

# A novel variant on chromosome 7q22.3 associated with mean platelet volume, counts, and function

\*Nicole Soranzo,<sup>1,2</sup> \*Augusto Rendon,<sup>3</sup> \*Christian Gieger,<sup>4</sup> Chris I. Jones,<sup>5</sup> Nicholas A. Watkins,<sup>3</sup> Stephan Menzel,<sup>6</sup> Angela Döring,<sup>4</sup> Jonathan Stephens,<sup>3</sup> Holger Prokisch,<sup>7,8</sup> Wendy Erber,<sup>3,9</sup> Simon C. Potter,<sup>1</sup> Sarah L. Bray,<sup>10</sup> Philippa Burns,<sup>3</sup> Jennifer Jolley,<sup>3</sup> Mario Falchi,<sup>2,11</sup> Brigitte Kühnel,<sup>4</sup> Jeanette Erdmann,<sup>12</sup> Heribert Schunkert,<sup>12</sup> Nilesh J. Samani,<sup>13</sup> Thomas Illig,<sup>4</sup> Stephen F. Garner,<sup>3</sup> Angela Rankin,<sup>3</sup> Christa Meisinger,<sup>4</sup> John R. Bradley,<sup>14</sup> Swee Lay Thein,<sup>6</sup> Alison H. Goodall,<sup>13</sup> †Tim D. Spector,<sup>2</sup> †Panos Deloukas,<sup>1</sup> and †Willem H. Ouwehand<sup>1,3</sup>

<sup>1</sup>Human Genetics Department, Wellcome Trust Sanger Institute, Hinxton, Cambridge, United Kingdom; <sup>2</sup>Department of Twin Research and Genetic Epidemiology, King's College London, London, United Kingdom; <sup>3</sup>Department of Haematology, University of Cambridge and National Health Service (NHS) Blood and Transplant, Cambridge, United Kingdom; <sup>4</sup>Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany; <sup>5</sup>School of Biological Sciences, University of Reading, Reading, United Kingdom; <sup>6</sup>Molecular Haematology, King's College London, London, United Kingdom; <sup>7</sup>Institute of Human Genetics, Technical University Munich, Munich, Germany; <sup>8</sup>Institute of Human Genetics, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany; <sup>9</sup>Haematology Department, Cambridge University Hospitals NHS Foundation Trust, Cambridge, United Kingdom; <sup>10</sup>Medical Research Council (MRC) Biostatistics Unit, University Forvie Site, Cambridge, United Kingdom; <sup>11</sup>Genomic Medicine, Imperial College London, London, United Kingdom; <sup>12</sup>Medizinische Klinik II, Universität zu Lübeck, Lübeck, Germany; <sup>13</sup>Cardiovascular Sciences, University of Leicester, Leicester, United Kingdom; and <sup>14</sup>Department of Medicine, University of Cambridge School of Clinical Medicine, Addenbrooke's Hospital, Cambridge, United Kingdom

**Mean platelet volume (MPV) and platelet count (PLT) are highly heritable and tightly regulated traits. We performed a genome-wide association study for MPV and identified one SNP, rs342293, as having highly significant and reproducible association with MPV (per-G allele effect  $0.016 \pm 0.001$  log fL;  $P < 1.08 \times 10^{-24}$ ) and PLT (per-G effect  $-4.55 \pm 0.80$   $10^9/L$ ;  $P < 7.19 \times 10^{-8}$ ) in 8586 healthy subjects. Whole-genome expression analysis in the 1-MB region**

**showed a significant association with platelet transcript levels for *PIK3CG* ( $n = 35$ ;  $P = .047$ ). The G allele at rs342293 was also associated with decreased binding of annexin V to platelets activated with collagen-related peptide ( $n = 84$ ;  $P = .003$ ). The region 7q22.3 identifies the first QTL influencing platelet volume, counts, and function in healthy subjects. Notably, the association signal maps to a chromosome region implicated in my-**

**eloid malignancies, indicating this site as an important regulatory site for hemato-poiesis. The identification of loci regulating MPV by this and other studies will increase our insight in the processes of megakaryopoiesis and proplatelet formation, and it may aid the identification of genes that are somatically mutated in essential thrombocytosis. (Blood. 2009; 113:3831-3837)**

## Introduction

Platelets are anucleate blood cell fragments that play a key role in maintaining primary hemostasis and in wound healing. Mean platelet volume (MPV) and platelet count (PLT) are tightly regulated and inversely correlated in the healthy population. Increased MPV represents a strong, independent predictor of postevent outcome in coronary disease and myocardial infarction,<sup>1,2</sup> and changes of either parameter outside the normal ranges are routinely used to ascertain and manage a large number of clinical conditions.

Studies in rodents, primates, and twins have confirmed that blood cell quantitative traits, such as MPV and PLT, have high heritability levels.<sup>3-5</sup> Genome-wide association studies that used dense genotyping in thousands of subjects have greatly improved the resolution of such complex polygenic traits in humans.<sup>6,7</sup> We therefore reasoned that it should be possible to identify novel quantitative trait loci (QTLs) for MPV and PLT by a similar approach. The identification of such loci will provide new insights

in the complex cellular processes of megakaryopoiesis and platelet formation. In addition, it may contribute to our understanding of premalignant conditions such as polycythemia vera and particularly essential thrombocytosis (ET), in which somatically acquired mutations in the *JAK2* gene explain only a fraction of cases.<sup>8</sup>

Megakaryocytes (MKs), the platelet precursor cells, originate from pluripotent hematopoietic stem cells (HSCs) in a stepwise process of fate determination and proliferation controlled by the cytokine thrombopoietin<sup>9</sup> and several extracellular matrix proteins (ECMPs).<sup>10-12</sup> After migration of the MK progenitors from the HSC niche to the bone marrow, platelets are formed via the tightly regulated process of proplatelet formation (PPF) from mature polyploid MKs.<sup>13</sup> Transendothelial proplatelet ends are exposed to shear, triggering the final step of this process.<sup>14</sup> The molecular machineries underlying megakaryopoiesis and PPF have been studied in some detail, but many aspects are ill understood.<sup>15</sup> Most information has been obtained from studies on the genetic basis of

Submitted October 23, 2008; accepted December 24, 2008. Prepublished online as *Blood* First Edition paper, February 12, 2009; DOI 10.1182/blood-2008-10-184234.

\*N.S., A. Rendon, and C.G. contributed equally to this study.

†T.D.S., P.D., and W.H.O. contributed equally to this study.

The online version of this article contains a data supplement.

The publication costs of this article were defrayed in part by page charge payment. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 USC section 1734.

© 2009 by The American Society of Hematology

rare inherited syndromes of abnormal megakaryopoiesis and platelet formation.<sup>16,17</sup> Further insight has been gained from gene knockout experiments in mice and RNAi gene silencing in human HSCs and MKs, which have shown the crucial role of the gene *MEIS1*,<sup>18</sup> microRNAs (miR-150 and miR-146a),<sup>19,20</sup> and *NFE21* in fate determination and PPF.

To date, common genetic variants in humans controlling MPV have not been identified. Here, we report on the first genetic locus on chromosome 7q22.3, associated with MPV and, intriguingly, also with differences in platelet function. Notably, the 65-kilobase (kb) recombination interval harboring the association signal is located in a chromosomal region frequently rearranged in myeloid malignancies.<sup>22</sup>

## Methods

### Population samples

The discovery sample consisted of 1221 healthy donors from the UK Blood Services Common Control (UKBS-CC1) genotyped with the Affymetrix (Santa Clara, CA) 500K Array.<sup>23</sup> Four replication samples were also assessed, including 1050 subjects from the TwinsUK adult twin registry (www.TwinsUK.ac.uk), genotyped with the HumanHap300 chip (Illumina, San Diego, CA) and imputed using genotype data from the HapMap project<sup>24</sup>; 1601 subjects from the Kooperative Gesundheitsforschung in der Region Augsburg (KORA) cohort from the region of Augsburg in Germany, genotyped with the same Affymetrix 500K Array<sup>25</sup>; a second panel of 1304 subjects from the UKBS-CC collection (UKBS-CC2); and 3410 subjects from the Cambridge BioResource (CBR). Characteristics of the sample collections and genotyping are given in Document S1 (available on the *Blood* website; see the Supplemental Materials link at the top of the online article). Samples and phenotype information were obtained from all study subjects with written and signed consent in accordance with the Declaration of Helsinki. Ethics committee approval was obtained for each of the study cohorts according to the national rules and regulations that are in place in the different member states of the European Union.

### Statistical analyses

Associations of blood traits with genotypes were performed with linear regression on natural log-transformed (MPV, red blood cell count [RBC], and white blood cell count [WBC]) or untransformed (MCH, MCHC, PLT, MCV, HB) variables after adjustment for age, sex, and instrument (TwinsUK). Analyses were performed with the software PLINK<sup>26</sup> (UKBS-CC1, UKBS-CC2, CBR) or MACH2QTL (KORA<sup>27</sup>). Association analyses in the TwinsUK collection were performed with a score test and variance components to adjust for family structure with the use of the software MERLIN.<sup>28</sup> Meta-analysis statistics were obtained with a weighted  $z$  statistics method, where weights were proportional to the square root of the number of subjects examined in each sample and selected such as the squared weights sum to be 1.<sup>29</sup> Calculations were implemented in the METAL package.<sup>27</sup> Calculations of effect sizes for quantitative traits were performed with the inverse variance method and custom scripts in the R environment (www.r-project.org).

### Gene expression analysis

For this study, we focused on the 1-MB region centered on rs342293, which contains 6 annotated genes. For platelets, mRNA from 35 subjects spanning the range of platelet reactivities in the Platelet Function Cohort of 500 healthy subjects and with known rs342293 genotype were assessed with the Illumina HumanWG-6 v1 Expression BeadChip arrays (Illumina) as described previously.<sup>30</sup> In short, leukocyte-depleted platelet concentrates were obtained by apheresis from the 35 donors, and EDTA (pH 8.0, final concentration 5 mM) was added, and residual leukocytes were removed by centrifugation at 140g for 15 minutes. An aliquot of the resulting platelet-rich plasma (PRP) was used for leukocyte counting by flow cytometry

(LeukoCOUNT; Becton Dickinson, Oxford, United Kingdom), and the remaining PRP was centrifuged at 1500g for 10 minutes to pellet the platelets. The supernatant was then removed, and the platelets were lysed in TRIzol (10 mL per  $4 \times 10^{10}$  platelets; Invitrogen, Paisley, United Kingdom), and RNA was isolated following the manufacturers' instructions. RNA quality, purity, and quantity were assessed by nonreducing agarose gel electrophoresis and duplex reverse-transcription-polymerase chain reaction (RT-PCR) to determine the levels of the pan-leukocyte marker CD45 (*PTPRC*) as described previously.<sup>31</sup> Before microarray analysis, total RNA was further purified with the RNeasy MinElute Cleanup kit (QIAGEN, Crawley, United Kingdom).

Erythroblast (EB) and MK transcriptome data were obtained from an earlier study.<sup>32</sup> In short, paired samples of both types of precursors were obtained by culturing of CD34<sup>+</sup> HSCs purified from cord blood. After approximately 8 days of culture in the presence of specific growth factors, EBs and MKs underwent fluorescence-activated cell sorting (FACS) with monoclonal antibodies against glycophorin A (anti-CD235a) and the integrin  $\alpha$ Ib (anti-CD41), respectively. Purified RNA was amplified and labeled with the Illumina TotalPrep RNA Amplification Kit (Applied Biosystems, Austin, TX). Biotinylated cRNA was applied to Illumina HumanWG-6 v1 Expression BeadChips and hybridized overnight at 58°C. Chips were washed, detected, and scanned according to the manufacturer's instructions.

The corresponding probes for the 6 genes in the 1-MB interval centered on rs342293 are *PBEF1* (GI\_5031976-I), *FLJ36031* (GI\_31343528-S), *PIK3CG* (GI\_21237724-S), *PRKAR2B* (GI\_4506064-S), *HBP1* (GI\_21361410-S), and *COG5* (GI\_32481217-A). Array processing was performed under the R/BioConductor environment (www.bioconductor.org) with the "lumi" package.<sup>33</sup> Probe intensities were first transformed to stabilize the variance across the intensity range with the VST transform from lumi.<sup>33</sup> This transformation makes use of the multiple measurements per probe particular to the Illumina platform. It approaches a log<sub>2</sub> transformation at high intensities but performs better at lower signal intensities. Subsequently, the transformed signals were median-normalized. The probe detection level was assessed with the use of Illumina's detection  $P$  value, where at least 18 of the 35 samples were required to have a detection of  $P$  less than .01. This level roughly corresponds to a normalized signal intensity of 7.7. Tests of association with rs342293 genotype were performed with linear regression and permutation testing ( $n = 10\,000$ ) with the software PLINK.<sup>26</sup>

### Bioinformatics prediction of altered transcription factor binding sites

The search for potentially altered transcription factor binding sites was performed as follows. The 4-bp regions surrounding each allele of rs342293 (20-bp extension in each direction) were searched for transcription factor binding sites present in TRANSFAC 7.0 public (http://www.gene-regulation.com/) with MotifScanner<sup>34</sup> with a prior of 0.1 and a background model EPD-3 (third-order Markov chain based on eukaryotic promoter sequences). Motifs that were called for only one of the alleles were selected. This list was further reduced to 2 elements (*EVII* and *GATA*) by only considering motifs for which the putative transcription factors were detected above background levels in the genome-wide expression profiles of MKs obtained as detailed in the previous paragraph.

### Measurement of platelet response

Platelet reactivity was analyzed by whole-blood flow cytometry in a panel of 500 healthy control donors from the Platelet Function Cohort (PFC) as described previously.<sup>35</sup> Briefly, healthy regular blood donor subjects were recruited from the NHS Blood and Transplant clinic in Cambridge. Blood was processed within 10 minutes of venesection, and 5  $\mu$ L citrated whole blood was added to 50  $\mu$ L HBS containing aspirin (100  $\mu$ M), hirudin (10 U mL<sup>-1</sup>), and, for cross-linked collagen-related-peptide (CRP-XL)-stimulated samples, apyrase (4 U mL<sup>-1</sup>). Samples were coincubated for 20 minutes at room temperature with either ADP ( $10^{-8}$ ,  $10^{-7}$ ,  $10^{-6}$  M) or CRP-XL (0.01, 0.1, 1  $\mu$ g mL<sup>-1</sup>), and with FITC-antifibrinogen, PE-anti-CD62P, or FITC-labeled annexin V. The reaction was stopped by hundred-fold dilution in formyl saline (0.2% formaldehyde in 0.9% NaCl). Negative

**Table 1. Association of rs342293 with mean platelet volume in the 5 studies and meta-analysis summary statistics for the blood traits investigated**

	Subjects, n	G allele frequency	Mean $\pm$ SD level of MPV, fL			Per-G allele effect (SE), log (fL)*	P
			GG	GC	CC		
<b>Cohort</b>							
UKBS-CC1	1221	0.46	9.62 $\pm$ 1.02	9.50 $\pm$ 1.03	9.21 $\pm$ 0.88	0.022 (0.004)	2.21 $\times 10^{-7}$
TwinsUK	1050	0.46	11.29 $\pm$ 1.10	11.06 $\pm$ 1.06	10.94 $\pm$ 0.93	0.014 (0.004)	3.08 $\times 10^{-4}$
KORA F3 500K	1601	0.45	8.81 $\pm$ 1.02	8.71 $\pm$ 0.91	8.56 $\pm$ 0.90	0.014 (0.004)	1.6 $\times 10^{-4}$
CBR	3410	0.45	9.63 $\pm$ 0.93	9.02 $\pm$ 0.91	8.93 $\pm$ 0.91	0.013 (0.002)	2.03 $\times 10^{-7}$
UKBS-CC2	1304	0.46	9.76 $\pm$ 1.09	9.39 $\pm$ 1.01	9.30 $\pm$ 1.02	0.023 (0.004)	3.29 $\times 10^{-8}$
Total	8586	—	—	—	—	0.016 (0.001)	1.08 $\times 10^{-24}$
<b>Meta-analysis of other blood cell indices</b>							
Platelet count, $\times 10^9/L$	—	—	—	—	—	-4.553 (0.802)	7.19 $\times 10^{-8}$
Hemoglobin concentration, g/dL	—	—	—	—	—	-0.020 (0.013)	.119
RBC count, log $10^{12}/L$	—	—	—	—	—	-0.002 (0.001)	.070
WBC count, log $10^9/L$	—	—	—	—	—	-0.001 (0.004)	.730
MCH, pg	—	—	—	—	—	0.018 (0.025)	.509
MCHC, g/dL	—	—	—	—	—	-0.006 (0.011)	.516
MCV, fL	—	—	—	—	—	0.087 (0.064)	.163

MPV indicates mean platelet volume; SD, standard deviation; UKBS-CC, UK Blood Services Common Controls; KORA, Kooperative Gesundheitsforschung in der Region Augsburg; CBR, Cambridge BioResource; RBC, red blood cell; WBC, white blood cell; MCH, mean corpuscular hemoglobin; MCHC, mean corpuscular hemoglobin concentration; and MCV, mean RBC volume.

\*After adjustment for age and sex, the proportion of total variation in MPV explained by rs342293 is UKBS-CC1, 2.25%; UKBS-CC2, 2.33%; TwinsUK, 1.27%; CBR, 0.79%; and KORA F3, 0.89%.

controls for the P-selectin antibodies were set with an appropriate isotype control and for the antifibrinogen antibody with samples incubated with the antibody in the presence of 6 mM EDTA. Platelets were then washed, incubated with FITC-labeled secondary antibody for 30 minutes, washed again, and resuspended in PBS containing 0.32% trisodium citrate and 0.25% bovine serum albumin (BSA). Flow cytometric analysis was performed on a Beckman Coulter EPICS Profile XL (Beckman Coulter, High Wycombe, United Kingdom). The percentage of positive cells (annexin V and fibrinogen binding) and median fluorescence intensity (MFI; P-selectin expression) were recorded. Measurements were then adjusted for confounders by logit transformation and by fitting a multiple linear regression model to the response, with sample processing time as a continuous predictor and type of donor as a categorical predictor.

## Results

### Genome-wide associations with MPV

We performed a genome-wide association analysis for MPV at 361 352 SNPs profiled with the Affymetrix Gene Chip 500K mapping array set<sup>36</sup> in 1221 healthy donors from the UKBS-CC1 collection with available phenotype.<sup>23</sup> After applying stringent QC criteria, only one genomic region displayed association with  $P$  less than  $10^{-5}$  at multiple SNPs (rs342214, rs342242, rs342275, rs342293;  $P = 1.32 \times 10^{-6}$ - $2.21 \times 10^{-7}$ ). Because the 4 SNPs were highly correlated ( $r^2 = 0.94$  in the HapMap CEU sample), we selected the SNP with the strongest evidence for association, rs342293, for replication in additional population samples. These included 2 of our population-based cohorts with available genome-wide typing results for MPV, namely (1) 1050 subjects from the TwinsUK adult twin registry, genotyped with the HumanHap300 chip (Illumina) and imputed with genotype data from the HapMap project<sup>24</sup>; (2) 1601 subjects from the KORA cohort from the region of Augsburg in Germany, genotyped with the same Affymetrix 500K array.<sup>25</sup> In addition, we genotyped rs342293 in 2 further collections: a second panel of 1304 subjects from the UKBS-CC2 and 3410 subjects from the CBR. The characteristics of the collections are described in the Document S1. The association of rs342293 with MPV was strongly replicated in all the collections tested, with

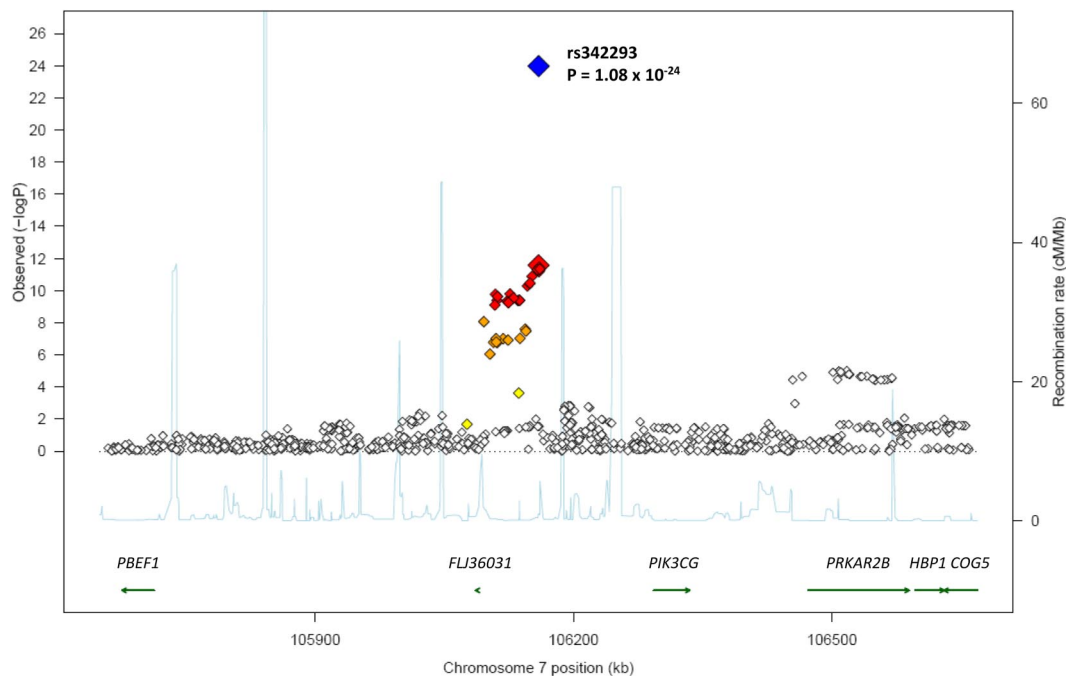
a  $P$  value of  $1.08 \times 10^{-24}$  in the combined sample of 8586 subjects (Table 1). All collections displayed associations of similar magnitude, with a total increase of 0.016 log fL (95% CI, 0.013-0.018) per copy of the minor G allele. On average, genetic variation at this locus explains 1.5% (range, 0.78%-2.45%) of total variance in MPV in a model adjusted for sex and age.

PLT and MPV are known to be negatively correlated, and this was replicated in all the sample collections investigated (average correlation coefficient  $r = -0.39$ ). When tested, we found that the minor (G) allele at rs342293 was also significantly associated with reduced PLT ( $P = 7.19 \times 10^{-8}$ ; per G-allele effect,  $-4.55 \text{ } 10^9/L$ ; 95% CI:  $-6.12$  to  $-2.98$ ; Table S2A). The corresponding platelet count levels (mean  $\pm$  SD) for the 3 genotypes in the CBR sample were GG =  $242.8 \pm 55.2 \text{ } 10^9/L$ , GC =  $247 \pm 54.0 \text{ } 10^9/L$ , CC =  $249.3 \pm 56.4 \text{ } 10^9/L$ . Association of rs342293 with PLT conditioning on MPV, however, was not significant ( $P = .472$ ; Table S3), indicating that the effect of rs342293 on count is fully explained by its association with volume.

Because MKs and EBs share a common bipotent precursor cell, we investigated whether rs342293 has a similar effect on red cell volume and number. Genotypes at rs342293 were not associated with any of the red blood cell traits tested (Table S2), indicating that the effect of this SNP is unique to regulatory processes in the megakaryocytic lineage.

### SNP rs342293 modifies gene expression

We fine-mapped the region by meta-analysis of 918 genotyped and imputed SNPs from 3 genome-wide scans over a 1-MB interval centered on rs342293 (Figure 1). The strongest association signal was detected at rs342293. The hypothetical gene *FLJ36031* was most proximal to the SNP at a distance of 70 889 bp, and 5 additional genes were identified within the 1-MB interval studied (Figure 1). We assayed expression of the 6 genes in stem cell-derived MKs and EBs with whole-genome Illumina HumanWG-6 (v1) Expression BeadChip arrays.<sup>30,32</sup> All 6 genes were transcribed in MKs (Figure 2A). Remarkably, *PIK3CG* was robustly transcribed in MKs but not in EBs.



**Figure 1. Results of genome-wide association scan.** Regional plot of mean platelet volume association results for the chromosome 7 loci across the 3 GWAS. Meta-analysis  $\log_{10} P$  values are plotted as a function of genomic position (Build 36). The  $P$  value for the lead SNP rs342293 is denoted by large red (GWAS) and blue (combined discovery and replication) diamonds. Proxies are indicated with diamonds of smaller size, with colors assigned based on the pairwise  $r^2$  values with the lead SNP in the HapMap CEU sample: red ( $r^2 > 0.8$ ), orange ( $0.5 < r^2 < 0.8$ ), or yellow ( $0.2 < r^2 < 0.5$ ). White indicates either no LD with the lead SNP ( $r^2 < 0.2$ ), or loci in which such information was not available. Recombination rate estimates (HapMap Phase II) are given in light blue, Refseq genes (NCBI) in green.

We tested the effect of rs342293 on gene transcription in platelets from 35 subjects obtained with the same Illumina HumanWG-6 (v1) arrays as detailed in “Methods.”<sup>30</sup> Because of RNA decay in platelets, signal intensities for platelet RNAs were on average 2-fold lower than signal intensities in MKs (Figure 2A). In addition, approximately 50% of the transcripts present in MKs could not be detected in platelets. We observed that the transcripts of 2 of the 6 genes, *PIK3CG* and *PRKAR2B*, were robustly detected in platelet RNA samples, and that subjects carrying the GG genotype had significantly higher *PIK3CG* mRNA levels ( $8.73 \pm 0.11$ ) compared with GC ( $8.44 \pm 0.10$ ) and CC ( $8.33 \pm 0.12$ ) genotypes (permutation  $P = .047$ ; Figure 2B). We did not observe a significant association with *PRKAR2B* expression (permutation  $P = .12$ ), although we note that GG homozygotes on average had higher expression ( $11.99 \pm 0.15$ ) compared with CG ( $11.5 \pm 0.14$ ) and CC ( $11.45 \pm 0.28$ ) subjects.

It is possible that the observed effect of rs342293 on *PIK3CG* transcript levels is the result of *trans*-acting elements. We therefore predicted the effect of each allele of rs342293 for potentially altered transcription factor binding sites. A region of 41 bases centered on rs342293 was used for a search in the TRANSFAC 7.0 database<sup>37</sup> with MotifScanner as described in “Methods.” Motifs that were called for only one of the alleles were selected, leaving 12 potential transcription factor binding sites that were altered by the SNP. We further investigated the expression of all the 12 transcription factors in MKs<sup>32</sup> and found that only 2 were robustly expressed in MKs, *GATA-1/2* and *EVII*. In both cases the C > G change leads to the loss of the putative binding site.

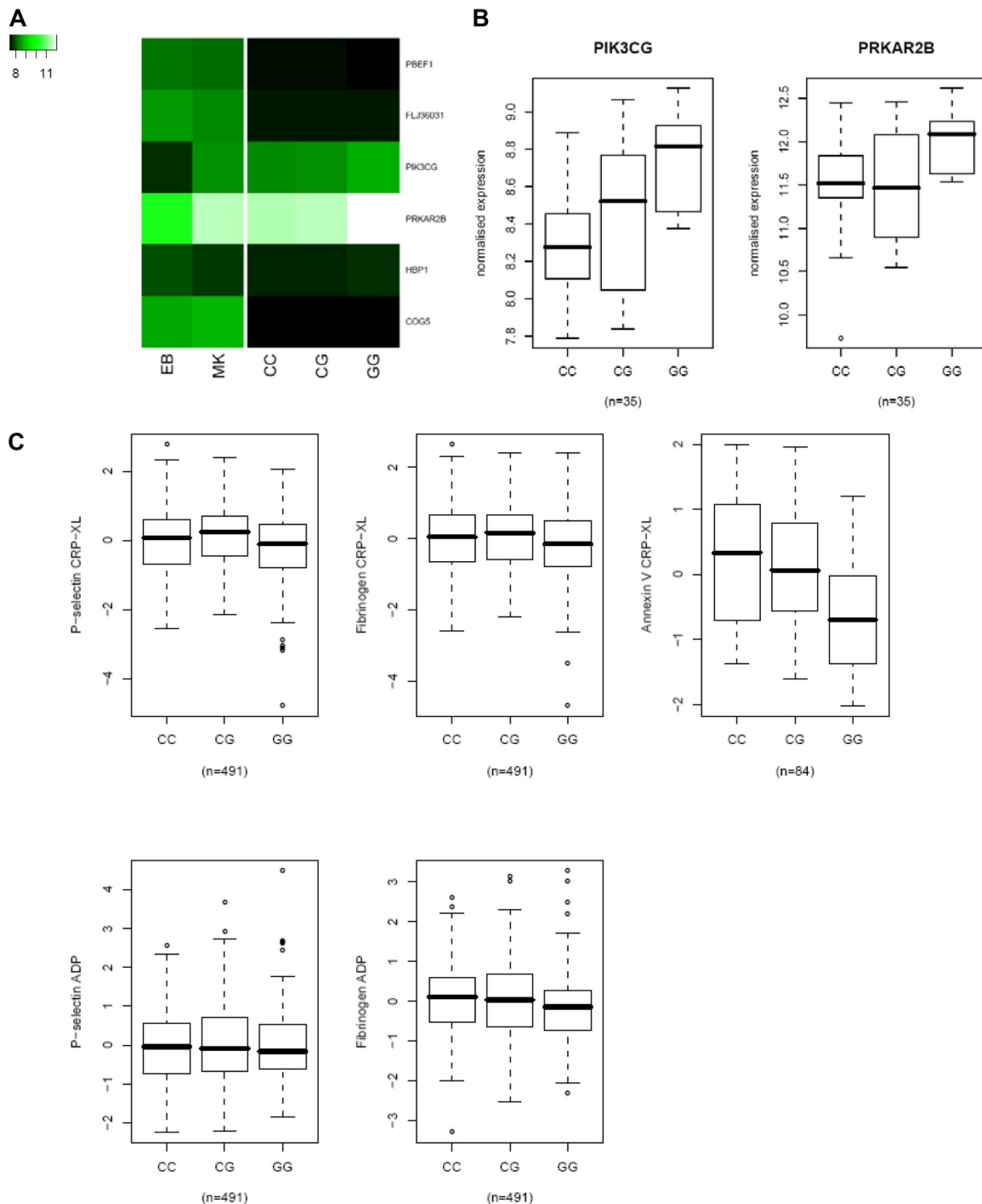
#### SNP rs342293 modifies the level of annexin V binding to CRP-XL-activated platelets

The regulation of MPV is integrally linked to megakaryopoiesis and PPF, which depends on signals received by MKs from cytokines and ECMPs. The interaction with fibrinogen and collagen is considered

central to this process, with the former driving platelet formation through the  $\alpha$ Ib $\beta$ 3 integrin<sup>11,12</sup> and the latter having an inhibitory effect mediated via the  $\alpha$ 2 $\beta$ 1 integrin. Further outside-in signals may be emanating from the interaction of collagen with its activatory and inhibitory signaling receptors glycoprotein (GP) VI and LAIR1, respectively.<sup>10,38</sup> To determine whether rs342293 genotypes exert an effect on the signaling pathway downstream of the GPVI/Fc $\gamma$  chain complex, we interrogated the platelet function phenotypes in our PFC of 500 healthy subjects. Two measures of platelet function were tested, namely fibrinogen binding (a measure of  $\alpha$ Ib $\beta$ 3 activation) and P-selectin expression (a measure of  $\alpha$  granule release), after activation of platelets with either the GPVI-specific collagen mimetic CRP-XL or ADP.<sup>35</sup> We observed that subjects carrying the G allele showed a trend of decreased fibrinogen binding and P-selectin expression after activation of platelets with CRP-XL ( $P = .05$  and  $.08$ , respectively; Figure 2C). No effect of genotype was observed when platelets were activated with ADP ( $P = .32$  and  $.15$ ). We next tested association with annexin V binding to CRP-XL-activated platelets from 84 subjects from the same cohort. Expression of negatively charged phospholipids reported by annexin V binding is a reliable measure of membrane asymmetry and platelet prothrombotic potential.<sup>39</sup> We found that homozygotes for the G allele at rs342293, which is associated with higher MPV levels, on average had fewer platelets binding annexin V compared with GC and CC genotypes ( $P = .003$ ; Figure 2C). This indicates that the level of expression of phosphatidylserine after activation by CRP-XL is modified by this SNP, an observation consistent with the observed trend of reduced fibrinogen binding.

## Discussion

We performed a genome-wide association analysis for MPV in 1221 healthy subjects followed by replication in 7365 subjects from 4 sample collections and identified the first common (minor allele



**Figure 2. Association of rs342293 with platelet expression and function.** (A) Heatmap of mRNA expression for 6 genes within a 1-MB interval. VST-transformed signal intensities were median-normalized, and values were averaged across biologic replicates in erythroblasts (EB; n = 4), megakaryocytes (MK; n = 4), and platelets (n = 35, whereby expression values were averaged based on rs342293 genotype). Signal intensities obtained with platelets were normalized independently from the MK and EB samples. (B) Boxplots of associations between rs342293 genotype and mRNA levels in 35 platelet samples, shown for the 2 genes with expression above detection level (*PIK3CG* and *PRKAR2B*). (C) Association of rs342293 genotypes with measurements of platelet reactivity in the Platelet Function Cohort. (Top) Standardized residuals of logit transformed levels of P-selectin expression (left), and fraction of platelets binding fibrinogen (middle) or annexin V (right) in response to the collagen mimetic CRP-XL. (Bottom) P-selectin expression and fibrinogen binding in response to ADP. Horizontal bars indicate averages and error bars indicate 1 standard deviation.

frequency [MAF], 0.45-0.46 in Northern Europeans) QTL robustly associated with MPV levels in the general population. The genetic association signal was localized in a relatively small (65 kb) intergenic region that did not contain genes, indicating that regulation of platelet volume and function are probably mediated through *cis*- or *trans*-regulatory effects. The observation that rs342293 modifies the response of platelets

to CRP-XL and the volume of platelets, but not the volume of red cells, indicates that the effect of the SNP is restricted to the MK lineage.

Our data, however, do not provide conclusive evidence on which genes may underlie variation in MPV and function. All 6 genes within the arbitrarily chosen 1-MB interval centered on rs342293 are transcribed in MKs, a surprising finding given that only approximately 40%

of all human genes are transcribed in MKs.<sup>32</sup> The transcript level of *PIK3CG* and *PRKAR2B* in platelets was robust enough to answer the question whether the implicated SNP exerts an effect on the respective transcript levels. This showed a nominally significant effect for *PIK3CG* but not for *PRKAR2B* (Figure 2B). Interestingly, to note that subjects homozygous for the G allele had higher *PRKAR2B* transcript levels than CG and CC subjects.

Both these genes are biologically plausible candidates for genes affecting platelet volume and function. *PICK3CG* encodes for the  $\gamma$  chain of pi3/pi4-kinase, which is responsible for the synthesis of phosphatidylinositol-3,4,5-trisphosphate (PtdIns[3,4,5]P<sub>3</sub>, PIP<sub>3</sub>) by conversion of PtdIns-3,4-bisphosphate.<sup>40</sup> The formation of PIP<sub>3</sub> by PIK3CG in MKs and platelets is a key step in the signal transduction cascades leading to the release of calcium from the intracellular stores and controlling cellular movement, adhesion, and contraction, all 3 essential to megakaryopoiesis and PPF. It is also an essential intermediary in the regulation of phospholipase C (PLC $\gamma$ 2) in MKs and platelets in response to collagen.<sup>41</sup> PI3K $\gamma$ -deficient mice display impaired aggregation after stimulation with ADP, but not after stimulation with collagen or thrombin, and are protected from death caused by ADP-induced platelet-dependent thromboembolic vascular occlusion.<sup>42</sup> The observation that the *PIK3CG* transcript is absent from EBs (Figure 2A) is compatible with the lineage-restricted effect of rs342293 on cell volume. *PRKAR2B* encodes the  $\beta$  chain of cAMP-dependent protein kinase, which has an opposing effect to PIK3CG and reduces the release of intracellular calcium by the phosphorylation of ITPR3, the receptor for PIP<sub>3</sub>.<sup>43,44</sup>

In our study, the minor G allele at rs342293 was associated with increased MPV ( $P = 1.08 \times 10^{-24}$ ) and also with decreased platelet reactivity measured as the proportion of activated platelets binding annexin V ( $P = .003$ ) and the level of fibrinogen binding ( $P = .05$ ). The directionality of association between MPV and the number of annexin V–reactive platelets, measured indirectly through the association with rs342293, thus would appear to support a negative correlation between these 2 variables. This contradicts the positive correlation between platelet volume and thrombogenicity which has been suggested by several groups but remains disputed.<sup>2</sup>

The differential functional response of platelets of different genotype on activation with CRP-XL indicates that sequence variation in the 7q22.3 region probably most likely exerts its effect on molecular events downstream of the GPVI/Fc receptor  $\gamma$  chain complex. We consider it unlikely that the observed association between sequence variation at rs342293 and the platelet response to CRP-XL is at the level of GPVI membrane expression, because the binding of the anti-GPVI monoclonal antibody HY101 did not differ significantly between the 3 genotype groups from the PFC (data not shown, but see Jones et al<sup>35</sup>). There are several plausible explanations why no effect of rs342293 was observed on platelet activation by ADP.

First, ADP is a weaker agonist, and the range of values for fibrinogen binding in the PFC was less widely distributed than compared with the range observed after activation with CRP-XL.<sup>35</sup> Second, ADP is less effective in bringing about the expression of phosphatidylserine and annexin V binding.<sup>45</sup> Third, the effect of SNP rs342293 on function may be pathway specific. This is a reasonable assumption that has been validated in our recently completed study of 96 candidate genes encoding proteins in the ADP and CRP-XL pathways, in which we observed that sequence variation in 16 genes was significantly associated with the levels of platelet activation by ADP or CRP-XL ( $P \leq .005$ ). For 5 of these genes the effect was restricted to the CRP-XL activation pathway.<sup>46</sup>

It remains to be established whether the effect of rs342293 CRP-XL induced annexin V binding to platelets is relevant for the regulation of platelet volume during PPF. Previous studies have provided several lines of evidence that apoptosis and platelet formation are integrally linked.<sup>47</sup> The molecular events that lead to the expression of phosphatidylserine in CRP-XL–activated platelets are very similar to those observed in the process of apoptosis. We therefore reason that SNP rs342293 may modify the PPF process with the same directionality as observed for platelet function.

Alternatively, it is possible that the 65-kb region identified in this study harbors *trans*-acting regulatory elements relevant to hematopoiesis and especially megakaryopoiesis. We did obtain computational evidence that the C allele may introduce a binding site for GATA-1/2 and EVI1, a finding that will require experimental validation. Both GATA-1 and EVI1 are key transcription factors for megakaryopoiesis. *EVI1* is located on chromosome band 3q26, and rearrangements in this region can lead to its overexpression in hematopoietic cells leading to acute myeloid leukemias, which are frequently associated with dystrophic megakaryopoiesis.<sup>48</sup> GATA-1 knockdown mice have severe thrombocytopenia,<sup>49</sup> and rare nonsynonymous SNPs in the *GATA-1* gene result in X-linked macrothrombocytopenia, confirming the pivotal role of GATA-1 in megakaryopoiesis and platelet formation.<sup>50</sup>

In conclusion, we have identified the first QTL associated with variation in platelet volume, count, and function in the normal population. The discovery that common sequence variation with a MAF of 0.45 in this region modifies platelet phenotypes warrants further experiments to unravel the genetic mechanisms underlying the observed effects. The observation that chromosomal aberrations in the same region of chromosome 7 have been linked to myeloid malignancies<sup>22</sup> lends further support to the hypothesis that this region harbors important regulatory sites of hematopoiesis.

The 7q22.3 locus explains a small fraction (1.5%) of total variance in MPV attributable to genetic factors,<sup>5</sup> indicating that many additional QTLs are likely to be soon discovered through further genome-wide studies involving a larger number of subjects. This may become clinically relevant because the manipulation of platelet function and number by drugs is one of the successful pharmacologic interventions in cardiovascular diseases and ET. The identification of additional QTLs for blood cell indices will improve the understanding of the complexity of fate determination in hematopoiesis and the molecular machinery underlying platelet formation and function. This may in the future translate into a new generation of safer drugs.

## Acknowledgments

We thank the staff members of CBR for donor recruitment and related phenotyping. Detailed acknowledgments are reported in Document S2.

This work was supported by The Wellcome Trust (P.D., T.D.S., W.H.O., and N.S.), the European Union FP-5 (QLG2-CT-2002-01254; P.D., H.S., J.E., N.S., and W.H.O.), and the FP-6–funded Bloodomics Integrated Project (LSHM-CT-2004-503485; W.H.O.). Support for the UKBS-CC (W.H.O.) and TwinsUK (T.D.S.) were from The Wellcome Trust. The CBR is a resource of the Cambridge Biomedical Research Center and is funded by grants from the National Institute for Health Research, The Wellcome Trust, and the Juvenile Diabetes Research Foundation International. The MONICA/KORA Augsburg studies were funded by the Helmholtz Zentrum München, the German Federal

Ministry of Education and Research (BMBF), and the German National Genome Research Network (NGFN; A.D. and T.I.).

## Authorship

Contribution: T.I., T.D.S., W.H.O., J.R.B., H.S., N.J.S., and J.E. contributed DNA samples; P.D., A. Rankin, J.S., T.I., P.B., H.P., N.A.W., S.C.P., and J.J. performed research (genotyping and gene expression); S.F.G., A.D., C.M., S.M., S.L.T., A.H.G., C.I.J., and W.E. performed phenotyp-

ing; N.S., A. Rendon, C.G., S.L.B., M.F., and B.K. analyzed the data; and N.S., T.D.S., W.H.O., S.M., C.G., A. Rendon, P.D., and A.H.G. wrote the paper.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

Correspondence: Willem H. Ouwehand, Department of Haematology, University of Cambridge, Long Rd, Cambridge CB2 2PT, United Kingdom; e-mail: who1000@cam.ac.uk; or Nicole Soranzo, Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, United Kingdom; e-mail: ns6@sanger.ac.uk.

## References

- Martin JF, Bath PM, Burr ML. Influence of platelet size on outcome after myocardial infarction. *Lancet*. 1991;338:1409-1411.
- Boos CJ, Lip GYH. Assessment of mean platelet volume in coronary artery disease – what does it mean? *Thrombosis Res*. 2007;120:11-13.
- Evans DM, Frazer IH, Martin NG. Genetic and environmental causes of variation in basal levels of blood cells. *Twin Res*. 1999;2:250-257.
- Bertin A, Mahaney MC, Cox LA, et al. Quantitative trait loci for peripheral blood cell counts: a study in baboons. *Mamm Genome*. 2007;18:361-372.
- Mahaney MC, Brugnara C, Lease LR, Platt OS. Genetic influences on peripheral blood cell counts: a study in baboons. *Blood*. 2005;106:1210-1214.
- Lettre G, Jackson AU, Gieger C, et al. Identification of ten loci associated with height highlights new biological pathways in human growth. *Nat Genet*. 2008;40:584-591.
- Weedon MN, Lango H, Lindgren CM, et al. Genome-wide association analysis identifies 20 loci that influence adult height. *Nat Genet*. 2008;40:575-583.
- Campbell PJ, Green AR. The myeloproliferative disorders. *N Engl J Med*. 2006;355:2452-2466.
- Wendling F, Maraskovsky E, Debili N, et al. cMpl ligand is a humoral regulator of megakaryocytopoiesis. *Nature*. 1994;369:571-574.
- Sabri S, Jandrot-Perrus M, Bertoglio J, et al. Differential regulation of actin stress fiber assembly and proplatelet formation by alpha2beta1 integrin and GPVI in human megakaryocytes. *Blood*. 2004;104:3117-3125.
- Larson MK, Watson SP. Regulation of proplatelet formation and platelet release by integrin alpha IIb beta3. *Blood*. 2006;108:1509-1514.
- Ghevaert C, Salsmann A, Watkins NA, et al. A nonsynonymous SNP in the ITGB3 gene disrupts the conserved membrane-proximal cytoplasmic salt bridge in the alphaIIb beta3 integrin and cosegregates dominantly with abnormal proplatelet formation and macrothrombocytopenia. *Blood*. 2008;111:3407-3414.
- Avecilla ST, Hattori K, Heissig B, et al. Chemokine-mediated interaction of hematopoietic progenitors with the bone marrow vascular niche is required for thrombopoiesis. *Nat Med*. 2004;10:64-71.
- Italiano JE Jr, Lecine P, Shivdasani RA, Hartwig JH. Blood platelets are assembled principally at the ends of proplatelet processes produced by differentiated megakaryocytes. *J Cell Biol*. 1999;147:1299-1312.
- Kaushansky K. Historical review: megakaryopoiesis and thrombopoiesis. *Blood*. 2008;111:981-986.
- Nurden AT. Qualitative disorders of platelets and megakaryocytes. *J Thromb Haemost*. 2005;3:1773-1782.
- Cattaneo M. Inherited platelet-based bleeding disorders. *J Thromb Haemost*. 2003;1:1628-1636.
- Hisa T, Spence SE, Rachel RA, et al. Hematopoietic, angiogenic and eye defects in Meis1 mutant animals. *EMBO J*. 2004;23:450-459.
- Lu J, Guo S, Ebert BL, et al. MicroRNA-mediated control of cell fate in megakaryocyte-erythrocyte progenitors. *Dev Cell*. 2008;14:843-853.
- Garcia P, Frampton J. Hematopoietic lineage commitment: miRNAs add specificity to a widely expressed transcription factor. *Dev Cell*. 2008;14:815-816.
- Shivdasani RA, Rosenblatt MF, Zucker-Franklin D, et al. Transcription factor NF-E2 is required for platelet formation independent of the actions of thrombopoietin/MGDF in megakaryocyte development. *Cell*. 1995;81:695-704.
- Kratz CP, Emerling BM, Bonifas J, et al. Genomic structure of the PIK3CG gene on chromosome band 7q22 and evaluation as a candidate myeloid tumor suppressor. *Blood*. 2002;99:372-374.
- Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. 2007;447:661-678.
- Richards JB, Rivadeneira F, Inouye M, et al. Bone mineral density, osteoporosis, and osteoporotic fractures: a genome-wide association study. *Lancet*. 2008;371:1505-1512.
- Doring A, Gieger C, Mehta D, et al. SLC2A9 influences uric acid concentrations with pronounced sex-specific effects. *Nat Genet*. 2008;40:430-436.
- Purcell S. PLINK: a toolset for whole-genome association and population-based linkage analysis. *Am J Hum Genet*. 2007;81:559-575.
- Li Y, Abecasis GR. Mach 1.0: rapid haplotype reconstruction and missing genotype inference. *Am J Hum Genet*. 2006;2290:579.
- Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin—rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet*. 2002;30:97-101.
- Loos RJ, Lindgren CM, Li S, et al. Common variants near MC4R are associated with fat mass, weight and risk of obesity. *Nat Genet*. 2008;40:768-775.
- Burns P, Gusnanto A, Macaulay IC, et al. Identification of variation in the platelet transcriptome associated with glycoprotein 6 haplotype. *Platelets*. 2008;19:258-267.
- Bugert P, Dugrillon A, Gunaydin A, Eichler H, Kluter H. Messenger RNA profiling of human platelets by microarray hybridization. *Thromb Haemost*. 2003;90:738-748.
- Macaulay IC, Tijssen MR, Thijssen-Timmer DC, et al. Comparative gene expression profiling of in vitro differentiated megakaryocytes and erythroblasts identifies novel activatory and inhibitory platelet membrane proteins. *Blood*. 2007;109:3260-3269.
- Lin SM, Du P, Huber W, Kibbe WA. Model-based variance-stabilizing transformation for Illumina microarray data. *Nucleic Acids Res*. 2008;36:e11.
- Coessens B, Thijs G, Aerts S, et al. INCLUSiVe: a web portal and service registry for microarray and regulatory sequence analysis. *Nucleic Acids Res*. 2003;31:3468-3470.
- Jones CI, Garner SF, Angenent W, et al. Mapping the platelet profile for functional genomic studies and demonstration of the effect size of the GP6 locus. *J Thromb Haemost*. 2007;5:1756-1765.
- Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet*. 2007;39:906-913.
- Gesellschaft für Biotechnologische Forschung mbH, Department of Genome Analysis. TRANSFAC. <http://www.gene-regulation.com>. Accessed August 2008.
- Lebbink RJ, de Ruyter T, Adelmeijer J, et al. Collagens are functional, high affinity ligands for the inhibitory immune receptor LAIR-1. *J Exp Med*. 2006;203:1419-1425.
- van Engeland M, Nieland LJ, Ramaekers FC, Schutte B, Reutelingsperger CP. Annexin V-affinity assay: a review on an apoptosis detection system based on phosphatidylserine exposure. *Cytometry*. 1998;31:1-9.
- Hawkins PT, Stephens LR. PI3Kgamma is a key regulator of inflammatory responses and cardiovascular homeostasis. *Science*. 2007;318:64-66.
- Pasquet JM, Gross BS, Gratacap MP, et al. Thrombopoietin potentiates collagen receptor signaling in platelets through a phosphatidylinositol 3-kinase-dependent pathway. *Blood*. 2000;95:3429-3434.
- Hirsch E, Bosco O, Tropel P, et al. Resistance to thromboembolism in PI3Kgamma-deficient mice. *FASEB J*. 2001;15:2019-2021.
- Tertyshnikova S, Fein A. Inhibition of inositol 1,4,5-trisphosphate-induced Ca<sup>2+</sup> release by cAMP-dependent protein kinase in a living cell. *Proc Natl Acad Sci U S A*. 1998;95:1613-1617.
- Wojcikiewicz RJ, Luo SG. Phosphorylation of inositol 1,4,5-trisphosphate receptors by cAMP-dependent protein kinase. Type I, II, and III receptors are differentially susceptible to phosphorylation and are phosphorylated in intact cells. *J Biol Chem*. 1998;273:5670-5677.
- Holme PA, Solum NO, Brosstad F, Egberg N, Lindahl TL. Stimulated Glanzmann's thrombasthenia platelets produced microvesicles. Microvesiculation correlates better to exposure of procoagulant surface than to activation of GPIIb-IIIa. *Thromb Haemost*. 1995;74:1533-1540.
- Jones CI, Bray S, Garner SF, et al. A functional genomics approach reveals novel quantitative trait loci associated with platelet signaling pathways. *Blood*. In press.
- De Botton S, Sabri S, Daugas E, et al. Platelet formation is the consequence of caspase activation within megakaryocytes. *Blood*. 2002;100:1310-1317.
- Buonamici S, Li D, Chi Y, et al. EVI1 induces myelodysplastic syndrome in mice. *J Clin Invest*. 2004;114:713-719.
- McDevitt MA, Shivdasani RA, Fujiwara Y, Yang H, Orkin SH. A "knockdown" mutation created by cis-element gene targeting reveals the dependence of erythroid cell maturation on the level of transcription factor GATA-1. *Proc Natl Acad Sci U S A*. 1997;94:6781-6785.
- Freson K, Devriendt K, Matthijs G, et al. Platelet characteristics in patients with X-linked macrothrombocytopenia because of a novel GATA1 mutation. *Blood*. 2001;98:85-92.