CHAPTER 5

Functional follow-up of the platelet volume and function locus at chromosome 7q22.3.

This chapter is in parts based on the following publication:

Paul, D.S., et al. (2011). Maps of open chromatin guide the functional follow-up of genome-wide association signals: application to hematological traits. PLoS Genet. 7, e1002139.

Collaboration note:

Section 5.2: DNA amplification by PCR and capillary sequencing in 643 individuals was performed at the Wellcome Trust Sanger Institute (WTSI) by the ExoSeq/ExoCan facility[1]. I designed the experiments and reviewed potential SNP positions.

Section 5.4: The platelet sample cohort of 24 healthy individuals was collected by Suthesh Sivapalaratnam[2], Hanneke Basart[2] and Mieke D. Trip[2]. Processing of raw expression and genotyping data was performed by Augusto Rendon[3–6]. Expression QTL data in macrophages and monocytes, as well as in LCLs, adipose and skin, were generated as part of the Cardiogenics and MuTHER consortia, respectively. I performed the eQTL analysis with help from Tsun-Po Yang[1].

Section 5.5: The breeding, maintaining and processing of the mice used in this study was performed by Katta Hautaviita[1], Jonna Tallila[1], Jacqui White[1], the staff from the WTSI's Research Support Facility (RSF) and Mouse Genetics Project. James P. Nisbet[1] performed total RNA extraction and whole-genome gene expression profiling. I processed and analysed the expression data, and performed functional ontology classification and canonical pathway analysis.

Section 5.6: The protein-protein interaction network of core proteins was created in collaboration with Stuart Meacham[3,4]. I retrieved the orthologous human genes and their respective proteins, and interpreted all results.

[1]Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK; [2]Department of Vascular Medicine, Academic Medical Centre Amsterdam, Amsterdam, The Netherlands; [3]Department of Haematology, University of Cambridge, Cambridge, UK; [4]National Health Service (NHS) Blood and Transplant, Cambridge, UK; [5]Biostatistics Unit, Medical Research Council, Cambridge, UK; [6]NIHR Biomedical Research Centre, Cambridge, UK.

5.1. Introduction

In **Chapters 3** and **4**, I identified putative functional variants associated with haematological traits by intersecting trait-associated SNPs with maps of open chromatin in myeloid cell types. Among many, I identified an NDR harbouring rs342293 at chromosome 7q22.3 in the megakaryocytic cell line CHRF-288-11 ('MK cells') but not in the erythroblastoid cell line K562 ('EB cells') (**Chapter 3**). This common non-coding sequence variant is known to be associated with platelet volume and function. In **Chapter 4**, I demonstrated that this NDR was also present in primary megakaryocytes, and not in erythroblasts and monocytes, confirming the initial findings from cell lines.

Cell type-specific open chromatin regions are likely to play a role in regulation of cell type-specific gene expression. Therefore, sequence variants in NDRs restricted to a cell type may have the potential to impact traits linked to that specific cell type, for example megakaryocytes/platelets and platelet volume. However, functional studies at individual loci are crucial to prove that these non-coding sequence variants do indeed have a functional effect on the regulatory element, for example by disruption/introduction of transcription factor binding sites, and the phenotypic trait of interest.

The third objective of this thesis was to perform a proof-of-principle experiment to define the molecular mechanism of a GWA locus associated with platelet volume and function. In this chapter, I use *in silico* transcription factor binding site prediction, electrophoretic mobility shift assays (EMSAs), expression QTL (eQTL) analyses, transcription factor chromatin immunoprecipitation (ChIP) and mouse models, to elucidate the molecular basis of the association with the platelet phenotype at the 7q22.3 locus.

5.2. Identification of rs342293 as the only likely putative functional candidate at the 7q22.3 locus

The 65 kb recombination interval associated with platelet volume and function at chromosome 7q22.3 (Soranzo, Rendon, et al., 2009; Soranzo, Spector, et al., 2009; **Figure 5-1 A**) exhibited a total of six distinct NDRs in MK and EB cells. Of these six NDRs, two were common to both cell types and located at an evolutionary conserved element 10 kb upstream of the GWA index SNP rs342293 and at the promoter region of *FLJ36031*, three were specific to EB cells, and one was specific to MK cells (**Figure 5-1 B**).

The MK-specific NDR contained SNPs associated with mean platelet volume (MPV): rs342293 (MAF=0.48, Phase II+III HapMap, CEU population) and its best proxy rs342294 (r^2=1.0, Phase II+III HapMap, CEU). Data from the 1000 Genomes Project (Pilot 1, CEU; The 1000 Genomes Project Consortium, 2010) revealed 34 SNPs in LD ($r^2 \geq 0.8$) with rs342293, and confirmed that only rs342293 and rs342294 fall into the MK-specific NDR (**Appendix, Table 8-5**). This NDR was absent in both FAIRE-chip data sets (8 and 12 min formaldehyde cross-linking conditions) in EB cells (**Figure 5-1 C**). Irrespective of LD to rs342293, there were no sequence variants reported by the 1000 Genomes Project (Pilot 1, CEU) within sites of open chromatin in MK cells at the recombination interval.

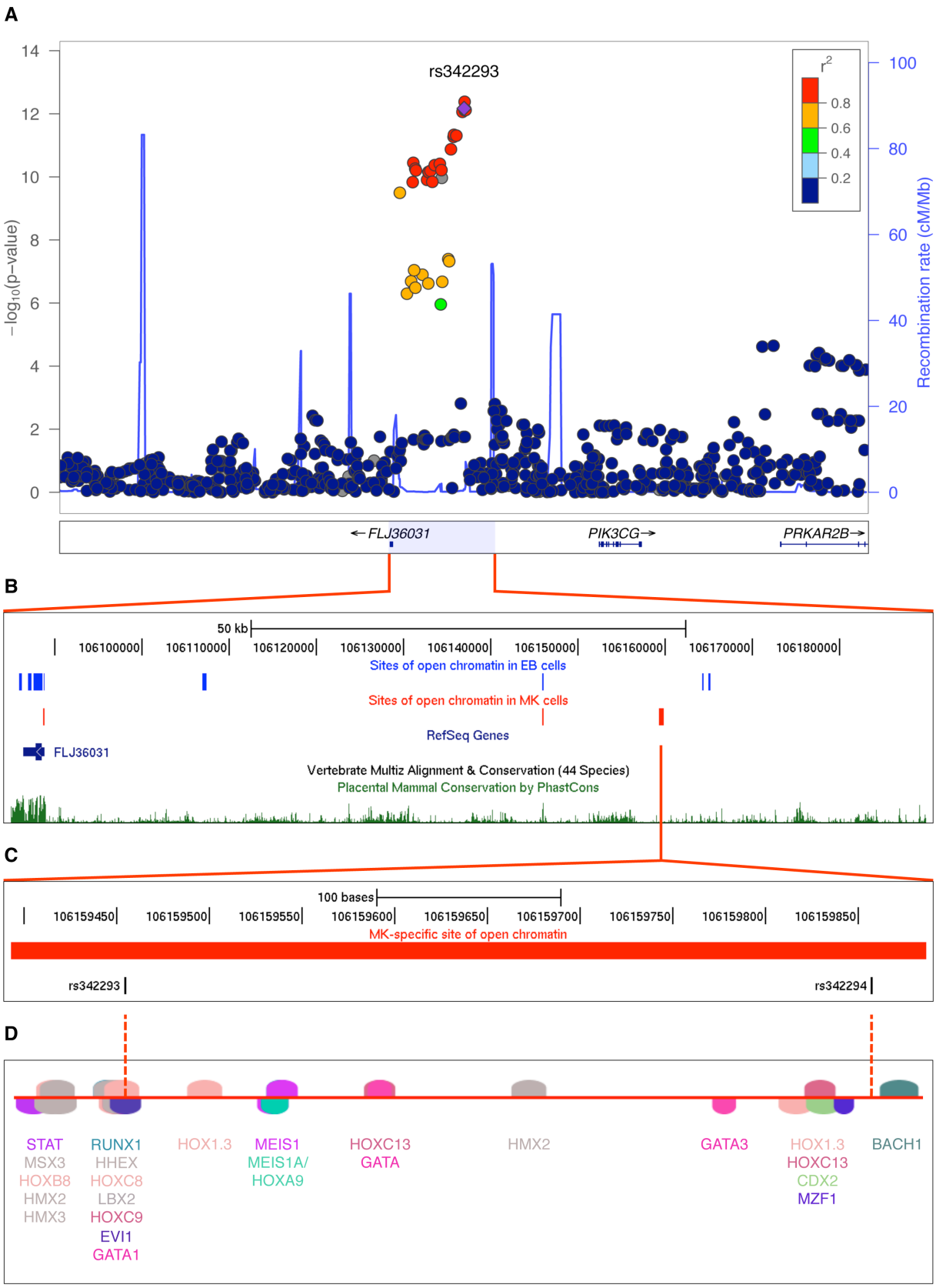


Figure 5-1. Functional follow-up of the 7q22.3 locus associated with platelet volume and function. (**A**) Regional plot of the 7q22.3 locus showing data from the discovery GWA meta-analysis as reported by Soranzo, Spector, et al., 2009. The SNP rs342293 is indicated in purple (P=6.75x10^{-13}). Values for

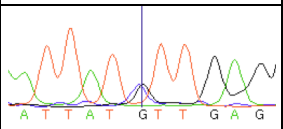
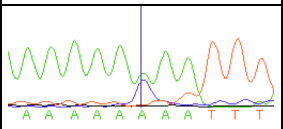
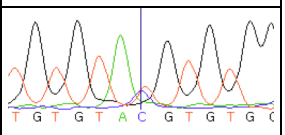
r^2 were based on the 1000 Genomes Project (Pilot 1, CEU). The data was plotted with LocusZoom v1.1 (http://csg.sph.umich.edu/locuszoom/). (**B**) Sites of open chromatin marked by FAIRE across the locus in the CHRF-288-11 (MK) and K562 (EB) cell lines (data uploaded as UCSC Genome Browser custom tracks). (**C**) Site of open chromatin found in MK but not EB cells harbouring the common variants rs342293 and rs342294. (**D**) *In silico* annotation of transcription factor binding sites at the open chromatin region described in (C). The predicted binding events are shown as transcription factor matrices (MatInspector v8.01). Note that some predicted binding sites are overlapping and not visible.

Next, I performed an *in silico* analysis of transcription factor binding sites at the 7q22.3 locus (**Appendix, Table 8-5**). Of the 34 SNPs in LD with rs342293, four (rs342240, rs342247, rs342292 and rs342293) disrupted an *in silico* predicted transcription factor binding site. However, only rs342293 was located within an experimentally verified site of open chromatin in MK cells. Among these four SNPs, rs342293 was the most strongly associated with mean platelet volume (Soranzo, Spector, et al., 2009).

The SNP rs342293 was located within overlapping predicted binding sites for the transcription factors BARX2, EVI1, GATA1, HHEX, HOXC8, HOXC9 and LBX2 (**Figure 5-1 D**). Based on the HaemAtlas data, of these seven transcription factors only EVI1, GATA1 and HHEX are expressed in megakaryocytes (**Figure 5-5 B**). The C>G substitution leads to disruption of the predicted binding sites of EVI1 and GATA1 (**Figure 5-2 A**). It is worth noting that *in silico* analysis predicted a RUNX1 transcription factor binding site only 5 bp apart from the EVI1-like binding site (discussed in **Section 5.3**).

To obtain the full spectrum of sequence variation at this region, we sequenced the NDR (chr7:106,159,393–106,159,887, build: hg18; 494 bp) in 643 healthy individuals of the Cambridge BioResource. No other common or low-frequency variants were detected, although a rare, possibly private, SNP was detected in one individual as a heterozygous position (chr7:106,159,601A>C) located in a polyA-region (**Table 5-1**).

Table 5-1. Resequencing of the MK-specific open chromatin region at chromosome 7q22.3. The low call rate for rs342294 was due to the location of the SNP at the very end of the sequence-tagged site (STS). MAFs from Phase II+III HapMap (CEU) and the 1000 Genomes Project (Pilot 1, CEU) were retrieved from dbSNP v131. Genomic coordinates were based on the human reference genome, build hg18.

	rs342293	private	rs342294
Calls	600/643 (93.3%)	633/643 (98.4%)	330/643 (51.3%)
Position	chr7:106,159,455	chr7:106,159,601	chr7:106,159,858
Genotype	C/G	A/C	T/C
MAF (observed)	0.4167	0.0008	0.3909
MAF (HapMap)	0.482	n/a	0.482
MAF (1K Genomes)	0.458	n/a	0.458
Sequencing traces			

Based on the above evidence and the absence of any additional common or low-frequency sequence variant at the MK-specific NDR, rs342293 remained the only likely putative functional candidate underlying the MPV association signal at the 7q22.3 locus. However, it cannot be excluded that additional functional variants may exist outside NDRs and exert their function through, for example, disruption of a yet to be identified transcription factor binding site or modulation of DNA methylation.

5.3. The alleles of rs342293 differentially bind the transcription factor EVI1

I then performed EMSAs in nuclear protein extracts from the megakaryocytic cell line CHRF-288-11. I observed a band shift with oligonucleotide probes harbouring rs342293 for both the ancestral allele rs342293-C and the alternative allele rs342293-G (**Figure 5-2 B**). However, the bands were unequally shifted suggesting differential protein binding properties at this position depending on the allele of rs342293 tested. The specific unlabelled competitors supported specificity of the retarded bands. In addition, the results suggested superior protein binding to the probe containing the C-allele. With supershift experiments using an EVI1 antibody, I confirmed binding of EVI1 to probes containing the ancestral C-allele, but not to those harbouring the alternative G-allele (**Figure 5-2 B**).

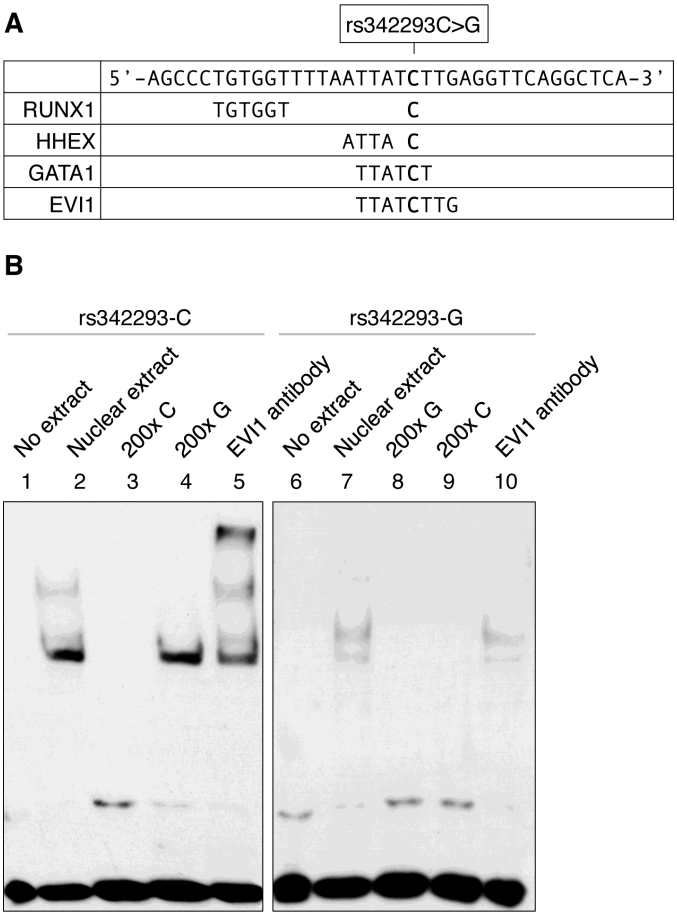


Figure 5-2. The effect of rs342293C>G on transcription factor binding. (**A**) Nucleotide sequence of the synthetic oligonucleotide probe harbouring rs342293 and *in silico* prediction of transcription factor binding sites. The major (C-) allele of rs342293 is shown in bold. The minor (G-) allele of rs342293 was predicted to disrupt the binding motifs of the transcription factors GATA1 (– strand) and EVI1 (–). RUNX1 (+) and HHEX (+) transcription factor binding sites were predicted to be located 9 bp and 1 bp apart from rs342293, respectively. (**B**) EMSAs in nuclear protein extracts from the megakaryocytic cell line CHRF-288-11 showed differential binding of the two alleles of rs342293. Competition reactions were performed with a 200-fold molar excess of unlabelled probes. Only the probe harbouring the reference allele was able to shift the protein-DNA complex when incubating with EVI1 antibodies.

Supershift experiments with a GATA1 antibody did not support *in vitro* binding of the GATA1 transcription factor to this site. I also confirmed RUNX1 transcription factor binding *in vitro* by demonstrating a supershift for both probes with a RUNX1 antibody (**Figure 5-3**).

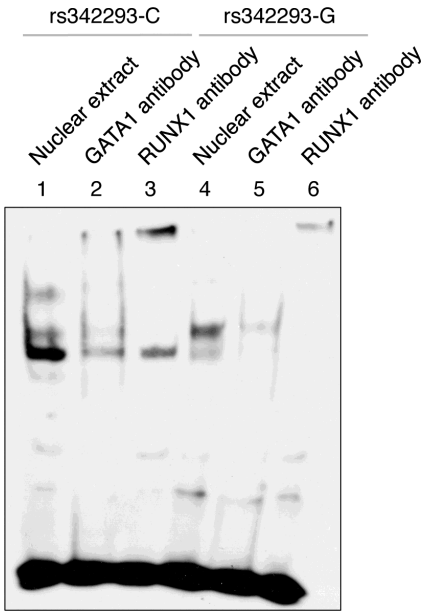


Figure 5-3. Gel shift assays in CHRF-288-11 nuclear protein extracts using GATA1 and RUNX1 antibodies. No supershift was observed for probes containing either rs342293-C or -G when incubating CHRF-288-11 nuclear protein extract with GATA1 antibodies. However, we showed evidence for RUNX1 transcription factor binding *in vitro*.

I further corroborated the *in silico* predictions and EMSA results by integrating data from genome-wide maps of transcription factor binding. Tijssen et al. performed chromatin immunoprecipitation combined with next-generation DNA sequencing (ChIP-seq) with GATA1 and RUNX1 antibodies in primary human megakaryocytes (Tijssen et al., 2011). No significant GATA1 but weak RUNX1 binding was observed at this locus (**Figure 5-4**).

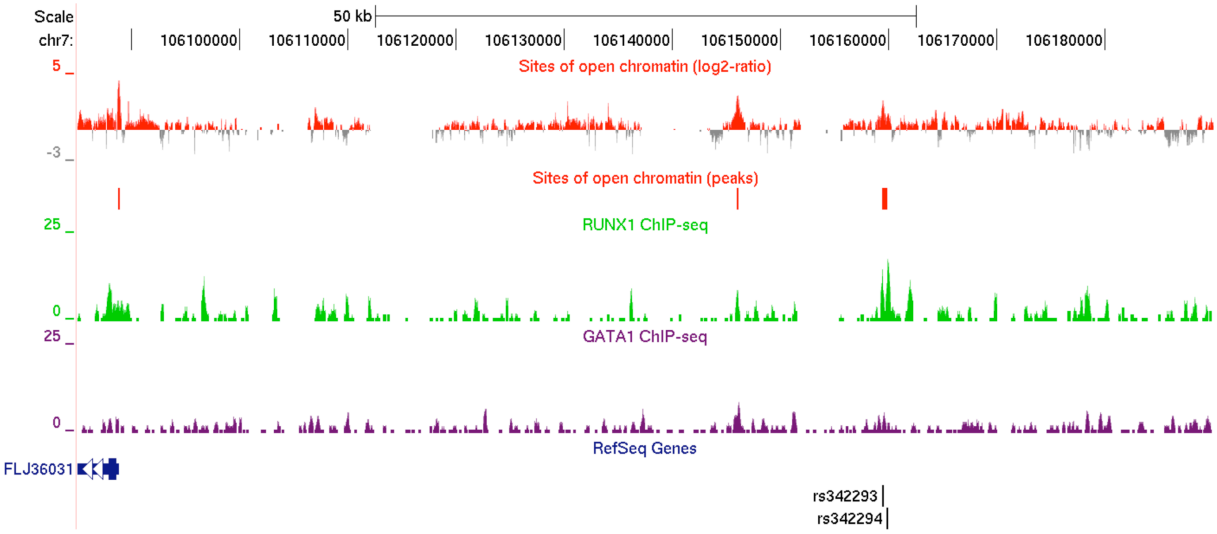


Figure 5-4. GATA1 and RUNX1 ChIP-seq profiles at the MK-specific open chromatin region at chromosome 7q22.3 in primary megakaryocytes. Data were transformed into density plots and displayed as UCSC Genome Browser custom tracks. Visual inspection of the 7q22.3 region showed no *in vivo* binding of GATA1, but weak RUNX1 binding. In total, the ChIP-seq data sets comprised 4,722 and 7,345 peaks for GATA1 and RUNX1, respectively (Tijssen et al., 2011).

5.4. The SNP rs342293 is associated with *PIK3CG* transcript levels in platelets and macrophages

A total of six protein-coding genes map within 1 Mb of the GWA index SNP rs342293, i.e. *COG5*, *FLJ36031* (also known as *CCDC71L*), *HBP1*, *NAMPT* (also known as *PBEF1*), *PIK3CG* and *PRKAR2B*. Whole-genome gene expression analysis on Illumina HumanWG-6 v1 Expression BeadChips revealed expression of all six genes in cord blood-derived megakaryocytes (**Figure 5-5 A**; Soranzo, Rendon, et al., 2009; Watkins et al., 2009). However, only *PIK3CG* and *PRKAR2B* are transcribed in platelets (Soranzo, Rendon, et al., 2009).

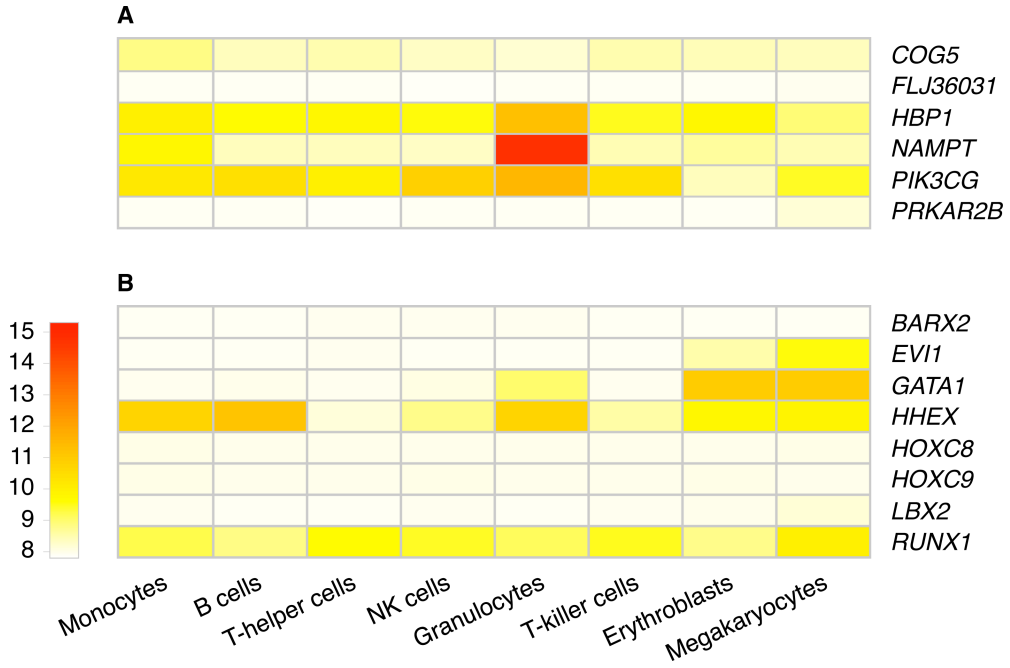


Figure 5-5. Gene expression profiles in differentiated human blood cells. Based on the HaemAtlas data (Watkins et al., 2009), the heat map shows normalised gene expression profiles of (**A**) genes within a 1 Mb interval of rs342293, and (**B**) genes encoding transcription factors predicted to bind DNA sequence motifs around rs342293.

Soranzo, Rendon, et al. investigated the association of rs342293 with transcript levels of both *PIK3CG* and *PRKAR2B* in platelets, and reported a weak eQTL association with *PIK3CG* transcript levels (permutation P=0.047) but not *PRKAR2B* (Soranzo, Rendon, et al., 2009). We replicated this eQTL association in an independent sample cohort of 24 healthy individuals (**Section 2.12**), obtaining the same genotypic effect for rs342293 (P=0.0542; **Figure 5-6 A**).

Next, we assessed the *PIK3CG*-eQTL in three different types of white blood cells, i.e. macrophages, monocytes and B cells (lymphoblastoid cell line, LCLs), as well as in different tissues, i.e. adipose (subcutaneous fat) and skin (**Appendix, Table 8-6**). We found an association between rs342293 and *PIK3CG* transcript abundance in macrophages (P=0.0018; **Figure 5-6 B**), but not in monocytes and LCLs. Further, we did not observe an association in adipose and skin tissues. Our data strengthened the previous evidence of rs342293 modulating *PIK3CG* transcript levels in platelets and showed that this eQTL was also present in macrophages. The association with *PIK3CG* transcript levels in macrophages but not monocytes suggested that the transcriptional effects may occur during late events of cellular differentiation (**Section 1.9**).

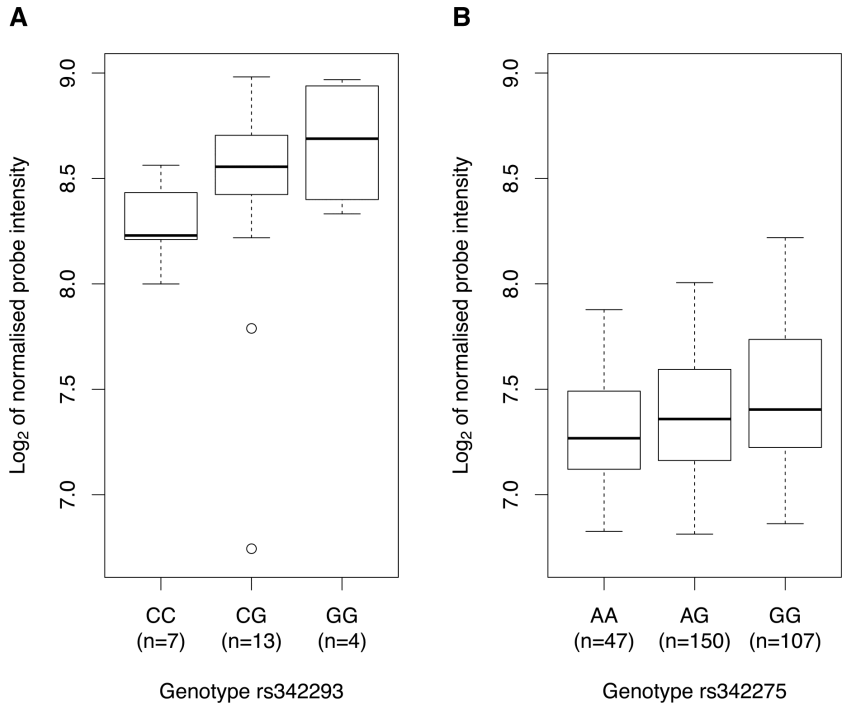


Figure 5-6. Association of rs342293 genotypes with *PIK3CG* transcript levels in (A) platelets and (B) macrophages. In platelets and macrophages, we observed a genotypic effect on *PIK3CG* transcript levels (P=0.0542 and P=0.0018, respectively). The associations are presented as box-and-whisker plots. In macrophages, the proxy SNP rs342275A>G (r^2=0.94, Phase II HapMap, CEU) was used for the eQTL analysis.

5.5. Gene expression profiles in whole blood of *Pik3cg*$^{-/-}$ mice

To further our understanding of the role of PIK3CG in platelets, we performed whole-genome gene expression profiling in whole blood of *Pik3cg* knockout mice (**Section 2.13**). Due to the need for relatively large quantities of RNA for whole-genome gene expression analysis on microarrays, we were

only able to assay RNA extracted from whole blood as opposed to platelets. I identified 220 differentially expressed genes between knockout (n=3) and wild type mice (n=3) with a fold-change of at least ±1.5. Functional ontology classification of these genes (**Appendix, Table 8-7**) revealed enrichment for 'regulation of biological characteristic', e.g. cell size and volume (GO term: 0065008; P=2.97x10^{-14}), and 'blood coagulation' (GO term: 0007596; P=7.87x10^{-12}). Notably, the 220 differentially expressed genes included *Gp1bb* (fold-change of -2.203 in *Pik3cg*$^{-/-}$ compared to wild type mice), *Gp5* (-3.211), *Gp9* (-0.996), *Gp6* (-0.711) and *Vwf* (-1.381). These five genes are transcribed in the megakaryocytic lineage in humans, based on the HaemAtlas data. The platelet glycoproteins GP1BB, GP5 and GP9, together with GP1BA, constitute the platelet membrane receptor for the plasma protein Von Willebrand Factor (VWF), which is encoded by *VWF* (Roth, 1991).

5.6. Canonical pathway enrichment analysis and protein-protein interaction network

To explore the signalling pathways of PIK3CG in humans, I analysed 191 human orthologs of the 220 differentially expressed genes between *Pik3cg*$^{-/-}$ and wild type mice. Canonical pathway enrichment analysis based on the curated gene sets of the Molecular Signatures Database (MSigDB) v3.6 (Subramanian et al., 2005) revealed that the top six enriched gene sets are related to platelets: 'platelet degranulation' (P=3.44x10^{-8}), 'platelet activation' (P=5.05x10^{-8}), 'formation of platelet plug' (P=2.85x10^{-7}), 'haemostasis' (P=7.48x10^{-6}), 'formation of fibrin clot and clotting cascade' (P=1.39x10^{-5}) and 'platelet adhesion to exposed collagen' (P=1.86x10^{-5}).

We constructed a protein-protein interaction network centred on the proteins encoded by the 191 transcripts described above. First-order interactors of these 'core' proteins were obtained from Reactome, an open-source, manually curated database of human biological pathways (Matthews et al., 2009; Croft et al., 2011). We filtered interactors on their expression levels in megakaryocytes (**Section 2.14**). The resulting network incorporated 45 core proteins centred on PIK3CG consisting of 642 nodes and 1067 edges (**Figure 5-7**).

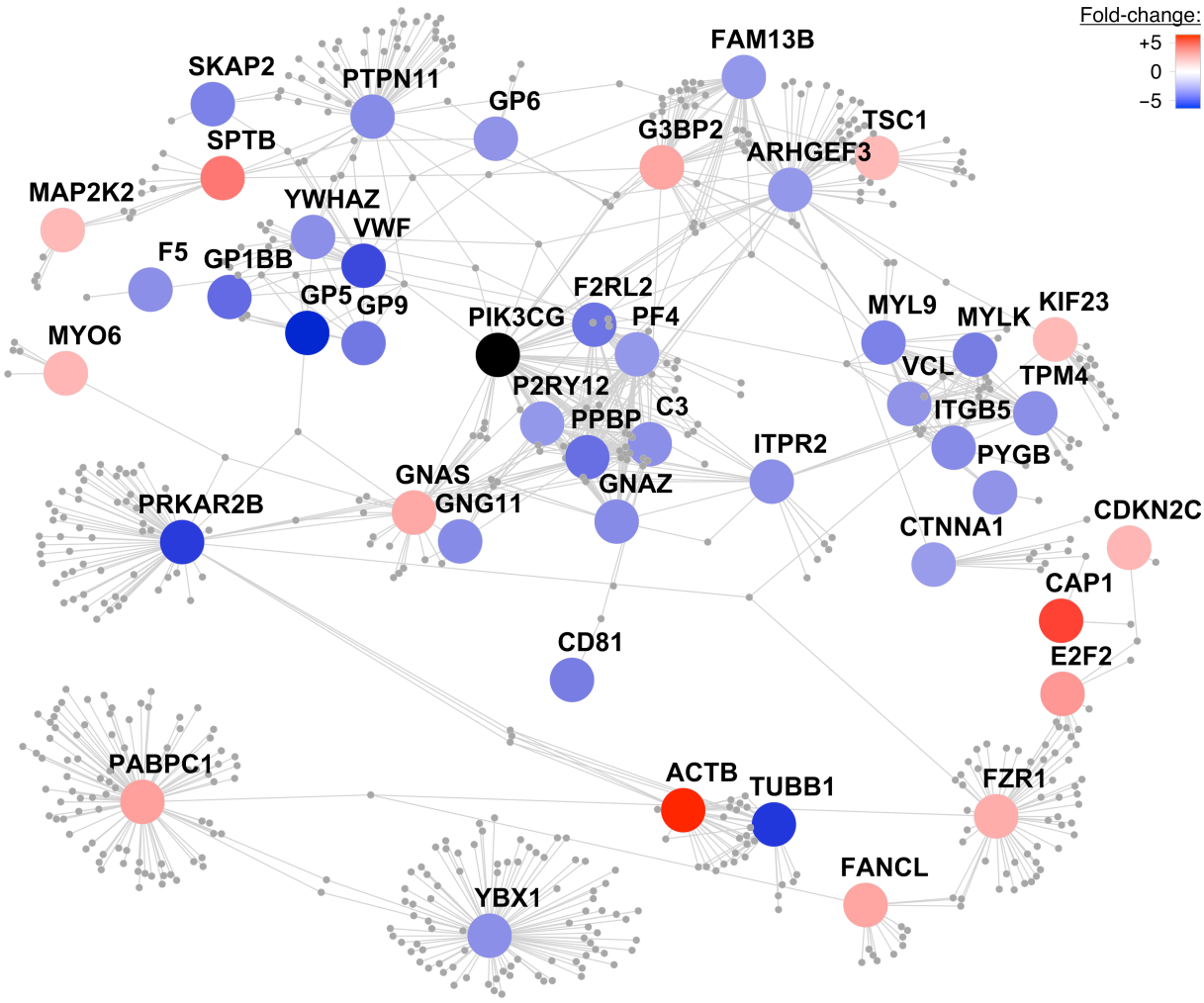


Figure 5-7. Protein-protein interaction network centred on PIK3CG. A protein-protein interaction network was constructed centring on the human orthologous proteins encoded by the 220 differentially expressed transcripts between *Pik3cg* knockout and wild type mice with a fold-change of at least ±1.5. The colour of these 'core' proteins represents the fold-change of gene transcript levels between *Pik3cg*$^{-/-}$ and wild type mice on a continuous scale from over- (red) to underexpression (blue) of transcripts in knockout mice. Core proteins, which did not exhibit first-order interactors in the Reactome database and were disconnected from the largest connected component, are not shown. First-order interactors of PIK3CG that form network nodes and edges are shown in grey and were obtained from Reactome. Interactors that are not expressed in megakaryocytes based on the HaemAtlas data were omitted. The resulting network consisted of 45 core proteins with 642 nodes and 1067 edges.

5.7. Discussion

We elucidated the molecular basis of the association between rs342293 and platelet volume and function at chromosome 7q22.3. We identified an NDR in MK but not EB cells containing the index

SNP rs342293 for this association. Resequencing of this NDR in 643 individuals provided strong evidence that rs342293 is the only putative causative variant in this region. However, resources such as the completed 1000 Genomes Project (The 1000 Genomes Project Consortium, 2010) will make the requirement for resequencing any identified NDR redundant with respect to identifying common variants (**Chapter 4**). I then demonstrated that the alleles of rs342293 differentially bind the transcription factor EVI1.

At the time, rs342293 was the first functional variant to be elucidated among the known platelet QTLs (Paul et al., 2011). Our data suggested a molecular mechanism by which a non-coding GWA index SNP modulates platelet phenotype. We note that our results implicate the original GWA tag SNP as the likely functional variant. Indeed, two studies identified the GWA index SNP for colorectal cancer at chromosome 8q24 to be a regulatory variant, and provided a mechanistic explanation for its association with disease risk (Pomerantz et al., 2009; Tuupanen et al., 2009).

Expression QTL data in platelets and macrophages provided statistical support that rs342293, or a variant in LD, affect *PIK3CG* gene expression levels. *PIK3CG* is transcribed in megakaryocytes but only weakly expressed in erythroblasts (Watkins et al., 2009; **Figure 5-5 A**), which is in agreement with the MK-specific properties of the identified NDR. However, additional work is required to scrutinise all possible targets of this regulatory module, in particular non-protein coding transcripts. *PIK3CG* is located 134 kb downstream of rs342293 and encodes the phosphoinositide-3-kinase γ-catalytic subunit. The lipid kinase PIK3CG (PI3Kγ) is a member of the class I PI3Ks and catalyses the conversion of phosphatidylinositol-4,5-bisphosphate (PtdIns(4,5)P2; PIP2) to phosphatidylinositol-3,4,5-trisphosphate (PtdIns(3,4,5)P3; PIP3) downstream of cell surface receptor activation (Wymann et al., 2003; Rückle et al., 2006; Hawkins & Stephens, 2007). In megakaryocytes and platelets, PIP3 is crucial in the collagen-induced regulation of phospholipase C and initiation of megakaryopoiesis and pro-platelet formation (Pasquet et al., 2000). Functional studies with *Pik3cg* knockout mice previously indicated a role in wound healing (Zhao et al., 2006), ADP-induced platelet aggregation and thrombosis (Hirsch et al., 2001; Schoenwaelder et al., 2007). It is also worth noting that PIK3CG has a prominent role in macrophage activation (Hawkins & Stephens, 2007).

The whole-genome gene expression profiling of *Pik3cg*$^{-/-}$ mice performed in this study showed differential expression of genes involved in important platelet-related pathways compared to wild type mice, most notably *Vwf* and its platelet membrane receptor components. A recent meta-analysis of GWA studies in 68,000 Northern Europeans showed that common sequence variants at the *VWF* and *GP1BA* loci exert an effect on platelet volume and count, respectively (Gieger et al., 2011). This recent

finding, together with the established knowledge that Mendelian mutations in the genes encoding the platelet VWF (Von Willebrand disease, type 2; OMIM: 613554) and its receptor (Bernard-Soulier syndrome; OMIM: 231200) cause giant platelets, give biological credence to the effects we observed of *Pik3cg* knockout on the transcription of these platelet genes. The constructed protein-protein interaction network, which was centred on PIK3CG, highlighted additional proteins implicated in severe platelet disorders, including TUBB1 (macrothrombocytopenia, autosomal dominant, TUBB1-related; OMIM: 613112), F5 (factor V deficiency; OMIM: 612309) and P2RY12 (bleeding disorder due to P2RY12 defect; OMIM: 609821). Therefore, it is plausible to assume that differences in the *PIK3CG* transcript levels in humans, based on the different alleles of rs342293, may lead to changes in the abundance of platelet membrane proteins that are key regulators of platelet formation.

Soranzo, Rendon, et al. previously showed an association of rs342293-G with decreased platelet reactivity in humans (Soranzo, Rendon, et al., 2009), assessed as the proportion of binding to annexin V and fibrinogen, as well as P-selectin expression, after activation of platelets with collagen-related peptide (CRP-XL). We also observed that rs342293 is likely to modify events downstream of signalling via the collagen signalling receptor glycoprotein VI, which is encoded by *GP6* (Jones et al., 2009). A recent GWA study showed the same locus at chromosome 7q22.3 to also be associated with epinephrine-induced platelet aggregation (Johnson et al., 2010). The index SNP in that study, rs342286, and rs342293 are in high LD (r^2=0.87, Phase II HapMap, CEU) making rs342293 the putative causative variant underlying both functional associations. PIP3 is required for the platelet signalling cascades triggered via an immunoreceptor tyrosine-based activation motif on the Fc receptor γ-chain (ITAM-FcR-γ) for collagen and G-protein-coupled receptor for epinephrine. Therefore, our observations are compatible with the notion that these platelet functional events are modified by differences in the *PIK3CG* transcript level and that silencing of *Pik3cg* in mice reduces the expression of the *Gp6* gene.

EVI1 (ecotropic viral integration site-1) encodes a protein with two zinc-finger domains (ZF1 and ZF2), which feature distinct DNA-binding specificities (Bartholomew et al., 1997). EVI1 mainly promotes haematopoietic differentiation into the megakaryocytic lineage (Shimizu et al., 2002). EVI1 was frequently reported to be a repressor of transcription that has the potential to recruit diverse regulatory proteins. For example, EVI1 antagonises the growth-inhibitory effect of transforming growth factor-β (TGF-β), a potent regulator of megakaryopoiesis (Sakamaki et al., 1999), by interacting with SMAD3 via ZF1, and inhibiting SMAD3 from binding to DNA (Kurokawa, Mitani, Irie, et al., 1998; Kurokawa, Mitani, Imai, et al., 1998). EVI1 contains domains that interact with RUNX1 (Runt-related transcription factor 1), which is the α-subunit of the transcription factor CBF (core binding factor). The interaction

of EVI1 with the DNA-binding domain Runt of *RUNX1* leads to the destabilisation of the DNA-RUNX1 complex and subsequent loss of RUNX1 function (Senyuk et al., 2007).

Based on the above data, I propose a model (**Figure 5-8**) in which the DNA sequence containing the C- but not the G-allele of rs342293 binds the transcription factor EVI1. Together with the transcription factor RUNX1, EVI1 may act as transcriptional repressor of *PIK3CG* in megakaryocytes. Individuals with homozygous rs342293-C have lower *PIK3CG* gene expression levels in platelets compared to subjects with homozygous rs342293-G. In *Pik3cg* knockout mouse models, transcripts that encode key proteins for platelet membrane biogenesis were found to be downregulated, which may ultimately affect the platelet phenotype.

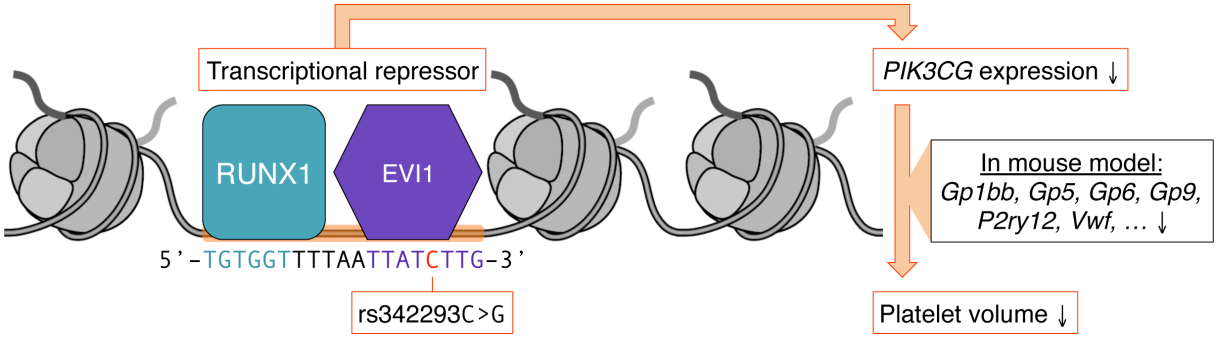


Figure 5-8. Model of the mechanism by which rs342293 may affect platelet volume. Based on our data, we propose that rs342293C>G affects EVI1 transcription factor binding at the megakaryocyte-specific site of open chromatin at chromosome 7q22.3. The underlying DNA sequence around rs342293 (indicated in red) is shown and transcription factor binding motifs for RUNX1 (turquoise) and EVI1 (purple) are highlighted. The transcriptional complex of EVI1 is a well-described repressor of gene transcription.