood.* Chronobiol Int, 1991. **8**(5): p. 336-51.

36. Park, S., et al., *Demonstration of single chain urokinase-type plasminogen activator on human platelet membrane.* Blood, 1989. **73**(6): p. 1421-5.

37. Kahr, W.H., et al., *Platelets from patients with the Quebec platelet disorder contain and secrete abnormal amounts of urokinase-type plasminogen activator.* Blood, 2001. **98**(2): p. 257-65.

38. Lenich, C., J.-N. Liu, and V. Gurewich, *Thrombin stimulation of platelets induces plasminogen activation mediated by endogenous urokinase-type plasminogen activator.* Blood, 1997. **90**(9): p. 3579-3586.

39. Bernik, M.B., *Increased plasminogen activator (urokinase) in tissue culture after fibrin deposition.* J Clin Invest, 1973. **52**(4): p. 823-34.

40. Baeten, K.M., et al., *Activation of single-chain urokinase-type plasminogen activator by platelet-associated plasminogen: a mechanism for stimulation of fibrinolysis by platelets.* J Thromb Haemost, 2010. **8**(6): p. 1313-22.

41. Ichinose, A., K. Fujikawa, and T. Suyama, *The activation of pro-urokinase by plasma kallikrein and its inactivation by thrombin.* J Biol Chem, 1986. **261**(8): p. 3486-9.

42. Petersen, L.C., *Kinetics of reciprocal pro-urokinase/plasminogen activation--stimulation by a template formed by the urokinase receptor bound to poly(D-lysine).* Eur J Biochem, 1997. **245**(2): p. 316-23.

43. GüNZLER, W.A., et al., *Structural relationship between human high and low molecular mass urokinase.* Hoppe-Seyler´ s Zeitschrift für physiologische Chemie, 1982. **363**(1): p. 133-142.

44. Kawano, T., K. Morimoto, and Y. Uemura, *Partial purification and properties of urokinase inhibitor from human placenta.* J Biochem, 1970. **67**(3): p. 333-42.

45. Kruithof, E.K., et al., *Purification and characterization of a plasminogen activator inhibitor from the histiocytic lymphoma cell line U-937.* J Biol Chem, 1986. **261**(24): p. 11207-13.

46. Gurewich, V., et al., *Effective and fibrin-specific clot lysis by a zymogen precursor form of urokinase (pro-urokinase). A study in vitro and in two animal species.* J Clin Invest, 1984. **73**(6): p. 1731-9.

47. Fleury, V., H. Lijnen, and E. Angles-Cano, *Mechanism of the enhanced intrinsic activity of single-chain urokinase-type plasminogen activator during ongoing fibrinolysis.* Journal of Biological Chemistry, 1993. **268**(25): p. 18554-18559.

48. Lee, S.W., V. Ellis, and D.A. Dichek, *Characterization of plasminogen activation by glycosylphosphatidylinositol-anchored urokinase.* Journal of Biological Chemistry, 1994. **269**(4): p. 2411-2418.

49. Manchanda, N. and B.S. Schwartz, *Lipopolysaccharide-induced modulation of human monocyte urokinase production and activity.* J Immunol, 1990. **145**(12): p. 4174-80.

50. Ellis, V. and K. Danø, *Potentiation of plasminogen activation by an anti-urokinase monoclonal antibody due to ternary complex formation. A mechanistic*

model for receptor-mediated plasminogen activation. Journal of Biological Chemistry, 1993. **268**(7): p. 4806-4813.

51. Pluskota, E., et al., *Integrin alphaMbeta2 orchestrates and accelerates plasminogen activation and fibrinolysis by neutrophils.* J Biol Chem, 2004. **279**(17): p. 18063-72.

52. Vaughan, D., E. Van Houtte, and D. Collen, *Urokinase binds to platelets through a specific saturable, low affinity mechanism.* Fibrinolysis, 1990. **4**(3): p. 141-146.

53. Dejouvencel, T., et al., *Fibrinolytic cross-talk: a new mechanism for plasmin formation.* Blood, 2010. **115**(10): p. 2048-56.

54. Webb, D.J., D. Nguyen, and S.L. Gonias, *Extracellular signal-regulated kinase functions in the urokinase receptor-dependent pathway by which neutralization of low density lipoprotein receptor-related protein promotes fibrosarcoma cell migration and matrigel invasion.* Journal of Cell Science, 2000. **113**(1): p. 123-134.

55. Konakova, M., F. Hucho, and W.D. Schleuning, *Downstream targets of urokinase-type plasminogen-activator-mediated signal transduction.* European Journal of Biochemistry, 1998. **253**(2): p. 421-429.

56. Tang, H., et al., *The urokinase-type plasminogen activator receptor mediates tyrosine phosphorylation of focal adhesion proteins and activation of mitogen-activated protein kinase in cultured endothelial cells.* Journal of Biological Chemistry, 1998. **273**(29): p. 18268-18272.

57. Busso, N., et al., *Induction of cell migration by pro-urokinase binding to its receptor: possible mechanism for signal transduction in human epithelial cells.* The Journal of Cell Biology, 1994. **126**(1): p. 259-270.

58. Nguyen, D.H., I.M. Hussaini, and S.L. Gonias, *Binding of urokinase-type plasminogen activator to its receptor in MCF-7 cells activates extracellular signal-regulated kinase 1 and 2 which is required for increased cellular motility.* Journal of Biological Chemistry, 1998. **273**(14): p. 8502-8507.

59. Ibanez-Tallon, I., et al., *In vivo analysis of the state of the human uPA enhancer following stimulation by TPA.* Oncogene, 1999. **18**(18): p. 2836-2845.

60. Ibañez-Tallon, I., et al., *Binding of Sp1 to the proximal promoter links constitutive expression of the human uPA gene and invasive potential of PC3 cells.* Blood, 2002. **100**(9): p. 3325-3332.

61. Verde, P., et al., *An upstream enhancer and a negative element in the 5' flanking region of the human urokinase plasminogen activator gene.* Nucleic acids research, 1988. **16**(22): p. 10699-10716.

62. Alfano, D., et al., *The urokinase plasminogen activator and its receptor.* Thromb Haemost, 2005. **93**: p. 205-11.

63. Ferrai, C., et al., *A transcription-dependent MNase-resistant fragment of the uPA promoter interacts with the enhancer.* Journal of Biological Chemistry, 2007.

64. Hansen, S.K., et al., *A novel complex between the p65 subunit of NF-kappa B and c-Rel binds to a DNA element involved in the phorbol ester induction of the human urokinase gene.* The EMBO journal, 1992. **11**(1): p. 205.

65. Cannio, R., P.S. Rennie, and F. Blasi, *A cell-type specific and enhancer-dependent silencer in the regulation of the expression of the human urokinase plasminogen activator gene.* Nucleic acids research, 1991. **19**(9): p. 2303-2308.

66. Stalder, M., et al., *Release of vascular plasminogen activator (v-PA) after venous stasis: electrophoretic-zymographic analysis of free and complexed v-PA.* Br J Haematol, 1985. **61**(1): p. 169-76.

67. Gyetko, M., et al., *Urokinase is required for the pulmonary inflammatory response to Cryptococcus neoformans. A murine transgenic model.* Journal of Clinical Investigation, 1996. **97**(8): p. 1818.

68. Carmeliet, P., et al., *Physiological consequences of loss of plasminogen activator gene function in mice.* Nature, 1994. **368**(6470): p. 419-424.

69. Kufrin, D., et al., *Antithrombotic thrombocytes: ectopic expression of urokinase-type plasminogen activator in platelets.* Blood, 2003. **102**(3): p. 926-933.

70. McKay, H., et al., *Bleeding risks associated with inheritance of the Quebec platelet disorder.* Blood, 2004. **104**(1): p. 159-165.

71. Lensch, M.W. and G.Q. Daley, *Origins of mammalian hematopoiesis: in vivo paradigms and in vitro models.* Current topics in developmental biology, 2004. **60**: p. 127-196.

72. Kondo, M., et al., *Biology of hematopoietic stem cells and progenitors: implications for clinical application.* Annual review of immunology, 2003. **21**(1): p. 759-806.

73. Eaves, C.J., *Hematopoietic stem cells: concepts, definitions, and the new reality.* Blood, 2015. **125**(17): p. 2605-2613.

74. Hordyjewska, A., Ł. Popiołek, and A. Horecka, *Characteristics of hematopoietic stem cells of umbilical cord blood.* Cytotechnology, 2015. **67**(3): p. 387-396.

75. Civin, C.I., et al., *Antigenic analysis of hematopoiesis. III. A hematopoietic progenitor cell surface antigen defined by a monoclonal antibody raised against KG-1a cells.* The Journal of Immunology, 1984. **133**(1): p. 157-165.

76. Orkin, S.H., *Diversification of haematopoietic stem cells to specific lineages.* Nature Reviews Genetics, 2000. **1**(1): p. 57-64.

77. Kondo, M., I.L. Weissman, and K. Akashi, *Identification of clonogenic common lymphoid progenitors in mouse bone marrow.* Cell, 1997. **91**(5): p. 661-672.

78. Akashi, K., et al., *A clonogenic common myeloid progenitor that gives rise to all myeloid lineages.* Nature, 2000. **404**(6774): p. 193-197.

79. Papatriantafyllou, M., *Haematopoiesis: A long cell cycle for myeloid differentiation.* Nat Rev Immunol, 2013. **13**(9): p. 616-617.

80. Tasneem, S., et al., *Effect of Quebec Platelet Disorder on Leukocyte Urokinase Plasminogen Activator (uPA) Levels.* ISTH 15 ABS-4056, 2015.

81. Veljkovic, D.K., et al., *Increased expression of urokinase plasminogen activator in Quebec platelet disorder is linked to megakaryocyte differentiation.* Blood, 2009. **113**(7): p. 1535-1542.

82. Yu, J. and J.A. Thomson, *1. Embryonic Stem Cells.* Regenerative medicine, 2006: p. 1.

83. Wen, Q., B. Goldenson, and J.D. Crispino, *Normal and malignant megakaryopoiesis.* Expert Rev Mol Med, 2011. **13**: p. e32.

84. Schulze, H., et al., *Characterization of the megakaryocyte demarcation membrane system and its role in thrombopoiesis.* Blood, 2006. **107**(10): p. 3868-3875.

85. Briddell, R.A., et al., *Characterization of the human burst-forming unit-megakaryocyte.* Blood, 1989. **74**(1): p. 145-151.

86. Patel, S.R., J.H. Hartwig, and J.E. Italiano Jr, *The biogenesis of platelets from megakaryocyte proplatelets.* Journal of Clinical Investigation, 2005. **115**(12): p. 3348.

87. Richardson, J.L., et al., *Mechanisms of organelle transport and capture along proplatelets during platelet production.* Blood, 2005. **106**(13): p. 4066-4075.

88. Tomer, A., L.A. Harker, and S.A. Burstein, *Purification of human megakaryocytes by fluorescence-activated cell sorting.* Blood, 1987. **70**(6): p. 1735-1742.

89. Tomer, A., L.A. Harker, and S.A. Burstein, *Flow cytometric analysis of normal human megakaryocytes.* Blood, 1988. **71**(5): p. 1244-1252.

90. Zimmet, J. and K. Ravid, *Polyploidy: occurrence in nature, mechanisms, and significance for the megakaryocyte-platelet system.* Experimental hematology, 2000. **28**(1): p. 3-16.

91. Italiano, J.E., et al., *Blood platelets are assembled principally at the ends of proplatelet processes produced by differentiated megakaryocytes.* The Journal of cell biology, 1999. **147**(6): p. 1299-1312.

92. Italiano, J. and R. Shivdasani, *Megakaryocytes and beyond: the birth of platelets.* Journal of Thrombosis and Haemostasis, 2003. **1**(6): p. 1174-1182.

93. Heijnen, H.F., et al., *Multivesicular bodies are an intermediate stage in the formation of platelet α-granules.* Blood, 1998. **91**(7): p. 2313-2325.

94. Youssefian, T., et al., *Platelet and megakaryocyte dense granules contain glycoproteins Ib and IIb-IIIa.* Blood, 1997. **89**(11): p. 4047-4057.

95. Cramer, E.M., et al., *Uncoordinated expression of fibrinogen compared with thrombospondin and von Willebrand factor in maturing human megakaryocytes.* Blood, 1989. **73**(5): p. 1123-1129.

96. Cramer, E.M., et al., *Alpha-granule pool of glycoprotein IIb-IIIa in normal and pathologic platelets and megakaryocytes.* Blood, 1990. **75**(6): p. 1220-1227.

97. Avecilla, S.T., et al., *Chemokine-mediated interaction of hematopoietic progenitors with the bone marrow vascular niche is required for thrombopoiesis.* Nature medicine, 2004. **10**(1): p. 64-71.

98. Kaushansky, K., *The molecular mechanisms that control thrombopoiesis.* Journal of Clinical Investigation, 2005. **115**(12): p. 3339.

99. Kaushansky, K., *Thrombopoietin and the hematopoietic stem cell.* Annals of the New York Academy of Sciences, 2005. **1044**(1): p. 139-141.

100. Guo, T., et al., *Megakaryopoiesis and platelet production: insight into hematopoietic stem cell proliferation and differentiation.* Stem Cell Investigation, 2015. **2**(2).

101. Tozawa, K., Y. Ono-Uruga, and Y. Matsubara, *Megakaryopoiesis.* Clinical and Experimental Thrombosis and Hemostasis, 2014. **1**(2): p. 54-58.

102. Zucker-Franklin, D. and S. Petursson, *Thrombocytopoiesis--analysis by membrane tracer and freeze-fracture studies on fresh human and cultured mouse megakaryocytes.* The Journal of cell biology, 1984. **99**(2): p. 390-402.

103. Choi, E., et al., *Platelets generated in vitro from proplatelet-displaying human.* Blood, 1995. **85**(2): p. 402-413.

104. Kaushansky, K., *Lineage-specific hematopoietic growth factors.* New England Journal of Medicine, 2006. **354**(19): p. 2034-2045.

105. Avraham, H., et al., *Effects of the stem cell factor, c-kit ligand, on human megakaryocytic cells.* Blood, 1992. **79**(2): p. 365-371.

106. Briddell, R.A., et al., *Effect of c-kit ligand on in vitro human megakaryocytopoiesis.* Blood, 1991. **78**(11): p. 2854-2859.

107. Debili, N., et al., *Effects of the recombinant hematopoietic growth factors interleukin-3, interleukin-6, stem cell factor, and leukemia inhibitory factor on the megakaryocytic differentiation of CD34+ cells.* Blood, 1993. **82**(1): p. 84-95.

108. Ikebuchi, K., et al., *Interleukin 6 enhancement of interleukin 3-dependent proliferation of multipotential hemopoietic progenitors.* Proceedings of the National Academy of Sciences, 1987. **84**(24): p. 9035-9039.

109. Fujiki, H., et al., *Role of human interleukin-9 as a megakaryocyte potentiator in culture.* Exp Hematol, 2002. **30**(12): p. 1373-80.

110. Elagib, K.E., et al., *RUNX1 and GATA-1 coexpression and cooperation in megakaryocytic differentiation.* Blood, 2003. **101**(11): p. 4333-4341.

111. Tsang, A.P., et al., *FOG, a multitype zinc finger protein, acts as a cofactor for transcription factor GATA-1 in erythroid and megakaryocytic differentiation.* Cell, 1997. **90**(1): p. 109-119.

112. Deutsch, V.R. and A. Tomer, *Megakaryocyte development and platelet production.* British journal of haematology, 2006. **134**(5): p. 453-466.

113. Rekhtman, N., et al., *PU. 1 and pRB interact and cooperate to repress GATA-1 and block erythroid differentiation.* Molecular and Cellular Biology, 2003. **23**(21): p. 7460-7474.

114. Gurbuxani, S., P. Vyas, and J.D. Crispino, *Recent insights into the mechanisms of myeloid leukemogenesis in Down syndrome.* Blood, 2004. **103**(2): p. 399-406.

115. Iwasaki, H., et al., *GATA-1 converts lymphoid and myelomonocytic progenitors into the megakaryocyte/erythrocyte lineages.* Immunity, 2003. **19**(3): p. 451-462.

116. Lemarchandel, V., et al., *GATA and Ets cis-acting sequences mediate megakaryocyte-specific expression.* Molecular and Cellular Biology, 1993. **13**(1): p. 668-676.

117. Stockley, J., et al., *Enrichment of FLI1 and RUNX1 mutations in families with excessive bleeding and platelet dense granule secretion defects.* Blood, 2013. **122**(25): p. 4090-4093.

118. Fiedler, K. and C. Brunner, *The role of transcription factors in the guidance of granulopoiesis.* Am J Blood Res, 2012. **2**(1): p. 57-65.

119. Zhang, P., et al., *Enhancement of hematopoietic stem cell repopulating capacity and self-renewal in the absence of the transcription factor C/EBPα.* Immunity, 2004. **21**(6): p. 853-863.

120. Diamandis, M., et al., *Evaluation of urokinase plasminogen activator in urine from individuals with Quebec platelet disorder.* Blood Coagul Fibrinolysis, 2008. **19**(5): p. 463-4.

121. Diamandis, M., et al., *Quebec platelet disorder is linked to the urokinase plasminogen activator gene (PLAU) and increases expression of the linked allele in megakaryocytes.* Blood, 2009. **113**(7): p. 1543-6.

122. Hayward, C.P. and G.E. Rivard, *Quebec platelet disorder.* 2011.

123. Veljkovic, D.K., et al., *Increased expression of urokinase plasminogen activator in Quebec platelet disorder is linked to megakaryocyte differentiation.* Blood, 2009. **113**(7): p. 1535-42.

124. Hayward, C., et al., *An autosomal dominant, qualitative platelet disorder associated with multimerin deficiency, abnormalities in platelet factor V, thrombospondin, von Willebrand factor, and fibrinogen and an epinephrine aggregation defect.* Blood, 1996. **87**(12): p. 4967-4978.

125. Tracy, P.B., et al., *Factor V (Quebec): a bleeding diathesis associated with a qualitative platelet Factor V deficiency.* Journal of Clinical Investigation, 1984. **74**(4): p. 1221.

126. Hayward, C.P., et al., *Studies of a second family with the Quebec platelet disorder: evidence that the degradation of the alpha-granule membrane and its soluble contents are not secondary to a defect in targeting proteins to alpha-granules.* Blood, 1997. **89**(4): p. 1243-53.

127. Janeway, C.M., et al., *Factor V Quebec revisited.* Blood, 1996. **87**(9): p. 3571-3578.

128. Hayward, C., et al., *Diagnostic utility of light transmission platelet aggregometry: results from a prospective study of individuals referred for bleeding disorder assessments.* Journal of Thrombosis and Haemostasis, 2009. **7**(4): p. 676-684.

129. Sheth, P.M., et al., *Intracellular activation of the fibrinolytic cascade in the Quebec Platelet Disorder.* Thromb Haemost, 2003. **90**(2): p. 293-8.

130. Diamandis, M., et al., *Insights into abnormal hemostasis in the Quebec platelet disorder from analyses of clot lysis.* J Thromb Haemost, 2006. **4**(5): p. 1086-94.

131. Hayward, C., et al., *Fibrinogen degradation products in patients with the Quebec platelet disorder.* British journal of haematology, 1997. **97**(2): p. 497-503.

132. Paterson, A.D., et al., *Persons with Quebec platelet disorder have a tandem duplication of PLAU, the urokinase plasminogen activator gene.* Blood, 2010. **115**(6): p. 1264-6.

133. Tanaka, M., et al., *Forkhead Family Transcription Factor FKHRL1 Is Expressed in Human Megakaryocytes REGULATION OF CELL CYCLING AS A DOWNSTREAM MOLECULE OF THROMBOPOIETIN SIGNALING.* Journal of Biological Chemistry, 2001. **276**(18): p. 15082-15089.

134. Kirito, K. and K. Kaushansky, *Transcriptional regulation of megakaryopoiesis: thrombopoietin signaling and nuclear factors.* Current opinion in hematology, 2006. **13**(3): p. 151-156.

135. Margueron, R. and D. Reinberg, *Chromatin structure and the inheritance of epigenetic information.* Nature Reviews Genetics, 2010. **11**(4): p. 285-296.

136. Shlyueva, D., G. Stampfel, and A. Stark, *Transcriptional enhancers: from properties to genome-wide predictions.* Nat Rev Genet, 2014. **15**(4): p. 272-86.

137. Lenhard, B., A. Sandelin, and P. Carninci, *Metazoan promoters: emerging characteristics and insights into transcriptional regulation.* Nature Reviews Genetics, 2012. **13**(4): p. 233-245.

138. Riethoven, J.-J.M., *Regulatory regions in DNA: promoters, enhancers, silencers, and insulators.* Computational Biology of Transcription Factor Binding, 2010: p. 33-42.

139. Roy, A.L. and D.S. Singer, *Core promoters in transcription: old problem, new insights.* Trends in biochemical sciences, 2015. **40**(3): p. 165-171.

140. Amano, T., et al., *Chromosomal dynamics at the Shh locus: limb bud-specific differential regulation of competence and active transcription.* Developmental cell, 2009. **16**(1): p. 47-57.

141. Arnone, M.I. and E.H. Davidson, *The hardwiring of development: organization and function of genomic regulatory systems.* Development, 1997. **124**(10): p. 1851-1864.

142. Brookes, E. and Y. Shi, *Diverse epigenetic mechanisms of human disease.* Annual review of genetics, 2014. **48**: p. 237-268.

143. Lan, F. and Y. Shi, *Epigenetic regulation: methylation of histone and non-histone proteins.* Science in China Series C: Life Sciences, 2009. **52**(4): p. 311-322.

144. Berger, S.L., et al., *An operational definition of epigenetics.* Genes & development, 2009. **23**(7): p. 781-783.

145. Jones, P.A. and S.B. Baylin, *The fundamental role of epigenetic events in cancer.* Nat Rev Genet, 2002. **3**(6): p. 415-28.

146. Suzuki, M.M. and A. Bird, *DNA methylation landscapes: provocative insights from epigenomics.* Nat Rev Genet, 2008. **9**(6): p. 465-76.

147. Berdasco, M. and M. Esteller, *DNA methylation in stem cell renewal and multipotency.* Stem Cell Res Ther, 2011. **2**(5): p. 42.

148. Munoz, P., M.S. Iliou, and M. Esteller, *Epigenetic alterations involved in cancer stem cell reprogramming.* Mol Oncol, 2012. **6**(6): p. 620-36.

149. Jones, P.A. and D. Takai, *The role of DNA methylation in mammalian epigenetics.* Science, 2001. **293**(5532): p. 1068-70.

150. Straussman, R., et al., *Developmental programming of CpG island methylation profiles in the human genome.* Nat Struct Mol Biol, 2009. **16**(5): p. 564-71.

151. Klose, R.J. and A.P. Bird, *Genomic DNA methylation: the mark and its mediators.* Trends Biochem Sci, 2006. **31**(2): p. 89-97.

152. Shilatifard, A., *Chromatin modifications by methylation and ubiquitination: implications in the regulation of gene expression.* Annu Rev Biochem, 2006. **75**: p. 243-69.

153. Bogdanovic, O. and G.J. Veenstra, *DNA methylation and methyl-CpG binding proteins: developmental requirements and function.* Chromosoma, 2009. **118**(5): p. 549-65.

154. Bibikova, M., et al., *High density DNA methylation array with single CpG site resolution.* Genomics, 2011. **98**(4): p. 288-295.

155. Song, Q., et al., *A reference methylome database and analysis pipeline to facilitate integrative and comparative epigenomics.* 2013.

156. Bannister, A.J. and T. Kouzarides, *Regulation of chromatin by histone modifications.* Cell Res, 2011. **21**(3): p. 381-95.

157. Wang, Z., et al., *Combinatorial patterns of histone acetylations and methylations in the human genome.* Nature genetics, 2008. **40**(7): p. 897-903.

158. Ng, S., et al., *Dynamic protein methylation in chromatin biology.* Cellular and molecular life sciences, 2009. **66**(3): p. 407-422.

159. Xhemalce, B., M.A. Dawson, and A.J. Bannister, *Histone modifications.* Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2011.

160. Kim, J., et al., *Tudor, MBT and chromo domains gauge the degree of lysine methylation.* EMBO reports, 2006. **7**(4): p. 397-403.

161. Du, J., et al., *DNA methylation pathways and their crosstalk with histone methylation.* Nature Reviews Molecular Cell Biology, 2015. **16**(9): p. 519-532.

162. Young, M.D., et al., *ChIP-seq analysis reveals distinct H3K27me3 profiles that correlate with transcriptional activity.* Nucleic Acids Res, 2011. **39**(17): p. 7415-27.

163. Creyghton, M.P., et al., *Histone H3K27ac separates active from poised enhancers and predicts developmental state.* Proc Natl Acad Sci U S A, 2010. **107**(50): p. 21931-6.

164. Sims, R.J., 3rd and D. Reinberg, *Processing the H3K36me3 signature.* Nat Genet, 2009. **41**(3): p. 270-1.

165. Pekowska, A., et al., *A unique H3K4me2 profile marks tissue-specific gene regulation.* Genome Res, 2010. **20**(11): p. 1493-502.

166. Orford, K., et al., *Differential H3K4 methylation identifies developmentally poised hematopoietic genes.* Developmental cell, 2008. **14**(5): p. 798-809.

167. Barski, A., et al., *High-resolution profiling of histone methylations in the human genome.* Cell, 2007. **129**(4): p. 823-837.

168. Heintzman, N.D., et al., *Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome.* Nat Genet, 2007. **39**(3): p. 311-8.

169. Schmidt, D., et al., *ChIP-seq: using high-throughput sequencing to discover protein-DNA interactions.* Methods, 2009. **48**(3): p. 240-8.

170. Hatami, J., et al., *Proliferation extent of CD34+ cells as a key parameter to maximize megakaryocytic differentiation of umbilical cord blood-derived hematopoietic stem/progenitor cells in a two-stage culture protocol.* Biotechnology Reports, 2014. **4**: p. 50-55.

171. Tijssen, M.R., et al., *Genome-wide analysis of simultaneous GATA1/2, RUNX1, FLI1, and SCL binding in megakaryocytes identifies hematopoietic regulators.* Dev Cell, 2011. **20**(5): p. 597-609.

172. Martens, J.H. and H.G. Stunnenberg, *BLUEPRINT: mapping human blood cell epigenomes.* haematologica, 2013. **98**(10): p. 1487-1489.

173. Kim, T.H., et al., *A high-resolution map of active promoters in the human genome.* Nature, 2005. **436**(7052): p. 876-80.

174. Koch, C.M., et al., *The landscape of histone modifications across 1% of the human genome in five human cell lines.* Genome Res, 2007. **17**(6): p. 691-707.

175. Kim, T.H., et al., *Direct isolation and identification of promoters in the human genome.* Genome Res, 2005. **15**(6): p. 830-9.

176. Schmittgen, T.D. and K.J. Livak, *Analyzing real-time PCR data by the comparative CT method.* Nature protocols, 2008. **3**(6): p. 1101-1108.

177. Andrews, S., *FastQC: A quality control tool for high throughput sequence data.* Reference Source, 2010.

178. Li, H. and R. Durbin, *Fast and accurate short read alignment with Burrows-Wheeler transform.* Bioinformatics, 2009. **25**(14): p. 1754-60.

179. Heinz, S., et al., *Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities.* Mol Cell, 2010. **38**(4): p. 576-89.

180. Landt, S.G., et al., *ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia.* Genome Res, 2012. **22**(9): p. 1813-31.

181. Zhang, Y., et al., *Model-based analysis of ChIP-Seq (MACS).* Genome Biol, 2008. **9**(9): p. R137.

182. Kent, W.J., et al., *The human genome browser at UCSC.* Genome research, 2002. **12**(6): p. 996-1006.

183. Shen, L., et al., *diffReps: detecting differential chromatin modification sites from ChIP-seq data with biological replicates.* PLoS One, 2013. **8**(6): p. e65598.

184. Liao, Y., G.K. Smyth, and W. Shi, *featureCounts: an efficient general purpose program for assigning sequence reads to genomic features.* Bioinformatics, 2014. **30**(7): p. 923-30.

185. Love, M.I., W. Huber, and S. Anders, *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.* Genome Biol, 2014. **15**(12): p. 550.

186. Bocker, M.T., et al., *Genome-wide promoter DNA methylation dynamics of human hematopoietic progenitor cells during differentiation and aging.* Blood, 2011. **117**(19): p. e182-9.

187. Consortium, E.P., *An integrated encyclopedia of DNA elements in the human genome.* Nature, 2012. **489**(7414): p. 57-74.

188. Li, G., et al., *Chromatin Interaction Analysis with Paired-End Tag (ChIA-PET) sequencing technology and application.* BMC Genomics, 2014. **15 Suppl 12**: p. S11.

189. Van Steensel, B. and J. Dekker, *Genomics tools for unraveling chromosome architecture.* Nature biotechnology, 2010. **28**(10): p. 1089-1095.

190. Fullwood, M.J., et al., *An oestrogen-receptor-&agr;-bound human chromatin interactome.* Nature, 2009. **462**(7269): p. 58-64.

191. Rüdiger, M., et al., *Differential actin organization by vinculin isoforms: implications for cell type-specific microfilament anchorage.* FEBS letters, 1998. **431**(1): p. 49-54.

192. Xu, W., H. Baribault, and E.D. Adamson, *Vinculin knockout results in heart and brain defects during embryonic development.* Development, 1998. **125**(2): p. 327-337.

193. Zemljic-Harpf, A.E., et al., *Heterozygous inactivation of the vinculin gene predisposes to stress-induced cardiomyopathy.* The American journal of pathology, 2004. **165**(3): p. 1033-1044.

194. Mitsios, J.V., et al., *What is vinculin needed for in platelets?* Journal of Thrombosis and Haemostasis, 2010. **8**(10): p. 2294-2304.

195. Kim, S., N.K. Yu, and B.K. Kaang, *CTCF as a multifunctional protein in genome regulation and gene expression.* Exp Mol Med, 2015. **47**: p. e166.

196. Chernukhin, I., et al., *CTCF interacts with and recruits the largest subunit of RNA polymerase II to CTCF target sites genome-wide.* Mol Cell Biol, 2007. **27**(5): p. 1631-48.

197. Chacon, D., et al., *BloodChIP: a database of comparative genome-wide transcription factor binding profiles in human blood cells.* Nucleic acids research, 2013: p. gkt1036.

198. Mikkola, H.K., et al., *Haematopoietic stem cells retain long-term repopulating activity and multipotency in the absence of stem-cell leukaemia SCL/tal-1 gene.* Nature, 2003. **421**(6922): p. 547-551.

199. Hart, A., et al., *Fli-1 is required for murine vascular and megakaryocytic development and is hemizygously deleted in patients with thrombocytopenia.* Immunity, 2000. **13**(2): p. 167-177.

200. Chadwick, L.H., *The NIH Roadmap Epigenomics Program data resource.* Epigenomics, 2012. **4**(3): p. 317-24.

201. Buenrostro, J.D., et al., *Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position.* Nat Methods, 2013. **10**(12): p. 1213-8.

202. Cong, L., et al., *Multiplex genome engineering using CRISPR/Cas systems.* Science, 2013. **339**(6121): p. 819-823.

203. Jinek, M., et al., *RNA-programmed genome editing in human cells.* Elife, 2013. **2**: p. e00471.

204. Yang, L., et al., *CRISPR-Cas-mediated targeted genome editing in human cells*, in *Gene Correction*. 2014, Springer. p. 245-267.

205. Cho, S.W., et al., *Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease.* Nature biotechnology, 2013. **31**(3): p. 230-232.

206. Mali, P., et al., *RNA-guided human genome engineering via Cas9.* Science, 2013. **339**(6121): p. 823-826.

8. APPENDIX

Experiment ID	Cell	Sample ID	Factor	"# Raw Reads"	"# Mapped Reads"	"% Mapped"	"# Uniquely Mapped Reads"	"% Uniquely Mapped"	PBC	NRF	NSC	RSC
AS3.13.17	CD66b	P126-QPD	H3K27Ac	16725154	15662217	93.64	14114438	84.39	0.65	0.63	1.76639	1.886598
AS3.13.9	CD66b	P132-QPD	H3K27Ac	19519900	17140695	87.81	15278489	78.27	0.92	0.91	2.100229	1.312399
AS3.13.19	CD66b	P183-QPD	H3K27Ac	15028259	14286784	95.07	12703397	84.53	0.98	0.97	1.50674	1.185977
AS3.13.4	CD66b	C102-CTL	H3K27Ac	11626788	10731786	92.3	9746389	83.83	0.67	0.69	1.503072	1.390279
AS3.13.13	CD66b	C51-CTL	H3K27Ac	15952588	14486691	90.81	13183468	82.64	0.93	0.91	2.990927	1.279185
AS3.13.23	CD66b	C117-CTL	H3K27Ac	13737028	12975342	94.46	11362424	82.71	0.95	0.95	1.480705	1.196916
AS3.13.18	CD66b	P126-QPD	H3K36me3	16590183	15971081	96.27	14466442	87.2	0.96	0.95	1.318757	1.367249
AS3.13.10	CD66b	P132-QPD	H3K36me3	19361269	18136149	93.67	16639191	85.94	0.93	0.92	1.386387	1.320886
AS3.13.20	CD66b	P183-QPD	H3K36me3	17350467	16824190	96.97	15166410	87.41	0.96	0.96	1.243112	1.169048
AS3.13.5	CD66b	C102-CTL	H3K36me3	13066235	12725613	97.39	11550033	88.4	0.97	0.97	1.252753	1.236882
AS3.13.14	CD66b	C51-CTL	H3K36me3	9096840	8455506	92.95	7781765	85.54	0.96	0.96	1.363774	1.354411
AS3.13.24	CD66b	C117-CTL	H3K36me3	18765481	18226449	97.13	16612447	88.53	0.94	0.93	1.314837	1.343097
AS3.13.7	CD66b	P126-QPD	H3K4me2	14481773	13481729	93.09	12660753	87.43	0.94	0.93	2.17332	1.122111
AS3.13.11	CD66b	P132-QPD	H3K4me2	17883780	16267644	90.96	15209554	85.05	0.92	0.91	2.646803	1.181675
AS3.13.21	CD66b	P183-QPD	H3K4me2	22831803	21587421	94.55	20145071	88.23	0.92	0.91	2.210537	1.145823
AS3.13.6	CD66b	C102-CTL	H3K4me2	15995430	15207028	95.07	14230807	88.97	0.94	0.94	2.167111	1.137377
AS3.13.25	CD66b	C117-CTL	H3K4me2	19014930	18069224	95.03	16839698	88.56	0.93	0.92	2.128812	1.10826
AS3.13.8	CD66b	P126-QPD	Input	15972241	14547712	91.08	12590765	78.83	0.99	0.98	1.030418	0.3720358
AS3.13.12	CD66b	P132-QPD	Input	17053554	15299494	89.71	13121210	76.94	0.99	0.98	1.050589	0.508091
AS3.13.22	CD66b	P183-QPD	Input	15733736	15025785	95.5	13269522	84.34	0.99	0.98	1.021062	0.4207339
AS3.13.3	CD66b	C102-CTL	Input	15532665	14821907	95.42	13165062	84.76	0.99	0.98	1.019056	0.3903596
AS3.13.16	CD66b	C51-CTL	Input	15416202	14285980	92.67	12569230	81.53	0.99	0.99	1.036201	0.5290722
AS3.13.26	CD66b	C117-CTL	Input	27686541	26207817	94.66	22873839	82.62	0.97	0.96	1.02134	0.5773602

Supplementary Table 1. **Summary of quality control data for ChIP-seq samples used for granulocyte library sequencing.** The total and uniquely aligned reads along with quality control metrics that represent successful ChIP-seq experiments; PBC > 0.5 and NRF > 0.8 that represent ChIP-seq library complexity, NSC and RSC > 1.05 and > 0.8 that represent ChIP-seq signal to noise ratio independent of peak calling for QPD and control granulocyte ChIP-seq experiments are displayed [180].

Experiment ID	Cell	Sample ID	Factor	"# Raw Reads"	"# Mapped Reads"	"% Mapped"	"# Uniquely Mapped Reads"	"% Uniquely Mapped"	PBC	NRF	NSC	RSC
AS14.112.1	CD41	P183-QPD	H3K27Ac	22034982	19352268	87.83	17360768	78.79	0.93	0.92	1.753947	1.177206
AS14.112.2	CD41	P10-QPD	H3K27Ac	21432689	18539386	86.5	16750403	78.15	0.92	0.91	2.035316	1.156563
AS14.132.3	CD41	P132-QPD	H3K27Ac	23401565	16012389	68.42	14535868	62.11	0.94	0.93	1.847324	1.104207
AS14.112.3	CD41	C102-CTL	H3K27Ac	21543575	18923709	87.84	17060391	79.19	0.91	0.91	1.763068	1.167512
AS14.112.4	CD41	C108-CTL	H3K27Ac	22182610	19884968	89.64	17912627	80.75	0.95	0.94	1.805926	1.119641
AS14.95.4	CD41	C112-CTL	H3K27Ac	19348640	16077604	83.09	14544766	75.17	0.89	0.89	1.85571	1.217377
AS14.127.1	CD41	P183-QPD	H3K36me3	14300933	13754835	96.18	11977715	83.75	0.99	0.99	1.085443	0.7909247
AS14.127.2	CD41	P10-QPD	H3K36me3	15565705	14996354	96.34	13026097	83.68	0.99	0.99	1.119006	0.8670586
AS14.132.5	CD41	P132-QPD	H3K36me3	19915501	15977407	80.23	13978988	70.19	0.99	0.98	1.026811	0.2648816
AS14.127.3	CD41	C102-CTL	H3K36me3	15541826	14846558	95.53	12998719	83.64	0.99	0.99	1.062116	0.7669196
AS14.132.6	CD41	C108-CTL	H3K36me3	22427624	18063363	80.54	15692728	69.97	0.99	0.98	1.021111	0.2425622
AS14.127.4	CD41	C112-CTL	H3K36me3	16763010	16067746	95.85	14061075	83.88	0.99	0.99	1.078971	0.8132736
AS14.127.5	CD41	P183-QPD	H3K4me2	19023573	18235942	95.86	16932890	89.01	0.96	0.96	1.700625	1.081229
AS14.127.6	CD41	P10-QPD	H3K4me2	21707503	21040255	96.93	19585674	90.23	0.94	0.94	1.928107	1.059712
AS14.132.7	CD41	P132-QPD	H3K4me2	23079870	18092556	78.39	16900051	73.22	0.96	0.96	1.634169	1.022634
AS14.127.7	CD41	C102-CTL	H3K4me2	18893220	17895567	94.72	16574939	87.73	0.96	0.96	1.615063	1.051664
AS14.132.8	CD41	C108-CTL	H3K4me2	27453398	20940677	76.28	19593908	71.37	0.95	0.94	1.746176	1.028862
AS14.95.9	CD41	P183-QPD	Input	18746183	16695301	89.06	14736659	78.61	0.99	0.98	1.030275	0.4731462
AS14.95.10	CD41	P10-QPD	Input	23101258	21220301	91.86	18753498	81.18	0.99	0.98	1.012521	0.2367745
AS14.132.9	CD41	P132-QPD	Input	27411718	20663164	75.38	18315210	66.82	0.99	0.98	1.009183	0.2039184
AS14.112.11	CD41	C102-CTL	Input	25095777	22817359	90.92	20260015	80.73	0.99	0.98	1.010051	0.2288737
AS14.95.11	CD41	C108-CTL	Input	19648584	18313497	93.21	16221463	82.56	0.99	0.98	1.019049	0.3434809
AS14.95.12	CD41	C112-CTL	Input	20848469	18922305	90.76	16674388	79.98	0.98	0.98	1.014133	0.2751542

Supplementary Table 2. Summary of quality control data for ChIP-seq samples used for MK library sequencing. The total and uniquely aligned reads along with quality control metrics that represent successful ChIP-seq experiments; PBC > 0.5 and NRF > 0.8 that represent ChIP-seq library complexity, NSC and RSC > 1.05 and > 0.8 that represent ChIP-seq signal to noise ratio independent of peak calling for QPD and control MK ChIP-seq experiments are displayed [180].