2SNP heritability and effects of genetic variants for neutrophil-to-lymphocyte and platelet-to-lymphocyte ratio

Bochao Danae Lin1, Elena Carnero-Montoro2, Jordana T Bell2, Dorret I Boomsma1, Eco J de Geus1,3, Rick Jansen4, Cornelia Kluit5, Massimo Mangino2,6, Brenda Penninx4, Tim D Spector2, Gonneke Willemsen1 and Jouke-Jan Hottenga1,3

Neutrophil-to-lymphocyte ratio (NLR) and platelet-to-lymphocyte ratio (PLR) are important biomarkers for disease development and progression. To gain insight into the genetic causes of variance in NLR and PLR in the general population, we conducted genome-wide association (GWA) analyses and estimated SNP heritability in a sample of 5901 related healthy Dutch individuals. GWA analyses identified a new genome-wide significant locus on the HBS1L-MYB intergenic region for PLR, which replicated in a sample of 2538 British twins. For platelet count, we replicated three known genome-wide significant loci in our cohort (at CCDC71L-PK3CG, BAK1 and ARHGEF3). For neutrophil count, we replicated the PSMD3 locus. For the identified top SNPs, we found significant cis and trans expression quantitative trait loci effects for several loci involved in hematological and immunological pathways. Linkage Disequilibrium score (LD) regression analyses for PLR and NLR confirmed that both traits are heritable, with a polygenic SNP heritability for PLR of 14.1%, and for NLR of 2.4%. Genetic correlations were present between ratios and the constituent counts, with the genetic correlation (r = 0.45) of PLR with platelet count reaching statistical significance. In conclusion, we established that two important biomarkers have a significant heritable SNP component, and identified the first genome-wide locus for PLR.

Journal of Human Genetics (2017) 62, 979–988; doi:10.1038/jhg.2017.76; published online 3 August 2017

INTRODUCTION

Both neutrophil-to-lymphocyte ratio (NLR) and platelet-to-lymphocyte ratio (PLR) have been suggested as novel and useful biomarkers for the diagnosis or prognostic prediction of diseases.1,2 A high NLR level was shown to be an independent predictor of mortality in patients undergoing cardiac revascularization3 and in patients with myocardial infarction.4 Elevated NLR levels were also related to a poor prognosis of various cancers, such as esophageal, pancreatic, lung, ovarian and hepatocellular cancer.5–7 Similar to NLR, PLR was also reported as an index for diagnosis or prognostic prediction of oncologic disorders and inflammatory diseases.8,9 NLR and PLR thus may serve as biomarkers in patient populations. However, studies of variation in these biomarkers within healthy populations are scarce. Recently, we showed that variation in NLR and PLR levels is owing to genetic influences, with a broad sense heritability of 36% for NLR and 64% for PLR, using a twin family epidemiological design.10 Here, we investigate if the significant heritability estimates can be explained by common single-nucleotide polymorphisms (SNP) and if we can identify the genes that play a role in these two blood ratios. We also investigate if our findings are unique to the two ratios or whether their count-components (that is, lymphocyte, neutrophil and platelet counts) show similar results.

No genome-wide association study (GWAS) has yet been published for NLR and PLR. However, GWASs on their subcomponents, the neutrophil, lymphocyte and platelet counts were carried out in different populations including European,11–14 African–American,15–18 Korean19,20 and Japanese populations.21,22 These GWASs for blood cell count in different cohorts have identified multiple genetic loci for blood cell components. For neutrophil count, the DARC gene promoter at 1q23.3 was identified in African–American populations23 and loci at 20p12 (PLCB4 gene)22 and 7q21.2 (CDK6 gene)24 were found in the Japanese population. The chromosomal region nearby PSMD3 on 17q21 was associated in a GWAS meta-analysis in both Japanese and European ancestry cohorts, but not in African–American cohorts. The variants at AK123889 on 6p21.33 were novel findings in a European ancestry cohort,15 and were also confirmed by meta-analysis.24 For lymphocyte count, two genetic variants nearby EFS1SL1 gene on 6p21 and LOC101929772 on 19p13

1Department of Biological Psychology, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands; 2Department of Twin Research and Genetic Epidemiology, Kings College London, London, UK; 3EMGO+ Institute for Health & Care Research, VU Medical Center, Amsterdam, The Netherlands; 4Department of Psychiatry, VU Medical Center, Amsterdam, The Netherlands; 5Good Biomarker Sciences, Leiden, The Netherlands and 6NIHR Biomedical Research Centre at Guy’s and St Thomas’ Foundation Trust, London, UK.

Correspondence: Dr JJ Hottenga, Department of Biological Psychology, Vrije Universiteit Amsterdam, van der Boechorststraat 1, Amsterdam 1081BT, The Netherlands.

E-mail: j.j.hottenga@vu.nl

Received 17 February 2017; revised 24 May 2017; accepted 13 June 2017; published online 3 August 2017
were identified. For platelet count, many loci were identified: SH2B3 on 12q24, ARHGEF3 on 3p14.3, ZBTB9-BAK1 on 6p21.31, KIAA0232 on 4p16.1, EGF on 4q25, PNPLA3 on 22q13.31 in the Korean population. ARHGEF3 on 3p14.3, PEAR1 on 1q23.1, BMP1A on 10q23.2, loci on 6p22, 7q11, 10q21, 11q13, 20q13 were detected in the African–American population and over 55 loci including CTC7, PIK3CG, ARHGEF3, BAK1 and HBS1L-MYB in the European population.

Some blood cell count loci show pleiotropy: they influence multiple hematological indices. For example, the genetic region nearby AK123889 on 6p21.33 was associated with neutrophil count, lymphocyte count and total white blood cell count, and the DARC promoter on 1q23.2 was associated with neutrophil count, monocyte count and total number of white blood cells. The intergenic promoter on 1q23.2 was associated with neutrophil count, monocyte count and to determine the genetic correlation between variants all autosomal SNPs, to examine polygenetic effects between NLR and PLR and neutrophil, lymphocyte and platelet counts. The variants were associated with total white blood cell count HBS1L-MYB and also with number of neutrophils, lymphocytes, erythrocytes, eosinophils, monocytes and platelets. Therefore, we also examined genetic effects across the ratios and constituent cell counts.

We conducted five GWASs to identify genetic variants associated with NLR, PLR and neutrophil, lymphocyte and platelet counts. The discovery cohort consisted of 5901 healthy participants from the Netherlands Twin Register (NTR) and replication of top results was sought in the TwinsUK cohort consisting of 2538 participants. Furthermore, all top SNPs, which showed a significant association with our phenotypes of interest, were selected for an expression quantitative trait loci (eQTL) analysis to test whether these variants have an effect on the gene expression level. For the ratios, we estimated the proportion of trait variance explained by significant SNPs from the GWAS and the variance explained by SNPs that were associated with lymphocyte, platelet and neutrophil counts. Using the summary statistics of the GWAS results, we applied linkage disequilibrium (LD) regression to determine the variance explained by all autosomal SNPs, to examine polygenic effects between NLR and PLR, and to determine the genetic correlation between variants affecting the two ratios, their subcomponents and LD-Hub published GWAS.

MATERIALS AND METHODS

Participants

All participants were registered with the Netherlands Twin Register (NTR) and had taken part in biobank projects conducted between 2004 and 2011. After removing outliers (defined as values outside mean ± 2.5×s.d. for NLR, PLR or their subcomponents), the sample size for PLR and NLR was 9434 individuals from 3411 families. We further excluded individuals who met one or more of these criteria: (1) illness in the sampling collection week (N=539); (2) values of CRP ≥ 15 mg/dL (N=287); (3) basophil count > 0.02 × 10^9/L (N=151); (4) report of chronic immune disease or cancer (N=83); and (5) anti-inflammatory medication, glucocorticoids or iron supplements (N=537). When linking these data to the genetic data, 6112 individuals had both phenotype and genotype data. After exclusion of 211 individuals with non-Dutch ancestry (based on genotype information), the sample size was 5901 individuals. Written informed consent was obtained from all participants and the Medical Ethics Committee of the VU Medical Centre approved the study protocols.

Blood sampling and cell counts

Blood samples were obtained during a home visit, or sometimes a work visit, between 7 and 10 a.m. Participants were instructed to fast overnight and to refrain from heavy physical exertion and medication use (if possible) in the morning before the visit. Smokers were asked to abstain from smoking at least 1 h prior to the visit. For fertile women without hormonal birth control, when possible, an appointment was made within the 2nd to 4th day of the menstrual cycle and women taking hormonal birth control were visited during the pill-free week. Peripheral venous blood samples were collected into multiple anticoagulant vacuum tubes. Within 3–6 h upon blood withdraw tubes were transported to the laboratory. During the visit, data were also collected on body composition, the presence of chronic diseases, medication use and smoking history.

The hematological profile, including the number of neutrophils, lymphocytes and platelets, was obtained from 2 ml EDTA tubes using the Coulter system (Coulter Corporation, Miami, FL, USA). NLR was calculated as the absolute neutrophil count (10^9/l) divided by the absolute lymphocyte count (10^9/l) and PLR was calculated as the absolute platelet count (10^9/l) divided by the absolute lymphocyte count (10^9/l).

Genotype data

For DNA isolation, we used the GENTRA Puregene DNA isolation kit. Genotyping was done on multiple chip platforms, with a number of overlapping participants. Chronologically the following platforms were used: Affymetrix Perlegen 5.0 (N=1718), Illumina 370 (N=424), Illumina 660 (N=1103), Illumina Omni Express 1 M (N=346) and Affymetrix 6.0 (N=3602). Genotype calls were made with the platform-specific software (Beadstudio for the Affymetrix platform and Beadstudio) for each specific array. Quality control was done within and between platforms and subsets. For each platform, the individual SNP markers were lifted over to build 37 (HG19) of the Human Reference genome, using the LiftOver tool (http://genome.sph.umich.edu/wiki/LiftOver). The data were then strand aligned with the 1000 Genomes GIANT phase1 release v3 20,101,123 SNPs INDELS SSV ALL panel. SNPs from each platform were removed if they had ambiguous locations, mismatching alleles with this imputation reference set or the allele frequencies differed > 0.20 compared with the reference. From each platform, SNPs were also excluded if meeting the following criteria: a minor allele frequency (MAF) < 1%, Hardy–Weinberg Equilibrium with P<0.00001 and call rate < 95%. Samples were excluded from the analysis when their expected sex did not match their genotyped sex, the genotype missing rate was above 10% or the PLINK 1.07 F inbreeding value was either > 0.10 or < -0.10. After these steps, the data of the individual arrays were merged into a single data set using PLINK 1.07. Within the merged set, identity by state sharing was calculated between all possible pairs of participants and compared with the known NTR family structures. Samples were removed if the data did not match their expected identity by state sharing. The concordance rate of DNA samples on multiple platforms for overlapping SNPs generally exceeded 99.0% after data cleaning. The Hardy–Weinberg Equilibrium, MAF- and the reference allele frequency difference < 0.20 filters were re-applied in the combined data. As a final step, SNPs with C/G and A/T allele combinations were removed when the imputed allele frequency difference > 0.3 from the 1000G reference allele, Hardy–Weinberg Equilibrium with P<0.01 or 0.80. Hardy–Weinberg Equilibrium was calculated on the 1000G reference allele frequency, MAF<0.01 or R^2<0.80. Hardy–Weinberg Equilibrium was calculated on the genotypic probability counts for the full sample, and SNPs were removed if the P-value < 0.00001. This left 6 010 458 SNPs for the GWAS.

Analyses. Generation of genetic relatedness matrices. GRMs with the values of the identity by state allele sharing for a given set of SNP markers between all possible pairs of individuals were calculated with the GCTA software.
removing SNPs that showed significant genotyping differences between platforms (P<0.0001); 6 009 498 SNPs were retained, which is sufficient for GRM estimation. The SNP data were transformed to best guess Plink binary format, and subsets were made for each of the 22 chromosomes. We generated 25 GRMs: one GRM containing only the significant GWAS SNPs for PLR from our own study, and one GRM containing the SNPs known to be involved in the cell counts. A third GRM was constructed for closely related individuals (identity by state > 0.05), pairwise relationship estimates smaller than 0.05 were set to 0 in this matrix. This matrix is used as second covariate matrix in the GWAS and heritability studies to account for the family structure. Including family members in the GWAS increases the power to detect genes, and using a mixed linear model correction as employed in GCTA, corrects for the statistical inflation that is caused by including the related members. Finally, 22 GRM matrices were made that include all autosomal SNPs, except for the one chromosome on which the SNP is present that is tested in the GWAS: the LOCO (Leave One Chromosome Out) strategy. These matrixes are used in the GWAS as covariates to remove any remaining statistical inflation due to subsample stratification.

GWAS. The first three Dutch principal components as were generated with the EIGENSOFT software were used as covariates in the GWAS (Supplementary Figures 1 and 2). Additional covariates were age, sex and genotype platform. For NLR and PLR as well as for the three sub-component counts we modeled the phenotypes as being influenced by SNP and these six covariates. Analyses were performed with the GCTA software running a mixed linear association model, including the LOCO GRMs for chromosome 1–22, and the close-related GRM. For the GWAS, the significance threshold was P-value < 5 x 10^-8.

GWAS replication. Replication of significant GWAS hits for NLR, PLR or individual blood cell counts, which were not previously found, was examined in TwinsUK. TwinsUK is an UK-based twin registry with a focus on the genetic and environmental etiology of age related complex traits and diseases. Samples from TwinsUK were genotyped using the Illumina Hap317K and Hap610K assays (Illumina, San Diego, USA) following standard procedures. Normalized intensity data were pooled and genotypes called on the basis of the Illumina’s algorithm. No calls were assigned if the most likely call was less than a posterior probability of 0.95. SNPs were excluded if they had a low call rate (< 95%) and/or Hardy–Weinberg P-value < 1 x 10^-4. Subjects were also removed if the sample call rate was < 95%, autosomal heterozygosity was outside the expected range, genotype concordance was over 97% with another sample and the sample was of lesser call rate. Imputation of genotypes was carried out using the software IMPUTE. The best guess Plink binary format data were used to conduct the replication analysis. The sample size of the TwinsUK data set was 2538 subjects with genetic and phenotypic information, after values outside mean ± 2.5 s.d. in the phenotype of interest were removed. We tested the association with the SNPs using a linear mixed model, in which the traits were regressed on the SNPs, whereas correcting for age and sex as fixed effects variables.

eQTL analysis. To determine the effects of the GWAS located genetic variants for both ratios as well as the constituent counts, we conducted eQTL analysis, using the NESDA-NTR Conditional eQTL Catalog (online accessible: https://eqtl.onderzoek.io). The details of the eQTL analysis are described in the Supplementary Method Material. In brief, eQTL effects were examined with a linear model approach using MatrixeQTL with expression level as dependent variable and SNP genotype values as independent variable. eQTL effects were defined as cis when probe set–SNP pairs were at distance < 1 Mb base pairs (Mb), and as trans when the SNP and the probe set were separated by more than 1 Mb on the genome according to the Human reference genome HG19. To determine whether the observed cis and trans effects may reflect causal mechanism we checked the LD of our top SNPs with the top SNPS identified for gene expression in the implicated genes. As gene expression is related to blood composition we repeated the analysis with and without correction for blood composition components (specifically mean corpuscular volume, red cell distribution width, and neutrophil, lymphocyte, monocyte, eosinophil, basophil and platelet counts).

SNP heritability and genetic correlations. The variance explained by the significant SNPs in our GWAS for PLR was estimated with the GCTA software. The variance explained in NLR and PLR was estimated with GCTA for the known loci from literature for neutrophil, platelet and lymphocyte blood cell counts. For each analysis we included family members and therefore included the closely related GRM under the Restricted Maximum Likelihood analysis procedure within GCTA. The variance explained by all SNPs was estimated by LD regression between our computed GWAS summary statistics effect sizes and the expected Hapmap 3 LD. To do this, we used the HapMap3 LD scores (NSNPs = 1 293 150), computed for each SNP based on the LD observed in European ancestry individuals from 1000 Genomes project (online accessible: https://github.com/bulik/lisd). The criteria of passing quality control for SNPs were the default by LD regression: imputation quality info > 0.90, MAF > 0.01. SNPs with invalid P-values (P > = 1 or P < 0) were excluded. In addition, variants that are not SNPs (for example, insertion–deletions), strand ambiguous SNPs and SNPs with duplicated RS numbers were also excluded. After quality control, the number of SNPs for these analyses reduced to 951 097.

Genetic correlations among the ratios and counts were estimated using LD regression. The principle of this technique is that the genetic correlation of two traits can be calculated by the slope from the LD regression on the product of effect sizes (z-score) for two phenotypes. The genetic correlations between published GWAS available online and our summary statistics were estimated with LD-Hub. For these regression analyses we selected the list of recommended SNPs from the website and extracted those from the GWAS results for the counts and ratios. Note that this list excludes the chromosome 6 MHC region. A total of 1 210 244 SNPs were used in the analyses. Finally, phenotypic Pearson correlations between PLR, NLR and the constituent cell counts were calculated with the SPSS 22 program.

RESULTS GWAS

Summary statistics for the phenotypes of interest are given in Table 1 and GWAS results for NLR and PLR are summarized in the Manhattan and QQ plots in Figures 1 and 2, respectively. The GWAS inflation factors (λ) were 0.9963005 for NLR and 1.020995 for PLR, indicating that there is no hidden stratification left in the studied GWAS sample. For NLR, no loci were found that reached genome-wide significance level. For PLR, there were 20 SNPs located between the HBS1L and MYB genes on chromosome 6q23.3 in the HBS1L–MYB region, which were significantly associated with the phenotype (in Figure 2 Manhattan, Table 2 descriptive and Figure 3 locus zoom).

The top SNP rs9376092 of this locus has a C allele, which significantly increases PLR level (β = 5.48, P = 2.75 x 10^-8). This SNP was also significantly associated with platelet count (β = 6.98, P = 4.05 x 10^-8), but not with lymphocyte count (β = 0.039, P = 0.008). In the TwinsUK sample, rs9376092 replicated with a similar effect for PLR (β = 4.766, P = 0.004) as well as platelet count (β = 6.053, P = 0.002).
Figure 1 (a) Manhattan and (b) QQ plot for the neutrophil–lymphocyte ratio (NLR) GWAS results with SNPs having a MAF > 0.01. A full color version of this figure is available at the Journal of Human Genetics journal online.

Figure 2 (a) Manhattan and (b) QQ plot for the platelet–lymphocyte ratio (PLR) GWAS results with SNPs having a MAF > 0.01. A full color version of this figure is available at the Journal of Human Genetics journal online.

Table 2 The significant SNPs associated in our study for PLR and the P-values for the platelet and lymphocyte counts, these SNPs are all located in the known intergenic HBS1L-MYB region

| RS number | CHR | BP   | Imputed | Imputation quality | R² | MAF  | B     | s.e. | P (LR)       | P (platelet count) | P (lymphocyte count) |
|-----------|-----|------|---------|-------------------|----|------|-------|------|--------------|--------------------|----------------------|
| rs9376092 | 6   | 135427144 | Yo | Imputed | 0.270 | 5.483 | 0.922 | 2.75E-09 | 4.06E-08 | 0.008 |
| rs4895440 | 6   | 135426558 | Yes | 0.998 | 0.271 | 5.478 | 0.922 | 2.83E-09 | 4.45E-08 | 0.007 |
| rs35999442 | 6 | 135424179 | Yes | 0.987 | 0.272 | 5.476 | 0.922 | 2.9E-09  | 2.30E-08 | 0.009 |
| rs4895441 | 6   | 135426573 | No | 0.270 | 5.472 | 0.922 | 2.95E-09 | 4.37E-08 | 0.007 |
| rs11799553 | 6 | 135422296 | Yes | 0.984 | 0.273 | 5.446 | 0.919 | 3.19E-09 | 2.64E-08 | 0.010 |
| rs776054  | 6   | 135418916 | No | 0.258 | 5.492 | 0.934 | 4.21E-09 | 3.10E-08 | 0.014 |
| NA  | 6   | 135418632 | Yes | 0.997 | 0.257 | 5.474 | 0.934 | 4.74E-09 | 3.10E-08 | 0.016 |
| rs9399137 | 6   | 135419018 | No | 0.257 | 5.468 | 0.934 | 4.89E-09 | 1.99E-08 | 0.016 |
| rs35786788 | 6 | 135419042 | Yes | 0.998 | 0.257 | 5.468 | 0.934 | 4.89E-09 | 1.99E-08 | 0.016 |
| rs9373124 | 6   | 135423209 | Yes | 0.983 | 0.274 | 5.365 | 0.918 | 5.24E-09 | 2.64E-08 | 0.011 |
| rs9389268 | 6   | 135419631 | Yes | 0.996 | 0.258 | 5.449 | 0.934 | 4.82E-09 | 2.64E-08 | 0.014 |
| rs9376091 | 6   | 135419636 | Yes | 0.996 | 0.258 | 5.449 | 0.934 | 4.82E-09 | 2.52E-08 | 0.014 |
| rs34164109 | 6 | 135421176 | Yes | 0.994 | 0.258 | 5.449 | 0.934 | 4.82E-09 | 2.52E-08 | 0.014 |
| rs9402685 | 6   | 135419688 | Yes | 0.985 | 0.258 | 5.407 | 0.933 | 6.93E-09 | 3.30E-08 | 0.015 |
| rs7758845 | 6   | 135428537 | Yes | 0.985 | 0.263 | 5.362 | 0.929 | 7.86E-09 | 5.06E-08 | 0.015 |
| rs9376090 | 6   | 135411228 | No | 0.252 | 5.402 | 0.937 | 8.29E-09 | 1.69E-08 | 0.021 |
| rs9389269 | 6   | 135427159 | No | 0.263 | 5.325 | 0.928 | 9.64E-09 | 4.51E-08 | 0.016 |
| rs9402686 | 6   | 135427817 | Yes | 0.985 | 0.263 | 5.325 | 0.928 | 9.64E-09 | 4.51E-08 | 0.029 |
| rs1331309 | 6   | 135406178 | Yes | 0.979 | 0.252 | 5.221 | 0.936 | 2.45E-08 | 3.93E-08 | 0.029 |
| rs9399136 | 6   | 135402339 | Yes | 0.969 | 0.250 | 5.179 | 0.936 | 3.23E-08 | 3.73E-08 | 0.033 |

Journal of Human Genetics
Here again, the SNP was not associated with lymphocyte count ($\beta = 0.014, P = 0.49$) (Table 3).

Manhattan and QQ plots for the GWAS of neutrophil, lymphocyte and platelet counts are given in Figures 4 to 6. For neutrophil counts we found significant associations ($P < 5 \times 10^{-8}$) for 65 SNPs in LD in the $PSMD3$ locus (Table 4). For lymphocyte count we did not detect any significant genetic associations. For platelet count, a locus in $CCDC71L-PIK3CG$ on 7q22.3 showed the strongest signal in our study ($P = 3.45 \times 10^{-10}$). We also detected genetic variants for platelet count within $ARHGEF3$, $BAK1$ and $HBS1L-MYB$. In Supplementary Table 1

---

**Table 3** The top SNP rs9376092 GWAS results in the NTR data and TwinsUK data

| Data set        | Alleles(A1/A2) | Frequency A1 | Beta     | s.e.    | P-value |
|-----------------|----------------|--------------|----------|---------|---------|
| NTR ($N = 5091$) | C/A            | 0.72         | PLR      | 5.484   | 0.923   | 2.75E-9 |
|                 |                |              | Platelet | 6.984   | 1.273   | 4.05E-8 |
|                 |                |              | Lymphocytes | -0.039 | 0.0148  | 0.008   |
| Twins UK ($N = 2538$) | C/A         | 0.73         | PLR      | 4.766   | 1.642   | 0.004   |
|                 |                |              | Platelets | 6.053   | 1.942   | 0.002   |
|                 |                |              | Lymphocytes | 0.014   | 0.020   | 0.49    |

Figure 3 Regional plot for the rs9376092 association with PLR level.

Figure 4 (a) Manhattan and (b) QQ plot for the neutrophil count GWAS results with SNPs having a MAF $> 0.01 (\lambda = 1.011742)$. A full color version of this figure is available at the Journal of Human Genetics journal online.
we report the known genetic variants from literature for the three blood cell counts of interest and their significance level as reported previously, together with the P-values obtained from our GWAS study. For neutrophil count we replicated the PSMD3 locus, which also showed an indication of association with PLR (\(P < 1.0 \times 10^{-3}\)). The AK123889 locus showed a similar pattern for PLR (\(P = 3.67 \times 10^{-4}\)), and this locus also had a P-value of 0.001 for lymphocyte count. For lymphocyte count, the known locus rs25224079 was marginally significant (\(P = 3.02 \times 10^{-3}\)), while this locus showed a stronger association with PLR (\(P = 5.65 \times 10^{-5}\)). We did not detect an association for lymphocyte count with the other known locus ESP151 (\(P = 0.107\)). For platelet count, our top hit CCD7C1L-PIK3CG was a replication of earlier studies and it was also associated with mean platelet volume.\(^{11,26}\) We also replicated the loci at ARHGGEF3, BAK1 and HBSIL-MYB, with the latter being associated with PLR as well. Furthermore five loci showed some signal at (\(P < 1.0 \times 10^{-5}\)) for platelet counts: PDLA5, MEF2C, JMD1C, rs7149242 and TAO1K. Other platelet count loci showed some association (\(P < 1.0 \times 10^{-3}\)) with related phenotypes: RCL1, JMD1C, rs7149242 and SNORD7-AP2B1 with PLR, and MICA with lymphocytes.

eQTL effects for significant SNPs

Whole blood cis and trans eQTL analysis was performed for the top significant SNPs per locus identified in the GWAS for PLR (in Table 3) and blood cell counts (in Table 4), with and without correcting for blood composition. The eQTL results are shown in Supplementary Table 2. Information on the function of the genes and the involved pathways was retrieved from the GeneCards website (online accessible: http://www.genecards.org ). Cis eQTLs were found for rs8081692: it increases GSDMB, MSL1 and KRT23 gene expression and decreases GSDMA expression. However, after blood components correction, only GSDMA gene expression was left upregulated by rs8081692. The locus rs169738 was found to increase HLA-DPB1 and decrease TAPBP and HLA-DPA1 expression, which remained after correcting for blood composition. For rs9376090, we detected a significant negative association with ALDH8A1 gene expression, but this SNP is not in LD with the top rs4646871 SNP of the ALDH8A1 gene.

Trans effects for both rs9376090 and rs9376092 were found to increase TMEM158 and HBE1 gene expression, and whereas the trans effects were alleviated when correcting for blood composition, they remained significant. In addition, some eQTLs for genes involved in platelet activation, signaling and aggregation pathways, were present for the uncorrected expression results but disappeared when correcting for blood composition: GNAS (for rs9376090), AQP9 and CREBS5 (for rs8081692). The top SNP rs11925835 nearby ARHGGEF3 gene was found to regulate several sets of genes involved in: (1) platelet activation, signaling and aggregation (ITGB3, PBPB, ITGA2B, PF4, GPR1A, PRKAR2B, CerS25, SELP, THBS1, GNG11, CLU, SPARC, F13A1, VCL, EHD3, CD9, PDGF4, MGLL, GUCY1A3, TXB2AR, MMN1); (2) immune system (TREM1, CD9, CD226); and (3) metabolism (PTGSI, VSG12, EVOVL7, MGLL, ALX012, MFAP3L and NDUFAD3). In addition to these genes, there were several eQTLs for genes that regulate cell division, proliferation, and differentiation such as ABPLM3, LMSM1, ccrf41, FHI1, MAX, RSU1, TSPAN9 and MTPN. Furthermore, some genes play a key role in hematopoietic stem cell differentiation pathways and lineage-specific markers, such as PEAR1 and CD226. For the majority of these genes the effect was alleviated after correction for blood composition. Some trans effects were no longer present after the correction, such as the effects for TPI1, EHD3, PDLM1, MGLL, LMNA, SL2A, ELOVL7, MGLL, TXB2AR, RSU1, MAFAP3L, NEXN, CMTM5, ALX012, PGRM4C, SEPT5, CDK2AP1, CD226, NDUFAF3, MMR1, TSPAN9 and MTPN.

**Table 4 Significant and known loci associated with neutrophil and platelet blood cell count within the NTR study**

| Count | RSNUMBER | CHR | BP     | GENE      | Imputed | Imputation quality R² | MAF | β      | s.e. | P-value |
|-------|----------|-----|--------|-----------|---------|-----------------------|-----|--------|------|---------|
| Neutrophil | rs8081692 | 17  | 38154595 | PSMD3     | yes     | 0.997                 | 0.37| 0.157  | 0.024| 1.77E-10 |
| Platelet | rs342213  | 7   | 106324612| CCDC71L-PIK3CG | yes     | 0.963                 | 0.433| -6.981 | 1.112| 3.45E-10 |
| Platelet | rs169738  | 6   | 33537546 | BAK1      | yes     | 0.993                 | 0.412| -6.824 | 1.151| 3.06E-09 |
| Platelet | rs11925835| 3   | 56865445 | ARHGGEF3  | yes     | 0.984                 | 0.387| -6.457 | 1.147| 1.84E-08 |
| Platelet | rs9376090 | 6   | 135411228| HBSIL-MYB | no      | 0.254                 | 7.296| 1.293  | 1.69E-08| |

SNP heritability and genetic variants for NLR and PLR

The SNP heritability of NLR and PLR was estimated at 2.39% (s.e. = 0.0816) and 14.12% (s.e. = 0.0844), respectively, using LD regression (Table 5). With GCTA the estimated variance explained by the known loci from literature was 0.52% (s.e. = 0.300) for NLR and 3.28% (s.e. = 0.700) for PLR within our study. Finally, the significant

**Table 5 LD regression results for NLR, PLR and the blood cell counts**

|          | NLR   | PLR   | Neutrophil | Lymphocyte | Platelet |
|----------|-------|-------|------------|------------|----------|
| Mean of \(\chi^2\) | 0.997 | 1.014 | 1.009 | 1.021 | 1.034 |
| \(\chi^2\) GC | 0.996 | 1.024 | 1.017 | 1.018 | 1.027 |
| \(H^2\) (s.e.) | 0.0239 (0.0816) | 0.1412 (0.0844) | 0.1537 (0.1011) | 0.1912 (0.0895) | 0.3615 (0.0877) |
| Intercept (s.e.) | 0.9943 (0.0067) | 0.9994 (0.0059) | 0.9932 (0.0081) | 1.0011 (0.0069) | 0.9974 (0.0065) |

Mean of \(\chi^2\): the mean of \(\chi^2\) statistic of markers in the GWAS sample. The difference between this value and the intercept indicates the contribution of polygenicity to the statistical inflation. \(\chi^2\) GC: the lambda genomic control inflation factor. \(H^2\) (s.e.): the estimated SNP heritability based on LD score regression with (standard error). Intercept (s.e.) the regression intercept with (standard error), where the value minus one, indicates the contribution of confounding in the studied sample.
SNPs for PLR, the single SHB1L-MYB region found in our study, explained 0.50% (s.e. = 0.600) of variance.

Significant positive phenotypic correlations were observed between nearly all counts with the exception of NLR and platelet count, and the significantly negatively correlated combinations NLR and lymphocyte count, PLR and neutrophil count, and PLR with lymphocyte count (Table 6). Significant and nearly significant genetic correlations were found between PLR and platelet count ($r = 0.4565$, $P = 0.0309$) and between PLR and lymphocyte count ($r = 0.4858$, $P = 0.0701$). All other genetic correlations were not significant. In Supplementary Table 3 the genetic correlations between the ratios, counts and all available GWAS phenotypes at May 2017 from LD-Hub are presented. There are clear genetic correlations between the consortium platelet count GWAS, and our PLR, platelet counts and neutrophil count. Furthermore, there is a relation between PLR and HDL cholesterol, and for NLR there is a genetic correlation with Crohn’s disease. For the counts separately correlations are present with several diabetes related traits, kidney disease, BMI, (over-)weight, coronary artery disease, autoimmune disease, smoking and lung function assuming a threshold $P$-value of 0.05.

**DISCUSSION**

We studied the genetic architecture of NLR and PLR as well as the genetic relationship between NLR, PLR and the corresponding immune cell counts. The intergenic HBS1L-MYB region is a well-known locus for hematological parameters such as red blood cell count, platelet count, hemoglobin level, MCHC level and blood-related diseases such as myeloproliferative neoplasms, beta-thalassemia and sickle cell anemia. We found this intergenic HBS1L-MYB region to be significantly associated with PLR. HBS1L-MYB intergenic variants reduce the transcription factor binding and affect long-range interactions with MYB and MYB expression levels. This region was first identified as a quantitative trait locus controlling fetal hemoglobin level and is associated with iron deficiency anemia, beta-thalassemia and sickle cell disease. The MYB gene encodes a protein with three HTH DNA-binding domains that functions as a transcription regulator. This protein plays an essential role in the regulation of hematopoiesis and lymphocyte differentiation. This gene can be aberrantly expressed, rearranged or undergo translocation in leukemia’s and lymphomas, and is thus considered to be a (proto-) oncogene. The HBSIL (Hsp70 subfamily B suppressor 1-like) gene encodes a member of the GTP-binding elongation factor family. A single-nucleotide polymorphism in exon 1 of HBSIL gene is significantly associated with severity in beta-thalassemia/hemoglobin E as found in a sequencing study and verified in several other studies. Recently, this gene has been associated with several traits, including erythrocyte and platelet count, and cholesterol level. A pleiotropic association study on a wide number of hematological traits found that rs9373124, also in the HBS1L-MYB region, was significantly associated with all of the evaluated hematological traits ($P < 0.005$), including white blood cell count, red blood cell count and platelet count.

The eQTL results show that some of the GWAS top SNPs for PLR, NLR and blood cell counts regulate the expression of genes, which are mainly involved in immune system pathways: platelet activation, signaling and aggregation; metabolism; cell division, proliferation and differentiation; and hematopoietic stem cell differentiation pathways and lineage-specific markers. These results provide new genetic targets for immune biomarkers and may inform future functional studies. In our GWAS study, SNPs with significant associations for NLR were not identified, which is consistent with the small SNP heritability found with LD regression analyses. Compared with PLR, NLR shows more phenotypic plasticity, because neutrophils are part of the immune response to viral infections, autoimmune diseases, acute-phase reactions and several drugs. Furthermore, compared with the longer lifespan of platelets (8–9 days), the lifespan of neutrophils is shorter (a few hours to max 5 days). The phenotype is therefore much more dependent on environmental effects, for example, the time of measurement and health state of the individual also indicated by our own heritability findings.

By selecting only healthy individuals, we may have excluded the identification of genetic loci that are related to both, disorders and an
**CONFLICT OF INTEREST**

The authors declare no conflict of interest.

**ACKNOWLEDGEMENTS**

This work was supported by: Genotype/phenotype database for genetic studies (ZonMW Middelroot (911-09-032); Database Twin register (NWO 575-25-006); Twin family database for behavior genetics and genomics studies (NWO 480-04-004); Genome-wide analyses of European twin and population cohorts (EU/QLRPT-2001-01254); Center for Medical Systems Biology (CMSB), Biobanking and Bimolecular Resources Research Infrastructure (BBMRI-NL) 184.021.007; GENOMEUTWIN/EU (QLG2-CT-2002-01254); NIH (NIHHEALTHF4-2007-201413); European Research Council (230374-GMI). TwinsUK is funded by the Wellcome Trust, Medical Research Council, European Union (EU) and the National Institute for Health Research (NIHR)-funded BioResource, Clinical Research Facility, and Biomedical Research Centre based at Guy’s and St Thomas’ NHS Foundation Trust in partnership with King’s College London. B Lin received a PhD grant (201206180099) from the China Scholarship Council. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed. No writing assistance was utilized in the production of this manuscript. We thank all the twin families who participated in the Netherlands Twin Register Biobank project and TwinsUK project.

1. Durmus, E., Kivrak, T., Gerin, F., Sunbul, M., Sari, I. & Erdogan, O. Neutrophil-to-Lymphocyte Ratio and Platelet-to-Lymphocyte Ratio are predictors of heart failure. Arq. Bras. Cardiol. 105, 606–613 (2015).
2. Lee, S., Oh, S. Y., Kim, S. H., Lee, J. H., Kim, M. C., Kim, K. H. et al. Prognostic significance of neutrophil lymphocyte ratio and platelet lymphocyte ratio in advanced gastric cancer patients treated with FOLFOX chemotherapy. BMC Cancer 13, 350 (2013).
3. Wang, X. D., Zhang, G. Y., Jiang, X. J., Zhu, H. L., Liu, Z. & Xu, L. Neutrophil to lymphocyte ratio in relation to risk of all-cause mortality and cardiovascular events among patients undergoing angiography or cardiac revascularization: a meta-analysis of observational studies. Atherosclerosis 234, 206–213 (2014).
4. He, J. Y., Li, J., Wang, Y. F., Hao, P. & Hua, Q. Neutrophil-to-lymphocyte ratio (NLR) predicts mortality and adverse-outcomes after ST-segment elevation myocardial infarction in Chinese people. Int. J. Exp. Pathol. 7, 4045–4056 (2014).
5. Kang, M. H., Go, S. I., Song, H. N., Lee, A., Kim, S. H., Kang, J. H. et al. The prognostic impact of the neutrophil-to-lymphocyte ratio in patients with small-cell lung cancer. Br J Cancer. 111, 452–460 (2014).
6. Templeton, A. J., McNamara, M. G., Vera-Badillo, F. E., Aneja, P., Ocana, A., Seruga, B. et al. Prognostic role of neutrophil to lymphocyte ratio (NLR) in solid tumors: systematic review and meta-analysis. Eur. J. Cancer 49, 5211–5211 (2013).
7. Yodiyong, H., Matsuda, A., Miyashita, M., Matsumoto, S., Sakurazawa, N., Yamada, M. et al. Prognostic significance of neutrophil-to-lymphocyte ratio and platelet-to-lymphocyte ratio in oncologic outcomes of esophageal cancer: a systematic review and meta-analysis. Ann. Surg. Oncol. 23, 646–654 (2016).
8. Zhou, X., Du, Y. P., Huang, Z. B., Xu, J., Guit, T. Z., Wang, J. et al. Prognostic value of PLR in various cancers: a meta-analysis. PloS ONE 9, e0101119 (2014).
A genome-wide meta-analysis identifies new gene functions in megakaryopoiesis and platelet formation.

[10] Lin, B. D., Hottenga, J. J., Abdellaoui, A., Dolan, C. V., de Geus, E. J., Kluit, C. et al. Causes of variation in the neutrophil-lymphocyte and platelet-lymphocyte ratios: a twin-family study. *Biomark. Med.* 10, 1061–1072 (2016).

[11] Soranzo, N., Spector, T. D., Mangino, M., Kuhnle, B., Rondón, A., Teumer, A. et al. Genome-wide meta-analysis identifies 22 loci associated with eight hematological parameters in the HaemGen consortium. *Nat. Genet.* 41, 1182–1190 (2009).

[12] Gieger, C., Radhakrishnan, A., Cvejic, A., Tang, W., Porcu, E., Pitsis, G. et al. New gene functions in megakaryopoiesis and platelet formation. *Nature Genet.* 48, 201–208 (2011).

[13] Shameer, K., Denny, J. C., Ding, K. Y., Jouni, H., Crosslin, D. R., de Andrade, M. et al. A genome- and phenotype-wide association study to identify genetic variants influencing platelet count and volume and their pleiotropic effects. *Hum. Genet.* 133, 95–104 (2014).

[14] Schick, U. M., Jain, D., Hodosky, C. J., Morrison, J. V., Davis, J. P., Brown, L. et al. Genome-wide association study of platelet count identifies ancestry-specific loci in Hispanic/Latino Americans. *Am. J. Hum. Genet.* 98, 229–242 (2016).

[15] Reiner, A. P., Lettre, G., Nalls, M. A., Ganesh, S. K., Mathias, R., Austin, M. A. et al. Genome-wide association study of white blood cell count in the health ABC and Jackson Heart Studies. *Am. J. Hum. Genet.* 98, 430–437 (2011).

[16] Qayyum, R., Becker, D. M., Faraday, N., Yanek, L. R., Leal, S. M. et al. Genetic association analysis highlights new loci that modulate hematological function. *PLoS Genet.* 7, e1001397 (2011).

[17] Kim, Y. K., Oh, J. H., Kim, Y. J., Hwang, M. Y., Moon, S., Low, S. K. et al. Genetic association analysis highlights new loci that modulate hematological function. *J. Clin. Invest.* 127, 1699–1710 (2014).
Thein, S. L., Menzel, S., Peng, X., Best, S., Jiang, J., Close, J. et al. Intergenic variants of HBS1L-MYB are responsible for a major quantitative trait locus on chromosome 6q23 influencing fetal hemoglobin levels in adults. Proc. Natl Acad. Sci. USA 104, 11346–11351 (2007).

Roy, P., Bhattacharya, G., Mandal, A., Dasgupta, U. B., Banerjee, D., Chandra, S. et al. Influence of BCL11A, HBS1L-MYB, HBBP1 single nucleotide polymorphisms and the HBG2 XMNI polymorphism on Hb F levels. Hemoglobin 36, 592–599 (2012).

Le Guen, T., Touzot, F., Andre-Schmutz, I., Lagresle-Peyrou, C., France, B., Kermasson, L. et al. An in vivo genetic reversion highlights the crucial role of MYB-Like, SWIRM, and MPN domains 1 (MYSM1) in human hematopoiesis and lymphocyte differentiation. J. Allergy Clin. Immunol. 136, 1619–1626 (2015).

Mets, E., Van der Meulen, J., Van Peer, G., Boice, M., Mestdagh, P., Van de Walle, I. et al. MicroRNA-193b-3p acts as a tumor suppressor by targeting the MYB oncogene in T-cell acute lymphoblastic leukemia. Leukemia 29, 798–806 (2015).

Srivastava, S. K., Bhardwaj, A., Arora, S., Singh, S., Azim, S., Tyagi, N. et al. MYB is a novel regulator of pancreatic tumour growth and metastasis. Br. J. Cancer 113, 1694–1703 (2015).

Sripichai, O., Whitacre, J., Munkongdee, T., Kumkhaek, C., Makarasara, W., Winichagoon, P. et al. Genetic analysis of candidate modifier polymorphisms in Hb E-beta O-thalassemia patients. Ann. N Y Acad. Sci. 1054, 433–438 (2005).

Danjou, F., Francavilla, M., Arni, F., Satta, S., Demartis, F. R., Perseu, L. et al. A genetic score for the prediction of beta-thalassemia severity. Haematologica 100, 452–457 (2015).

Pandit, R. A., Svasti, S., Sripichai, O., Munkongdee, T., Trinitsaysakorn, K., Winichagoon, P. et al. Association of SNP in exon 1 of HBS1L with hemoglobin F level in beta0-thalassemia/hemoglobin E. Int. J. Hematol. 88, 357–361 (2008).

Farrell, J. J., Sherva, R. M., Chen, Z. Y., Luo, H. Y., Chu, B. F., Ha, S. Y. et al. A 3-bp deletion in the HBS1L-MYB intergenic region on chromosome 6q23 is associated with HbF expression. Blood 117, 4935–4945 (2011).

Willer, C. J., Schmidt, E. M., Sengupta, S., Peloso, G. M., Gustafsson, S., Kanoni, S. et al. Discovery and refinement of loci associated with lipid levels. Nat. Genet. 45, 1274–1283 (2013).

Wright, H. L., Moots, R. J., Bucknall, R. C. & Edwards, S. W. Neutrophil function in inflammation and inflammatory diseases. Rheumatology (Oxford) 49, 1618–1631 (2010).

Pillay, J., den Braber, I., Vrisekoop, N., Kwast, L. M., de Baer, R. J., Borghans, J. A. et al. In vivo labeling with 2H2O reveals a human neutrophil lifespan of 5.4 days. Blood 116, 625–627 (2010).

Tak, T., Tessleraer, K., Pillay, J., Borghans, J. A. & Koenderman, L. What's your age again? Determination of human neutrophil half-lives revisited. J. Leukoc. Biol. 94, 595–601 (2013).

Donababu, P., Nagesh, N., Linga, V. G., Gundeti, S., Kutala, V. K., Reddanna, P. et al. Epistatic interactions between thiopurine methyltransferase (TPMT) and inosine triphosphatase pyrophosphatase (ITPA) variations determine 6-mercaptopurine toxicity in Indian children with acute lymphoblastic leukemia. Eur. J. Clin. Pharmacol. 68, 379–387 (2012).

Supplementary Information accompanies the paper on Journal of Human Genetics website (http://www.nature.com/jhg)