



SNP 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 Geus^{1,3}, Rick Jansen⁴, Cornelis Klufft⁵, Massimo Mangino^{2,6}, Brenda Penninx⁴, Tim D. Spector², Gonneke Willemsen¹, and Jouke-Jan Hottenga^{1,3,*}

¹Department of Biological Psychology, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands ²Department of Twin Research and Genetic Epidemiology, Kings College London, London SE1 7EH, 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 ⁶NIHR Biomedical Research Centre at Guy's and St Thomas' Foundation Trust, London SE1 9RT, UK

Abstract

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-PIK3CG*, *BAK1* and *ARHGEF3*). For neutrophil count, we replicated the *PSMD3* locus. For the identified top SNPs, we found significant cis and trans eQTL 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 polygenetic 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.

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

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use: http://www.nature.com/authors/editorial_policies/license.html#terms

Corresponding author. JJ Hottenga; j.j.hottenga@vu.nl, Mobile: +31657778267, Work: +31205988982, FAX: +31205988832.

Conflict of interest Statement
None declared.

diseases. A high NLR level was shown to be an independent predictor of mortality in patients undergoing cardiac revascularization and in patients with myocardial infarction⁴. Elevated NLR levels were also related to a poor prognosis of various cancers, such as esophageal, pancreatic, lung, ovarian and hepatocellular cancer. Similar to NLR, PLR was also reported as an index for diagnosis or prognostic prediction of oncologic disorders and inflammatory diseases. 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 due to genetic influences, with a broad sense heritability of 36% for NLR and 64% for PLR, using a twin-family epidemiological design¹⁰. Here, we investigate if the significant heritability estimates can be explained by common SNPs (single nucleotide polymorphisms) 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 (i.e. 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, African-American, Korean and Japanese populations. 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 populations and loci at 20p12 (*PLCB4* gene)²² and 7q21.2 (*CDK6* gene) 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 *AKI23889* on 6p21.33 were novel findings in a European ancestry cohort, and were also confirmed by meta-analysis. For lymphocyte count, two genetic variants nearby *EPS15L1* gene on 6p21 and *LOC101929772* on 19p13 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, *BMPRIA* on 10q23.2, loci on 6p22, 7q11, 10q21, 11q13, 20q13 were detected in the African-American population and over 55 loci including *CCDC71L-PIK3CG*, *ARHGEF3*, *BAK1* and *HBSIL-MYB* in the European population.

Some blood cell count loci show pleiotropy: they influence multiple hematological indices. For example, the genetic region nearby *AKI23889* 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 *HBSIL-MYB* variants were associated with total white blood cell count 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 participants³⁴. Furthermore, all top SNPs, which showed a significant association with our phenotypes of interest, were selected for an 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 LD regression to determine the variance explained by all autosomal SNPs, to examine polygenetic effects between NLR and PLR, and to determine the genetic correlation between variants affecting the two ratios, their subcomponents and LD-Hub published GWAS⁴⁰.

Material 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 $\pm 5 \times SD$ 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/L (N=287); 3) basophile count $> 0.02 \times 10^9/L$ (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 one hour 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 to 6 hours 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, 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=1,718), Illumina 370 (N=424), Illumina 660 (N=1,103), Illumina Omni Express 1 M (N=346) and Affymetrix 6.0 (N=3602). Genotype calls were made with the platform specific software (Birdsuite, APT-Genotyper, 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 SVS 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 more than 0.20 compared to the reference. From each platform, SNPs were also excluded if meeting the following criteria: a Minor Allele Frequency (MAF) <1%, Hardy-Weinberg Equilibrium (HWE) 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 PLINK1.07 F inbreeding value was either >0.10 or <-0.10.

After these steps, the data of the individual arrays were merged into a single dataset using PLINK 1.07⁴². Within the merged set, identity by state (IBS) sharing was calculated between all possible pairs of participants and compared to the known NTR family structures. Samples were removed if the data did not match their expected IBS sharing. The concordance rate of DNA samples on multiple platforms for overlapping SNPs generally exceeded 99.0% after data cleaning. The HWE, 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 MAF was between 0.35 and 0.50 to avoid incorrect strand alignment. Phasing of all samples and imputing cross-missing platform SNPs was done with MACH 1⁴³. The phased data were then imputed with MINIMAC⁴⁴ in batches of around 500 individuals for the autosomal genome using the above 1000G Phase I integrated reference panel for 561 chromosome chunks obtained by the CHUNKCHROMOSOME program. To avoid issues having SNPs from different platforms partly imputed and partly genotyped we took the re-imputed calls for all genotyped SNPs. After imputation of these SNPs, we generally find a high concordance between re-imputed SNPs and the original genotype (0.9868). The mean imputation quality R^2 metric is 0.38 (based on all 30,051,533 imputed autosomal SNPs). After imputation, SNPs were filtered based on the Mendelian error rate in families, which was calculated from the best guess genotypes in families (trios or sib-pairs with parents) using first GTOOL to calculate best guess genotypes and then PLINK 1.07 to analyze the data. SNPs were removed if the Mendelian error rate >0.02, the imputed allele frequency differed more than 0.15 from the 1000G reference allele frequency, $MAF < 0.01$ or $R^2 < 0.80$. HWE was calculated on the genotype 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

Genetic Relatedness Matrices (GRMs) with the values of the identity by state (IBS) allele sharing for a given set of SNP markers between all possible pairs of individuals were calculated with the GCTA software⁴⁶, after removing SNPs that showed significant genotyping differences between platforms ($p < 0.0001$); 6,009,498 SNPs were retained, which is sufficient for GRM estimation⁴⁶. 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 ($IBS > 0.05$), 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 matrixes 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 Leave One Chromosome Out (LOCO) 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 (PCs) as were generated with the EIGENSOFT software were used as covariates in the GWAS⁹. 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 (MLMA) including the LOCO GRMs for chromosome 1 to 22, and the close-related GRM⁵⁰. For the GWAs, the significance threshold was $p\text{-value} < 5 \times 10^{-8}$.

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 United Kingdom based twin registry with a focus on the genetic and environmental etiology of age related complex traits and diseases³⁴. 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 algorithm⁵². No calls were assigned if the most likely call was less than a posterior probability of 0.95. SNPs were excluded if they that had a low call rate ($< 95\%$) and/or Hardy-Weinberg P value $< 10^{-4}$. Subjects were also removed if the sample call rate was less than 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 IMPUTE⁵³. The best guess Plink binary format data was used to conduct the replication analysis. The sample size of the

TwinsUK dataset was 2538 subjects with genetic and phenotypic information, after values outside $\text{mean} \pm 5\text{SD}$ 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, while 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>)⁴. 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 *MatruxeQTL*⁵⁵ 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 $< 1\text{M}$ 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. Since 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, basophile 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 software⁴⁶. 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 (REML) analysis procedure within GCTA. Sex, age, genotype platform and three Dutch PCs were used as covariates. The variance explained by all SNPs was estimated by Linkage Disequilibrium (LD) regression between our computed GWAS summary statistics effect sizes and the expected Hapmap 3 LD. In order 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: <http://github.com/bulik/ldsc>). The criteria of passing quality control for SNPs were the default by LD regression: imputation quality $\text{info} > 0.90$, $\text{MAF} > 0.01$. SNPs with invalid P values ($P \geq 1$ or $P < 0$) were excluded. In addition, variants that are not SNPs (e.g., 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^{38–39}. 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-Hub⁴⁰. 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 analyses⁴⁰. 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 *HBSIL* and *MYB* genes on chromosome 6q23.3 in the *HBSIL-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 ($\beta=5.48$, $p=2.75 \times 10^{-9}$). This SNP was also significantly associated with platelet count ($\beta=6.98$, $p=4.05 \times 10^{-8}$), but not with lymphocyte count ($\beta=-.039$, $p=0.008$). In the TwinsUK sample, rs9376092 replicated with a similar effect for PLR ($\beta=4.766$, $p=0.004$) as well as platelet count ($\beta=6.053$, $p=0.002$). 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 *HBSIL-MYB*. In supplementary Table 1 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 ($p = 3.02 \times 10^{-4}$), while this locus showed a stronger association with PLR ($p = 6.56 \times 10^{-5}$). We did not detect an association for lymphocyte count with the other known locus *ESP15LI* ($p = 0.107$). For platelet count, our top hit *CCDC71L-PIK3CG* was a replication of earlier studies and it was also associated with mean platelet volume. We also replicated the loci at *ARHGEF3*, *BAK1* and *HBSIL-MYB*, with the latter being associated with PLR as well. Furthermore five loci showed some signal at ($p < 1.0 \times 10^{-3}$) for platelet counts: *PDIA5*, *MEF2C*, *JMD1C*, rs7149242 and *TAOK1*. Other platelet count loci showed some association ($p < 1.0 \times 10^{-3}$) with related phenotypes: *RCLI*, *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 effects 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 up regulated 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 while 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 *CREB5* (for rs8081692). The top SNP rs11925835 nearby *ARHGEF3* gene was found to regulate several sets of genes involved in: 1) platelet activation, signaling and aggregation (*ITGB3*, *PPBP*, *ITGA2B*, *PF4*, *GP1BA*, *PRKAR2B*, *C6orf25*, *SELP*, *THBS1*, *GNG11*, *CLU*, *SPARC*, *F13A1*, *VCL*, *EHD3*, *CD9*, *PDGFA*, *MGLL*, *GUCY1A3*, *TBXA2R*, *MMRN1*); 2) immune system (*TREML1*, *CD9*, *CD226*); and 3) metabolism (*PTGS1*, *VSIG2*, *EVOVL7*, *MGLL*, *ALOX12*, *MFAP3L*, and *NDUFAF3*). In addition to these genes, there were several eQTLs for genes that regulate cell division, proliferation, and differentiation such as *ABLIM3*, *LMSM1*, *c7orf41*, *FHL1*, *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 *TPM1*, *EHD3*, *PDLIM1*, *MGLL*, *LMNA*, *SLA2*, *ELOVL7*, *MGLL*, *TBXA2R*, *RSU1*, *MFAP3L*, *NEXN*, *CMTM5*, *ALOX12*, *PGRMC1*, *SEPT5*, *CDK2AP1*, *CD226*, *NDUFAF3*, *MMRN1*, *TSPAN9*, and *MTPN*.

SNP heritability and correlations among phenotypes

The SNP heritability of NLR and PLR was estimated at 2.39% (se = 0.0816) and 14.12% (se = 0.0844) respectively using LD regression (Table 5). With GCTA the estimated variance explained by the known loci from literature was 0.52% (se = 0.300) for NLR and 3.28% (se = 0.700) for PLR within our study. Finally, the significant SNPs for PLR, the single *SHBIL-MYB* region found in our study, explained 0.50% (se = 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 ($r=-0.4858$, $p=0.0701$) (Table 6). 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 *HBSIL-MYB* region is a well-known locus for hematological parameters such as red blood cell count, platelet count, hemoglobin level⁷, MCHC level, and blood related diseases such as myeloproliferative neoplasms⁹, beta-thalassemia⁶⁰ and sickle cell anaemia. We found this intergenic *HBSIL-MYB* region to be significantly associated with PLR. *HBSIL-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 (QTL) controlling fetal hemoglobin level and is associated with iron deficiency anemia, beta-thalassemia, and sickle cell disease³. 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 count⁷⁰ and cholesterol level. A pleiotropic association study on a wide number of hematological traits found that rs9373124, also in the *HBSIL-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 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 to PLR, NLR shows more phenotypic plasticity, because neutrophils are part of the immune response to viral infections, autoimmune diseases, acute-

phase reactions and several drugs⁷². Furthermore, compared to the longer lifespan of platelets (8–9 days), the life span of neutrophils is shorter (a few hours to max 5 days). The phenotype is therefore much more dependent on environmental effects, e.g. the time of measurement and health state of the individual also indicated by our own heritability findings¹⁰.

By selecting only healthy individuals, we may have excluded the identification of genetic loci that are related to both, disorders and an effect on the counts or ratios. With this particular step it is also unclear whether the identified SNPs are affecting the risk for the disorders for which PLR and NLR are supposed to be the biomarkers. Another question is if the effect of the SNPs on the ratios and immune response remains the same once a person gets affected. We have examined the heritability of the ratios in our full sample of individuals, and the point estimates were not different. Although, there is currently no direct association between the top SNP or the *HBS1L* locus and myocardial infarction in the GWAS catalog, there are clear genetic correlations between cardiovascular traits, diabetic traits, HDL cholesterol, weight, smoking, lung function and the examined counts and ratios. The link between *HBS1L* and cancer was already established due to translocations in the genomic region.

For both NLR and PLR a large part of the heritability is still not explained by common SNPs or genetic variants in LD. This may suggest that other genetic variants, such as rare variants and copy number variants need to be studied. Furthermore, the missing heritability might be high because of non-additive effects and genetic interactions, which are not taken into account with the current applied statistical models. Epistatic effects of genetic variants for hematological indexes are already found⁵. We thus assume that, especially for immune system phenotypes, gene-gene and gene-environment interactions need to be studied further.

The LD regression results show that polygenic effects, rather than confounding factors explain NLR and PLR variance in our study. We also demonstrated significant genetic correlation between PLR and platelet count, but none of the other correlations between ratios and cell counts were large enough to be significant. Since we found no SNP effects on NLR, it is not surprising that no genetic overlap between NLR and PLR is detected, although the genetic background of the lymphocyte count is expected to be affecting both ratios.

In summary, our study found the *HBS1L-MYB* locus to be associated with PLR level and with platelet count. In addition, we verified 3 additional known loci for platelet count (rs342213 in *CCDC71L-PIK3CG*, rs169738 nearby *BAK1* and rs11925835 nearby *ARHGEF3*) and one locus for neutrophil count (rs8081692 nearby *PSMD3*). We did not identify any locus or any significant SNP heritability for NLR. Although NLR and PLR are both utilized as predictive or prognostic biomarkers for the same diseases, and phenotypic correlations are present, there seems to be no major genetic overlap between the two biomarkers in our healthy population. The NLR and PLR responses associated with the disorders, thus likely represent the simultaneous influence of separate and multiple immune genetic pathways.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by: Genotype/phenotype database for genetic studies (ZonMW Middelgroot [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/QLRT-2001-01254); Center for Medical Systems Biology (CMSB), Biobanking and Biomolecular 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.

The authors thank all the twin families who participated in The Netherlands Twin Register Biobank project and TwinsUK project.

References

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.* 2015 [PubMed: 26536980]
2. Lee S, Oh SY, Kim SH, Lee JH, Kim MC, Kim KH, et al. Prognostic significance of neutrophil lymphocyte ratio and platelet lymphocyte ratio in advanced gastric cancer patients treated with FOLFOX chemotherapy. *Bmc Cancer.* 2013; 13:350. [PubMed: 23876227]
3. Wang XD, Zhang GY, Jiang XJ, Zhu HL, Lu 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.* 2014; 234:206–213. [PubMed: 24681815]
4. He JY, Li J, Wang YF, 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 Clin Exp Pathol.* 2014; 7:4045–4056.
5. Kang MH, Go SI, Song HN, Lee A, Kim SH, Kang JH, Jeong BK, Kang KM, Ling H, Lee GW. The prognostic impact of the neutrophil-to-lymphocyte ratio in patients with small-cell lung cancer. *Br J Cancer.* 2014; 111:452–460. [PubMed: 24921916]
6. Templeton AJ, McNamara MG, Vera-Badillo FE, Aneja P, Ocana A, Seruga B, et al. Prognostic role of neutrophil to lymphocyte ratio (NLR) in solid tumors: A systematic review and meta-analysis. *Eur J Cancer.* 2013; 49:S211–S211.
7. Yodying 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.* 2016; 23:646–654. [PubMed: 26416715]
8. Zhou X, Du YP, Huang ZB, Xu J, Qiu TZ, Wang J, et al. Prognostic value of PLR in various cancers: a meta-analysis. *Plos One.* 2014; 9:e0101119.
9. Turkmen K, Erdur FM, Ozcicek F, Ozcicek A, Akbas EM, Ozbicer A, et al. Platelet-to-lymphocyte ratio better predicts inflammation than neutrophil-to-lymphocyte ratio in end-stage renal disease patients. *Hemodial Int.* 2013; 17:391–396. [PubMed: 23522328]

10. Lin BD, Hottenga JJ, Abdellaoui A, Dolan CV, de Geus EJ, Klufft C, et al. Causes of variation in the neutrophil-lymphocyte and platelet-lymphocyte ratios: a twin-family study. *Biomark Med.* 2016;1061–1072.
11. Soranzo N, Spector TD, Mangino M, Kuhnel B, Rendon A, Teumer A, et al. A genome-wide meta-analysis identifies 22 loci associated with eight hematological parameters in the HaemGen consortium. *Nat Genet.* 2009; 41:1182–1190. [PubMed: 19820697]
12. Gieger C, Radhakrishnan A, Cvejic A, Tang W, Porcu E, Pistis G, et al. New gene functions in megakaryopoiesis and platelet formation. *Nature.* 2011; 480:201–208. [PubMed: 22139419]
13. Shameer K, Denny JC, Ding KY, Jouni H, Crosslin DR, de Andrade M, et al. A genome- and phenome-wide association study to identify genetic variants influencing platelet count and volume and their pleiotropic effects. *Hum Genet.* 2014; 133:95–109. [PubMed: 24026423]
14. Schick UM, Jain D, Hodonsky CJ, Morrison JV, Davis JP, Brown L, et al. Genome-wide association study of platelet count identifies ancestry-specific loci in Hispanic/Latino Americans. *Am J Hum Genet.* 2016; 98:229–242. [PubMed: 26805783]
15. Reiner AP, Lettre G, Nalls MA, Ganesh SK, Mathias R, Austin MA, et al. Genome-Wide Association Study of White Blood Cell Count in 16,388 African Americans: the Continental Origins and Genetic Epidemiology Network (COGENT). *Plos Genet.* 2011; 7:e1002108. [PubMed: 21738479]
16. Qayyum R, Snively BM, Ziv E, Nalls MA, Liu Y, Tang W, et al. A meta-analysis and genome-wide association study of platelet count and mean platelet volume in African Americans. *Plos Genet.* 2012; 8:e1002491. [PubMed: 22423221]
17. Qayyum R, Becker LC, Becker DM, Faraday N, Yanek LR, Leal SM, et al. Genome-wide association study of platelet aggregation in African Americans. *BMC Genet.* 2015; 16:58. [PubMed: 26024889]
18. Nalls MA, Couper DJ, Tanaka T, van Rooij FJ, Chen MH, Smith AV, et al. Multiple loci are associated with white blood cell phenotypes. *Plos Genet.* 2011; 7:e1002113. [PubMed: 21738480]
19. Kim YK, Oh JH, Kim YJ, Hwang MY, Moon S, Low SK, et al. Influence of genetic variants in EGF and other genes on hematological traits in Korean populations by a genome-wide approach. *Biomed Res Int.* 2015; 2015:914965. [PubMed: 26064965]
20. Oh JH, Kim YK, Moon S, Kim YJ, Kim BJ. Genome-wide association study identifies candidate Loci associated with platelet count in Koreans. *Genomics Inform.* 2014; 12:225–230. [PubMed: 25705162]
21. Okada Y, Hirota T, Kamatani Y, Takahashi A, Ohmiya H, Kumasaka N, et al. Identification of nine novel loci associated with white blood cell subtypes in a Japanese population. *Plos Genet.* 2011; 7:e1002067. [PubMed: 21738478]
22. Okada Y, Kamatani Y, Takahashi A, Matsuda K, Hosono N, Ohmiya H, et al. Common variations in PSMD3-CSF3 and PLCB4 are associated with neutrophil count. *Hum Mol Genet.* 2010; 19:2079–2085. [PubMed: 20172861]
23. Nalls MA, Wilson JG, Patterson NJ, Tandon A, Zmuda JM, Huntsman S, et al. Admixture mapping of white cell count: Genetic locus responsible for lower white blood cell count in the health ABC and Jackson Heart Studies. *Am J Hum Genet.* 2008; 82:81–87. [PubMed: 18179887]
24. Keller MF, Reiner AP, Okada Y, van Rooij FJ, Johnson AD, Chen MH, et al. Trans-ethnic meta-analysis of white blood cell phenotypes. *Hum Mol Genet.* 2014; 23:6944–6960. [PubMed: 25096241]
25. Li J, Glessner JT, Zhang H, Hou C, Wei Z, Bradfield JP, et al. GWAS of blood cell traits identifies novel associated loci and epistatic interactions in Caucasian and African-American children. *Hum Mol Genet.* 2013; 22:1457–1464. [PubMed: 23263863]
26. Soranzo N, Rendon A, Gieger C, Jones CI, Watkins NA, Menzel S, et al. A novel variant on chromosome 7q22.3 associated with mean platelet volume, counts, and function. *Blood.* 2009; 113:3831–3837. [PubMed: 19221038]
27. Ganesh SK, Zakai NA, van Rooij FJA, Soranzo N, Smith AV, Nalls MA, et al. Multiple loci influence erythrocyte phenotypes in the CHARGE Consortium. *Nat Genet.* 2009; 41:1191–U1148. [PubMed: 19862010]

28. Kamatani Y, Matsuda K, Okada Y, Kubo M, Hosono N, Daigo Y, et al. Genome-wide association study of hematological and biochemical traits in a Japanese population. *Nat Genet.* 2010; 42:210–215. [PubMed: 20139978]
29. Gudbjartsson DF, Bjornsdottir US, Halapi E, Helgadóttir A, Sulem P, Jonsdóttir GM, et al. Sequence variants affecting eosinophil numbers associate with asthma and myocardial infarction. *Nat Genet.* 2009; 41:342–347. [PubMed: 19198610]
30. Lo KS, Wilson JG, Lange LA, Folsom AR, Galarnau G, Ganesh SK, et al. Genetic association analysis highlights new loci that modulate hematological trait variation in Caucasians and African Americans. *Hum Genet.* 2011; 129:307–317. [PubMed: 21153663]
31. Menzel S, Qin JA, Vasavda N, Thein SL, Ramakrishnan R. Experimental generation of SNP haplotype signatures in patients with sickle cell anaemia. *Plos One.* 2010; 5 [PubMed: 20886046]
32. Sirota M, Willemsen G, Sundar P, Pitts SJ, Potluri S, Prifti E, et al. Effect of genome and environment on metabolic and inflammatory profiles. *Plos One.* 2015; 10:e0120898. [PubMed: 25853885]
33. Willemsen G, de Geus EJ, Bartels M, van Beijsterveldt CE, Brooks AI, Estourgie-van Burk GF, et al. The Netherlands Twin Register biobank: a resource for genetic epidemiological studies. *Twin Res Hum Genet.* 2010; 13:231–245. [PubMed: 20477721]
34. Andrew T, Hart DJ, Snieder H, de Lange M, Spector TD, MacGregor AJ. Are twins and singletons comparable? A study of disease-related and lifestyle characteristics in adult women. *Twin Res.* 2001; 4:464–477. [PubMed: 11780939]
35. Wright FA, Sullivan PF, Brooks AI, Zou F, Sun W, Xia K, et al. Heritability and genomics of gene expression in peripheral blood. *Nat Genet.* 2014; 46:430–437. [PubMed: 24728292]
36. Zaitlen N, Kraft P, Patterson N, Pasaniuc B, Bhatia G, Pollack S, et al. Using Extended Genealogy to Estimate Components of Heritability for 23 Quantitative and Dichotomous Traits. *Plos Genet.* 2013; 9 [PubMed: 23737753]
37. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet.* 2011; 88:76–82. [PubMed: 21167468]
38. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh PR, et al. An atlas of genetic correlations across human diseases and traits. *Nat Genet.* 2015; 47:1236–1241. [PubMed: 26414676]
39. Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, Patterson N, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet.* 2015; 47:291–295. [PubMed: 25642630]
40. Zheng J, Erzurumluoglu AM, Elsworth BL, Kemp JP, Howe L, Haycock PC, et al. LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics.* 2017 Jan 15; 33(2):272–279.
41. Boomsma DI, Willemsen G, Sullivan PF, Heutink P, Meijer P, Sondervan D, et al. Genome-wide association of major depression: description of samples for the GAIN Major Depressive Disorder Study: NTR and NESDA biobank projects. *Eur J Hum Genet.* 2008; 16:335–342. [PubMed: 18197199]
42. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007; 81:559–575. [PubMed: 17701901]
43. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet Epidemiol.* 2010; 34:816–834. [PubMed: 21058334]
44. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *Plos Genet.* 2009; 5 [PubMed: 19543373]
45. Liu EY, Li M, Wang W, Li Y. MaCH-admix: genotype imputation for admixed populations. *Genet Epidemiol.* 2013; 37:25–37. [PubMed: 23074066]
46. Yang JA, Lee SH, Goddard ME, Visscher PM. GCTA: A tool for genome-wide complex trait analysis. *Am J Hum Genet.* 2011; 88:76–82. [PubMed: 21167468]

47. Yang J, Zaitlen NA, Goddard ME, Visscher PM, Price AL. Advantages and pitfalls in the application of mixed-model association methods. *Nature Genetics*. 2014; 46:100–106. [PubMed: 24473328]
48. Abdellaoui A, Hottenga JJ, de Knijff P, Nivard MG, Xiao XJ, Scheet P, et al. Population structure, migration, and diversifying selection in the Netherlands. *Eur J Hum Genet*. 2013; 21:1277–1285. [PubMed: 23531865]
49. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*. 2006; 38:904–909. [PubMed: 16862161]
50. Tucker G, Loh PR, MacLeod IM, Hayes BJ, Goddard ME, Berger B, et al. Two-Variance-Component Model Improves Genetic Prediction in Family Datasets. *Am J Hum Genet*. 2015; 97:677–690. [PubMed: 26544803]
51. Pe'er I, Yelensky R, Altshuler D, Daly MJ. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet Epidemiol*. 2008; 32:381–385. [PubMed: 18348202]
52. Teo YY, Inouye M, Small KS, Gwilliam R, Deloukas P, Kwiatkowski DP, et al. A genotype calling algorithm for the Illumina BeadArray platform. *Bioinformatics*. 2007; 23:2741–2746. [PubMed: 17846035]
53. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet*. 2012; 44:955–959. [PubMed: 22820512]
54. Jansen R, Hottenga JJ, Nivard M, Abdellaoui A, Laport B, de Geus EJ, et al. Conditional eQTL Analysis Reveals Allelic Heterogeneity of Gene Expression. *Human Molecular Genetics*. in press. [PubMed: 28165122]
55. Shabalin AA. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics*. 2012; 28:1353–1358. [PubMed: 22492648]
56. van der Harst P, Zhang WH, Leach IM, Rendon A, Verweij N, Sehmi J, et al. Seventy-five genetic loci influencing the human red blood cell. *Nature*. 2012; 492:369.
57. Lettre G, Sankaran VG, Bezerra MAC, Araujo AS, Uda M, Sanna S, et al. DNA polymorphisms at the BCL11A, HBS1L-MYB, and beta-globin loci associate with fetal hemoglobin levels and pain crises in sickle cell disease. *P Natl Acad Sci USA*. 2008; 105:11869–11874.
58. Pistis G, Okonkwo SU, Traglia M, Sala C, Shin SY, Masciullo C, et al. Genome Wide Association Analysis of a Founder Population Identified TAF3 as a Gene for MCHC in Humans. *Plos One*. 2013; 8 [PubMed: 23935956]
59. Tapper W, Jones AV, Kralovics R, Harutyunyan AS, Zoi K, Leung W, et al. Genetic variation at MECOM, TERT, JAK2 and HBS1L-MYB predisposes to myeloproliferative neoplasms. *Nat Commun*. 2015; 6 [PubMed: 25849990]
60. He YY, Chen P, Lin WX, Luo JM. Analysis of the rs35959442 polymorphism in Hb E/beta-thalassemia in Guangxi province of the Republic of China. *Hemoglobin*. 2012; 36:166–169. [PubMed: 22191716]
61. Stadhouders R, Aktuna S, Thongjuea S, Aghajani-refah A, Pourfarzad F, van Ijcken W, et al. HBS1L-MYB intergenic variants modulate fetal hemoglobin via long-range MYB enhancers. *J Clin Invest*. 2014; 124:1699–1710. [PubMed: 24614105]
62. Thein SL, 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. *P Natl Acad Sci USA*. 2007; 104:11346–11351.
63. Roy P, Bhattacharya G, Mandal A, Dasgupta UB, 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*. 2012; 36:592–599. [PubMed: 23094636]
64. 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*. 2015; 136:1619–1626. [PubMed: 26220525]

65. 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*. 2015; 29:798–806. [PubMed: 25231743]
66. Srivastava SK, Bhardwaj A, Arora S, Singh S, Azim S, Tyagi N, et al. MYB is a novel regulator of pancreatic tumour growth and metastasis. *Brit J Cancer*. 2015; 113:1694–1703. [PubMed: 26657649]
67. Sripichai O, Whitacre J, Munkongdee T, Kumkhaek C, Makarasara W, Winichagoon P, et al. Genetic analysis of candidate modifier polymorphisms in Hb E-beta 0-thalassemia patients. *Ann N Y Acad Sci*. 2005; 1054:433–438. [PubMed: 16339693]
68. Danjou F, Francavilla M, Anni F, Satta S, Demartis FR, Perseu L, et al. A genetic score for the prediction of beta-thalassemia severity. *Haematologica*. 2015; 100:452–457. [PubMed: 25480500]
69. Pandit RA, Svasti S, Sripichai O, Munkongdee T, Triwitayakorn 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*. 2008; 88:357–361. [PubMed: 18839276]
70. Farrell JJ, Sherva RM, Chen ZY, Luo HY, Chu BF, Ha SY, et al. A 3-bp deletion in the HBS1L-MYB intergenic region on chromosome 6q23 is associated with HbF expression. *Blood*. 2011; 117:4935–4945. [PubMed: 21385855]
71. Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, et al. Discovery and refinement of loci associated with lipid levels. *Nat Genet*. 2013; 45:1274–1283. [PubMed: 24097068]
72. Wright HL, Moots RJ, Bucknall RC, Edwards SW. Neutrophil function in inflammation and inflammatory diseases. *Rheumatology (Oxford)*. 2010; 49:1618–1631.
73. Pillay J, den Braber I, Vrisekoop N, Kwast LM, de Boer RJ, Borghans JA, et al. In vivo labeling with 2H2O reveals a human neutrophil lifespan of 5.4 days. *Blood*. 2010; 116:625–627. [PubMed: 20410504]
74. Tak T, Tesselaar K, Pillay J, Borghans JA, Koenderman L. What's your age again? Determination of human neutrophil half-lives revisited. *J Leukoc Biol*. 2013; 94:595–601. [PubMed: 23625199]
75. Dorababu P, Nagesh N, Linga VG, Gundeti S, Kutala VK, Reddanna P, et al. Epistatic interactions between thiopurine methyltransferase (TPMT) and inosine triphosphate pyrophosphatase (ITPA) variations determine 6-mercaptopurine toxicity in Indian children with acute lymphoblastic leukemia. *Eur J Clin Pharmacol*. 2012; 68:379–387. [PubMed: 22009189]

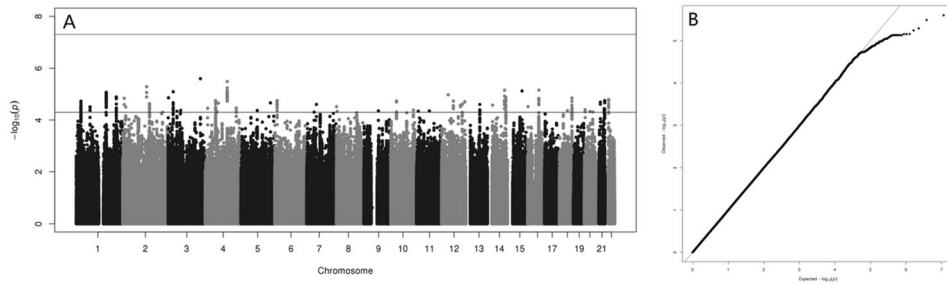


Figure 1.

A) Manhattan and B) QQ plot for the neutrophil-lymphocyte ratio (NLR) GWAS results with SNPs having a MAF > 0.01.

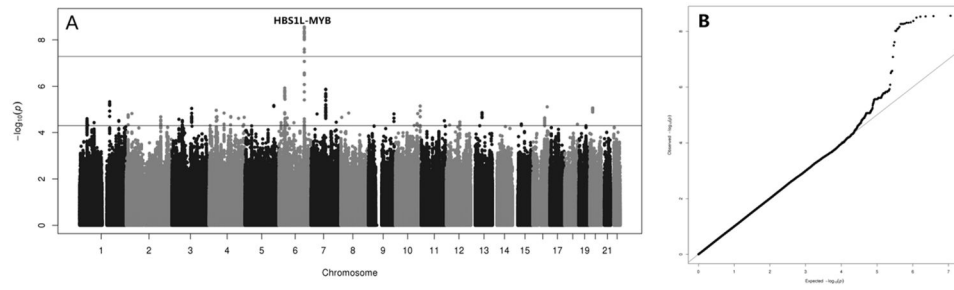


Figure 2.

A) Manhattan and B) QQ plot for the platelet-lymphocyte ratio (PLR) GWAS results with SNPs having a MAF > 0.01.

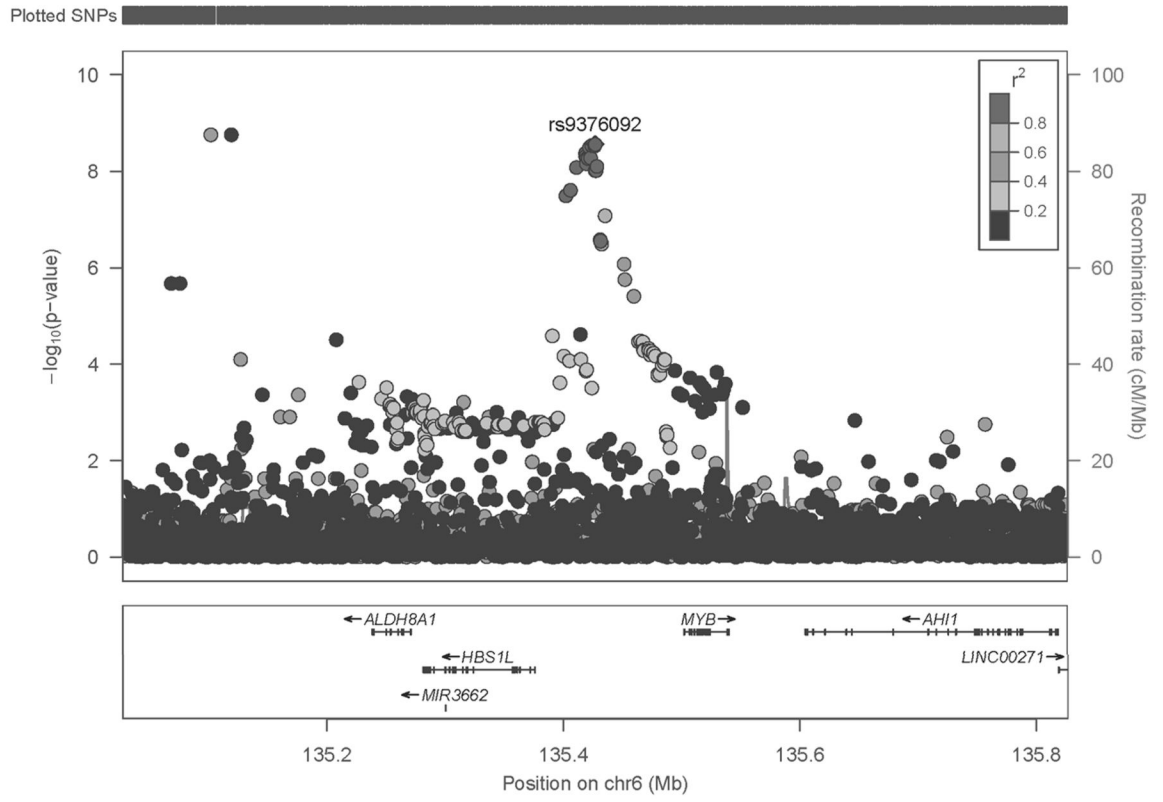


Figure 3.
Regional plot for the rs9376092 association with PLR level.

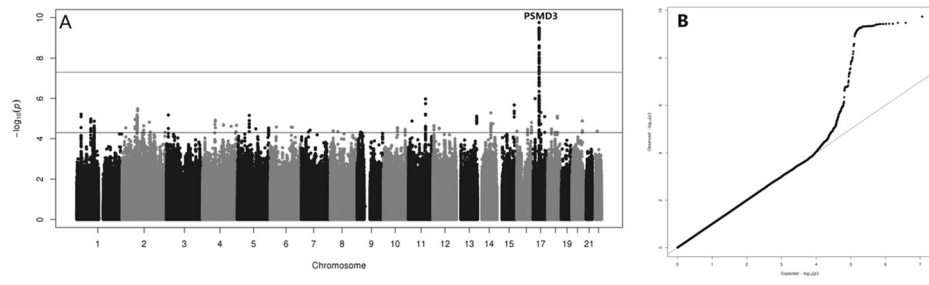


Figure 4.

A) Manhattan and B) QQ plot for the neutrophil count GWAS results with SNPs having a $MAF > 0.01$ ($\lambda = 1.011742$).

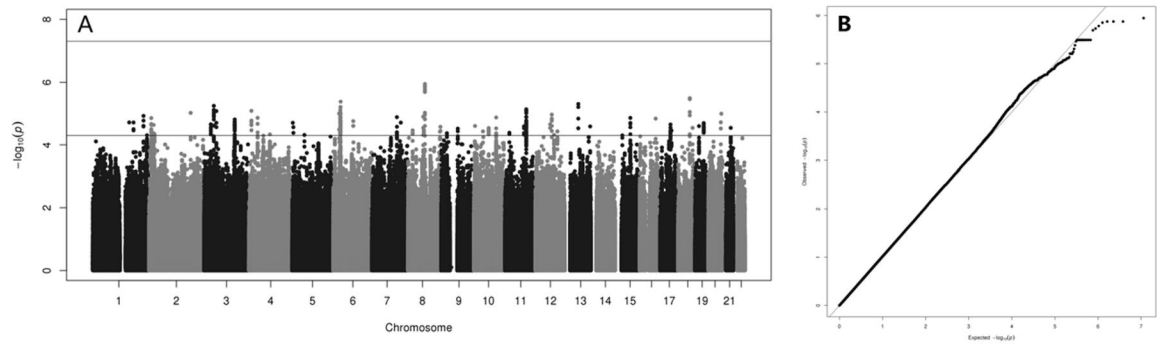


Figure 5.

A) Manhattan and B) QQ plot for the lymphocyte count GWAS results with SNPs having a $MAF > 0.01$ ($\lambda=1.022341$).

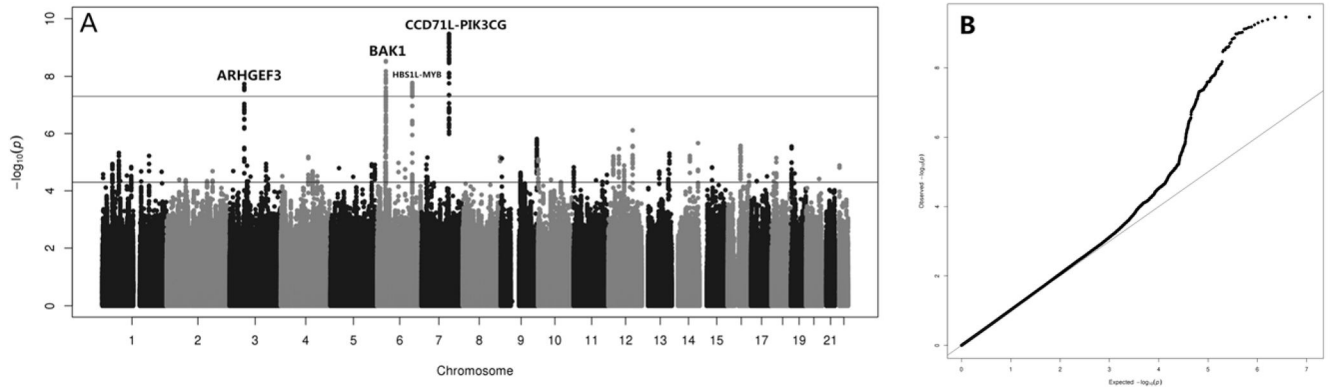


Figure 6.
A) Manhattan and B) QQ plot for the platelet count GWAS results with SNPs having a $MAF > 0.01$ ($\lambda = 1.018586$).

Table 1

Summary statistics of neutrophil-lymphocyte ratio (NLR) and platelet-lymphocyte ratio (PLR), the constituent blood cell count phenotypes and age in males and females.

	mean (SD) males (N=2250)	mean (SD) females (N=3651)	mean (SD) all (N=5091)
NLR	1.662 (.653)	1.615 (.690)	1.633 (.676)
PLR	116.354 (39.457)	124.770 (42.626)	121.561 (41.643)
Neutrophil	3.404 (1.146)	3.429 (1.256)	3.419 (1.215)
Lymphocyte	2.170 (.622)	2.264 (.699)	2.228 (.673)
Platelet	235.464 (52.147)	262.006 (60.187)	251.886 (58.684)
Age	43.60 (15.88)	42.26 (14.27)	42.77 (14.91)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

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 R ²	MAF	B	SE	P (PLR)	P(platelet count)	P(lymphocyte count)
rs9376092	6	135427144	no		0.270	5.483	0.922	2.75E-09	4.05E-08	0.008
rs4895440	6	135426558	yes	0.998	0.271	5.478	0.922	2.83E-09	4.45E-08	0.007
rs35959442	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
rs11759553	6	135422296	yes	0.984	0.273	5.446	0.919	3.19E-09	2.64E-08	0.010
rs7776054	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	5.42E-09	2.64E-08	0.014
rs9376091	6	135419636	yes	0.996	0.258	5.449	0.934	5.42E-09	2.52E-08	0.014
rs34164109	6	135421176	yes	0.994	0.258	5.449	0.934	5.42E-09	2.52E-08	0.014
rs9402685	9	135419688	yes	0.985	0.258	5.407	0.933	6.93E-09	3.30E-08	0.015
rs7758845	9	135428537	yes	0.985	0.263	5.362	0.929	7.86E-09	5.06E-08	0.015
rs9376090	9	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

Table 3

The top SNP rs9376092 GWAS results in the NTR data and TwinsUK data.

Dataset	Alleles(A1/A2)	Frequency A1	Beta	se	P
NTR (N=5091)	PLR	0.72	5.484	0.923	2.75E-9
	Platelet		6.984	1.273	4.05E-8
	Lymphocytes		-0.039	0.0148	0.008
TwinsUK (N=2538)	PLR	0.73	4.766	1.642	0.004
	Platelets		6.053	1.942	0.002
	Lymphocytes		0.014	0.020	0.49

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	SE	P
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	ARHGEF3	yes	0.984	0.387	-6.457	1.147	1.84E-08
Platelet	rs9376090	6	135411228	HBS1L-MYB	no		0.254	7.296	1.293	1.69E-08

Table 5

LD regression results for NLR, PLR and the blood cell counts.

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

Mean of χ^2 : the mean of χ^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. λ GC: the lambda genomic control inflation factor. H^2 (se): the estimated SNP heritability based on LD score regression with (standard error). Intercept (se) the regression intercept with (standard error), where the value minus one, indicates the contribution of confounding in the studied sample.

Table 6

Phenotypic and genetic correlations for NLR and PLR levels and blood cell counts.

Phenotype	Phenotypic correlation	Genetic correlation	NLR	PLR	NC	LC	PC
NLR			0	0.223	-0.153	0	0
PLR	0.491**		0.491**	0	-0.486	0.457*	0.457*
NC	0.651**	-0.073**	0.651**	-0.073**	0	-0.062	-0.062
LC	-0.481**	-0.678**	-0.481**	-0.678**	0.271**	0.644	0.644
PC	0.018	0.478**	0.018	0.478**	0.209**	0.218**	0.218**

In this table the phenotypic correlations are given below the diagonal. The genetic correlations are given above the diagonal. NC = neutrophil count, LC = lymphocyte count, PC = platelet count, NLR = Neutrophil to lymphocyte ratio, PLR = Platelet to lymphocyte ratio. P-values indicating the significance of the correlations are given with * p < 0.05, ** p < 0.001.