GWAS and PheWAS of Red Blood Cell Components in a Northern Nevadan Cohort

Karen A. Schlauch[1], Robert W. Read[1], Gai Elhanan[1], William J Metcalf[1], Anthony D. Slonim[2], Ramsey Aweti[3], Robert Borkowski[3], and Joseph J. Grzymski[1,2]

1. Applied Innovation Center, Renown Institute for Health Innovation, Desert Research Institute, Reno, NV, USA

2. Renown Health, Reno, NV, USA

3. 23andMe, Inc. Mountain View, CA, USA

Correspondence: Joseph J. Grzymski

***E-mail: joeg@dri.edu (JG)**

Abstract

In this study, we perform a full genome-wide association study (GWAS) to identify statistically significantly associated single nucleotide polymorphisms (SNPs) with three red blood cell (RBC) components and follow it with two independent PheWASs to examine associations between phenotypic data (case-control status of diagnoses or disease), significant SNPs, and RBC component levels. We first identified associations between the three RBC components: mean platelet volume (MPV), mean corpuscular volume (MCV), and platelet counts (PC), and the genotypes of approximately 500,000 SNPs on the Illumina Infimum® DNA Human OmniExpress-24 BeadChip using a single cohort of 4,700 Northern Nevadans. Twenty-one SNPs in five major genomic regions were found to be statistically significantly associated with MPV, two regions with MCV, and one region with PC, with $p<5x10^{-8}$. Twenty-nine SNPs and nine chromosomal regions were identified in 30 previous GWASs, with effect sizes of similar magnitude and direction as found in our cohort. The two strongest associations were SNP rs1354034 with MPV ($p=2.4x10^{-13}$) and rs855791 with MCV ($p=5.2x10^{-12}$).We then examined possible associations between these significant SNPs and incidence of 1,488 phenotype groups mapped from International Classification of Disease version 9 and 10 (ICD9 and ICD10) codes collected in the extensive electronic health record (EHR) database associated with Healthy Nevada Project consented participants. Further leveraging data collected in the EHR, we performed an additional PheWAS to identify associations between continuous red blood cell (RBC) component measures and incidence of specific diagnoses. The first PheWAS illuminated whether SNPs associated with RBC components in our cohort were linked with other hematologic phenotypic diagnoses or diagnoses of other nature. Although no SNPs from our GWAS were identified as strongly associated to other phenotypic components, a number of associations were identified with *p*-values ranging between $1x10^{-3}$ and $1x10^{-4}$ with traits such as respiratory failure, sleep disorders, hypoglycemia, hyperglyceridemia, GERD and IBS. The second PheWAS examined possible phenotypic predictors of

abnormal RBC component measures: a number of hematologic phenotypes such as thrombocytopenia, anemias, hemoglobinopathies and pancytopenia were found to be strongly associated to RBC component measures; additional phenotypes such as (morbid) obesity, malaise and fatigue, alcoholism, and cirrhosis were also identified to be possible predictors of RBC component measures.

Author Summary

The combination of electronic health records and genomic data have the capability to revolutionize personalized medicine. Each separately contains invaluable data; however, combined, the two are able to identify new discoveries that may have long-term health benefits. The Healthy Nevada Project is a non-profit initiative between Renown Medical Center and the Desert Research Institute in Reno, NV. The project has so far collected a cohort of 6,500 Northern Nevadans, with extensive medical electronic health records in the Renown Health database. Combining the genotypes of these participants with the clinical data, this study's aim is to find associations between genotypes (genes) and phenotypes (diagnoses and lab records). Here, we identify and examine clinical associations with red blood cell components such as platelet counts and mean platelet volume. These are components that have clinical relevance for several diseases, such as anemia, atherothrombosis and cancer. Our results from genome wide association studies mirror previous studies, and identify new associations. The extensive electronic health records enabled us to perform phenome wide associations to discover strong associations with hematologic components, as well as other important traits and diagnoses.

Introduction

The complete blood count (CBC) is a widely used medical diagnostic test that is a compilation of the number, size, and composition of various components of the hematopoietic system. Abnormal CBC measures may indicate illness or disease. Mean corpuscular volume (MCV), platelet count (PC), and

mean platelet volume (MPV) are specific CBC characteristics (hereby called RBC components), and linked to complex disorders such as anemia, alpha thalassemia and cardiovascular disease [1-5]. Platelets are involved in vascular integrity, wound healing, immune and inflammatory responses, and tumor metastasis; the role of platelets is also paramount in hemostasis and in the pathophysiology of atherothrombosis and cancer [6-12]. Additionally, abnormally high mean platelet volumes (MPV) are considered a predictor of post event outcome in coronary disease and myocardial infarction [13]. Furthermore, studies have shown that individuals living in higher altitudes have noted differences in red blood cell components than at sea level. At approximately 4,400 feet above sea level, Northern Nevada, where this study is conducted, is considered a high desert in the Sierra Nevada foothills. Alper showed that mean platelet volume (MPV) is 7.5% higher at altitudes greater than 4,000 feet than at sea level [14]. Similarly, Hudson showed a notable and statistically significant positive correlation with platelet counts (PC) and altitude [15], while mean corpuscular volume (MCV) was recorded as lower at higher altitudes than at sea level [16]. As RBCs help transport oxygen throughout the entire body, the identification of RBC-related genotypic mutations, especially in an RBC high-turnover environment is valuable. Lastly, the identification of genomic regions with roles in megakaryopoiesis and platelet formation, as well as neoplastic conditions like polycythemia vera and essential thrombocytosis (ET) [17,18], may help identify those that have a higher risk of certain complex RBC diseases.

Given the importance of these three RBC components, we conducted a study to identify both genetic and phenotypic associations with all three characteristics via GWASs and PheWASs. Our study begins with the Healthy Nevada Project, a single cohort formed in 2016 to investigate factors that may contribute to health outcomes in Northern Nevada. Its first phase provided 10,000 individuals in Northern Nevada with genotyping on the 23andMe 2016 Illumina Human OmniExpress-24 BeadChip platform at no cost. Renown Hospital is the largest hospital in the area, and 75% of these 10,000 individuals are cross-referenced in its extensive EHR database

As noted above, previous GWASs have identified significant genetic links with all three RBC components we examine in this study, MPV, MCV and PC [13,17-45]. Lin et al. 2007 identified a strong genetic link with MCV in region 11p15 using the Framingham cohort [19]; Kullo et al. 2010 leveraged EHR data from the Mayo Clinic to detect four genes strongly associated with at least one of the three RBC components [27]. Similarly, a number of regions were linked with PC in an African American cohort [35] and MPV [35]; Shameer detected five regions associated with PC and eight with MPV [18].

Our study first performed a genome-wide association study (GWAS) of 4,700 genotyped Northern Nevadans who have at least one recorded value for one of the three RBC components MPV, MCV and PC to examine the genetic component of these components. We found 38 SNPs to be statistically significantly associated ($p<5x10^{-8}$) to one of the three RBC components. Many of these associations were previously reported, yet our study did identify nine novel SNPs in six different regions. While there were few new associations discovered in our cohort, we identified several SNPs that fall within genes influencing megakaryocytes maturation, platelet volume, platelet signaling and diseases such as anemia. Further, with extensive linked electronic medical record (EMR) data, we had the ability to perform a PheWAS of 1,488 standard lab results (phenotypes) against each SNP found to be associated to RBC components in the Northern Nevadan cohort to examine pleiotropy. Additionally, we then examined the RBC components phenotypically, using linked electronic medical record (EMR) data to determine the relationship between measures of each component and a variety of clinical conditions recorded in patients. Many relevant and strongly statistically significant associations were identified, especially with hematologic components; other traits not currently shown to be linked to RBC components, such as obesity, alcoholism and cirrhosis, were also detected.

Results

Characteristics of cohort

We examined 4,700 genotyped individuals with at least one recorded RBC measure; 4,590 individuals in the cohort had measures for all three components. Table 1 describes the cohort with respect to gender, age, ethnic origin, and standardized value of each RBC component. Note that all values for each component were standardized to the most current lab test administered for that component via linear transformation. The normal (healthy) reference values to which all individual records were standardized are also presented in Table 1. The mean standardized RBC component values for each individual are available in Supplementary Table S1.

Table 1. Cohort Characteristics

Age (years)	47.24 ± 15.82	
Make (%)	1328 (28.24)	
African American (%)	53 (1.12)	
Asian (%)	100 (2.12)	
Caucasian (%)	4,202 (89.35)	
Latino (%)	138 (2.93)	
Native American (%)	30 (0.64)	
Pacific Islander (%)	11 (0.23)	
Unknown (%)	168 (3.6)	
	Standardized Component levels	**Normal Reference Ranges**
MPV (fL)	10.58 ± 0.98	[9, 12.9] (fL)
MCV (fL)	91.53 ± 4.47	[81.4, 97.8] (fL)
PC (K/uL)	251.57 ± 62.23	[164, 446] (K/uL)

Table of cohort characteristics. Continuous variables are presented as mean ± SD; categorical variables are presented as counts and percentages. All values were standardized to the reference ranges of the most recent administered laboratory test.

GWAS of RBC components

Using the average measures of each individual's MPV, PC and MCV lab records, a standard GWAS under the additive model with adjustments for gender, age and the first four principal components was

performed using *PLINK* 1.9. Genomic inflation coefficients (lambda) were computed for each cohort: 1.031 for MPV, 1.027 for PC, and 1.045 for MCV.

Any SNP with an association *p*-value of $p<5x10^{-8}$ was considered a statistically significant association, following current standards [28,32,46,47]. The percentage of phenotypic variance attributed to genetic variation was computed with a combination of *PLINK* and GCTA [48]: genetic variance was 35.3% for MCV; 32.2% for MPV; 20.7% for PC. The three individual GWAS studies identified a total of 38 SNPs that associated with a RBC component with statistical significance. Manhattan plots of the three GWAS results are presented in Supplemental Figure S1(A-C). As an example for the reader, we include in the manuscript (Fig. 1), a Manhattan plot for MCV.

MPV

A GWAS was performed on a cohort of 4,591 genotyped participants with MPV laboratory measures. We identified 21 SNPs across five different chromosomal regions that reached genome-wide significance ($p < 5x10^{-8}$; Table 2). Of these, 13 demonstrated previous associations in at least one other study, with six associated with RBC components (Supplementary Table S2) [13,17,18,25,28,30,33,35,49-58]. All five significant chromosomal regions were previously associated with MPV[17,18]. The fifth region 18q22.2, contains three SNPs associated in our cohort with average *p*-value $p=3.86x10^{-9}$, however none of the individual SNPs have been previously associated with MPV. Results are presented in Table 2.

Table 2. Statistically Significant GWAS SNPs

rsID	Chrom	Cyto Region	Associated Gene	Minor Allele	MAF	β	(SE)	GWAS *p*-value	Mutation Classification	RBC
rs10274553	chr3	p14.3	ARHGEF3	C	49.74	-0.1198	0.020	$3.82x10^{-9}$	intron	MPV
rs10509186	chr3	p14.3	ARHGEF3	T	45.55	-0.1187	0.021	$7.75x10^{-9}$	intron	MPV
rs10822186	chr7	q22.3	NA	G	49.22	-0.1107	0.020	$4.38x10^{-8}$	unknown	MPV
rs11130549	chr7	q22.3	NA	C	34.11	-0.1238	0.022	$9.94x10^{-9}$	unknown	MPV
rs12355784	chr7	q22.3	NA	A	45.46	-0.1184	0.021	$9.32x10^{-9}$	unknown	MPV

rs1354034	chr7	q22.3	NA	T	41.14	0.1546	0.021	$2.39x10^{-13}$	unknown	MPV
rs1788103	chr7	q22.3	NA	G	48.18	-0.1261	0.020	$5.15x10^{-10}$	unknown	MPV
rs1790588	chr7	q22.3	NA	C	48.01	-0.1273	0.020	$3.31x10^{-10}$	unknown	MPV
rs1790974	chr10	q21.3	JMJD1C	T	43.82	-0.1128	0.020	$3.32x10^{-8}$	intron	MPV
rs1935	chr10	q21.3	JMJD1C	G	45.58	-0.1135	0.021	$3.57x10^{-8}$	intron	MPV
rs201979226	chr10	q21.3	JMJD1C	C	48.78	0.1183	0.020	$5.89x10^{-9}$	intron, near-gene-5	MPV
rs342240	chr10	q21.3	JMJD1C	A	41.36	0.129	0.021	$3.49x10^{-10}$	intron, untranslated-3	MPV
rs342275	chr10	q21.3	JMJD1C	T	41.9	0.1292	0.020	$2.96x10^{-10}$	intron	MPV
rs342293	chr10	q21.3	JMJD1C	G	44.31	0.1325	0.020	$6.61x10^{-11}$	missense	MPV
rs342296	chr10	q21.3	REEP3	A	44.03	0.131	0.020	$1.04x10^{-10}$	intron	MPV
rs34818942	chr12	q24.31	WDR66	T	7.29	0.254	0.039	$7.77x10^{-11}$	intron	MPV
rs386614085	chr12	q24.31	RHOF	G	45.45	-0.1172	0.021	$1.21x10^{-8}$	intron	MPV
rs4379723	chr18	q22.2	CD226	C	45.45	-0.1172	0.021	$1.29x10^{-8}$	missense	MPV
rs763361	chr18	q22.2	CD226	T	47.24	-0.1273	0.020	$3.26x10^{-10}$	intron	MPV
rs7910927	chr18	q22.2	CD226	G	45.52	-0.1146	0.021	$2.68x10^{-8}$	intron	MPV
rs7961894	chr18	q22.2	DOK6	T	10.24	0.2221	0.033	$2.68x10^{-11}$	untranslated-3	MPV
rs218237	chr4	q12	NA	T	15.19	0.740	0.126	$5.07x10^{-9}$	unknown	MCV
rs9402686	chr6	q23.3	NA	A	24.89	0.647	0.104	$4.60x10x^{-10}$	unknown	MCV
rs7776054	chr6	q23.3	NA	G	24.28	0.645	0.104	$6.65x10^{-10}$	unknown	MCV
rs9399137	chr6	q23.3	NA	C	23.83	0.648	0.105	$7.82x10^{-10}$	unknown	MCV
rs7775698	chr6	q23.3	NA	T	24.23	0.642	0.104	$8.24x10^{-10}$	unknown	MCV
rs4895441	chr6	q23.3	NA	G	25.08	0.628	0.103	$1.27x10^{-9}$	unknown	MCV
rs111194878	chr6	q23.3	NA	A	25.41	0.622	0.103	$1.47x10^{-9}$	unknown	MCV
rs9373124	chr6	q23.3	NA	C	26.16	0.599	0.102	$5.14x10^{-9}$	unknown	MCV
rs855791	chr22	q12.3	TMPRSS6	A	44.2	-0.621	0.090	$5.23x10^{-12}$	missense	MCV
rs4820268	chr22	q12.3	TMPRSS6	G	46.56	-0.604	0.090	$2.65x10^{-11}$	coding-synon	MCV
rs5756504	chr22	q12.3	TMPRSS6	T	36.93	0.567	0.092	$7.77x10^{-10}$	intron	MCV
rs130624	chr22	q12.3	NA	G	42.77	0.549	0.090	$1.13x10^{-9}$	unknown	MCV
rs5756506	chr22	q12.3	TMPRSS6	C	36.92	0.563	0.092	$1.15x10^{-9}$	intron	MCV
rs386563505	chr22	q12.3	NA	A	40.75	0.525	0.091	$7.12x10^{-9}$	unknown	MCV
rs385893	chr9	p24.1	NA	T	49	-7.744	1.258	$8.04x10^{-10}$	unknown	PC
rs10974808	chr9	p24.1	RCL1	G	11.48	11.490	1.943	$3.53x10^{-9}$	intron	PC
rs423955	chr9	p24.1	NA	C	34.17	-7.387	1.325	$2.64x10^{-8}$	near-gene-5	PC

This table lists the statistically significant SNPs associated in our cohort with MPV, MCV, and PC. Effect sizes and their standard deviations are presented in fL per each copy of the minor allele. Raw *p*-values generated by the GWAS are presented.

MCV

A GWAS was performed on a cohort of 4,699 genotyped participants with MCV laboratory measures. There were 14 SNPS that were significantly associated with MCV (Table 2). These SNPs lie in three chromosomal regions: predominantly in 6q23.3 and 22q12.3. These two regions have detailed annotation and were linked previously with MCV (Supplementary Table S2) [20,27,32]. All but four of the SNPs are in non-coding regions. These four SNPs lie in *TMPRSS6*. The gene *TMPRSS6* codes for the protein matriptase-2, which is part of a signaling pathway that regulates blood iron levels [31]. The two SNPS rs855791 and rs4820268 showed the strongest association with MCV ($p<1x10^{-11}$). These two SNPS also lie in *TMPRSS6* and cause a missense and synonymous mutation, respectively. Results are presented in Table 2.

PC

A GWAS was performed on a cohort of 4,700 genotyped participants with PC laboratory measures. Three SNPs were identified with statistically significant ($p<5x10^{-8}$) links to PC values in our cohort, two of which were previously identified in other studies (Supplementary Table S2) [17,25,26,34]. The SNP rs10974808 is in the same cytoband region (9p24.1) as the others but has not been linked to PC. The three SNPs have different effects on PC: rs385893 and rs423955 have negative effect size (β=-7.744 and -7.387, respectively), while rs10974808 has a positive effect (β=11.490). The minor allele frequency of rs10974808 is much rarer (MAF=11.48%) compared to 49% for rs385893 and 31.17% for rs423955. Results are presented in Table 2.

Comparison to other GWAS studies

The Northern Nevada cohort had mean standardized MPV values of 10.58 ± 0.98 fL, comparable to levels reported in the Health ABC cohort described in Qayyum (10.9 ± 1.6 fL), and two European cohorts investigated in Geiger (10.53 ± 1.08, 10.83 ± 0.87) [28,35]. The Nevadan cohort had MCV values of 91.53 ± 4.5 fL, also comparable to those described in Kullo (90.5 ± 4.2 fL) and Ding in the

Mayo and Johns Hopkins Group Health Cooperative cohorts (90.53 ± 4.17 and 91.56 ± 4.49, respectively), as well as several European cohorts in Geiger (e.g., 91.5 ± 4.2, 91.4 ± 4.41, 91.1 ± 4.44, 92.0 ± 4.3) [27,28,32]. Mean standardized PC values in the Nevadan cohort (251 ± 62.23 K/uL) were very similar to many of the cohorts examined in Geiger (e.g., 258.6 ± 63.1, 252 ±71.7, 250.9 ±64.8, 247 ± 64.7) [28].

Our three GWAS results were in close correlation with many of the other studies. For example, the locus rs7961894 in the WDR66 gene on q24.31 was found associated to MPV in our cohort and in Meisinger as a top hit [24]. Effect sizes in Meisinger were larger than ours (1.03 vs. 0.22), but the number of minor alleles predicted an increase in MPV for both studies. Another SNP, rs342240, was one of our cohort's top associations with MPV, and was also identified by Shameer and Soranzo as significant links to MPV [17,18]. Similarly, locus rs385893 was identified as a possible predictor of PC by Soranzo and our cohort, with very similar notably large negative effect sizes (-6.24 and -7.74, respectively). Kullo also found SNP rs7775698 to be significantly associated to MCV, with similar positive effect sizes as our study (0.92 vs 0.56) [27]. Soranzo et al. identified rs9402686 as a top link with MCV, and again, effect sizes were similar to ours (0.82 vs 0.65) [17].

ANOVA

The mean component values across genotypes presented in Supplementary Table S2 correlate with negative and positive effect sizes: SNPs showing a negative effect size have a decrease in component values across the genotypes from left to right (homozygous in major allele, heterozygous, homozygous in minor allele). All ANOVA *p*-values of the significant SNPs identified in this study are significant, even after a simple Bonferonni correction (.05/38=0.001). A box and whisker figure of ANOVA results for the top hit SNP rs7961894 are shown in Supplementary Fig S2.

PheWAS of RBC components

The first PheWAS examined possible associations between significant SNPs identified in each RBC trait GWAS and 1,488 phenotypic groups. At significance levels $1x10^{-4}<p<1x10^{-3}$, putative associations of MCV-specific SNPs included respiratory failure; those with PC included GERD and other diseases of the esophagus. Our study also showed links with MPV-associated SNPs and skin cancer, hypoglycemia, hyperglyceridemia, IBS, among others. These associations are outlined in Supplementary Figs S3(A-C).

The second PheWAS investigated whether the 1,488 phenotype groups were associated with the levels of each RBC component; more specifically, the analysis identified whether the number of cases in a phenotype group was a predictor of the level of the component. For example, the PheWAS examining associations of MPV levels presented significant links with thrombocytopenia and purpura ($p<1x10^{-8}$). Interestingly, Vitamin D deficiency was also shown to be a predictor of MPV levels, although at a lower significance level ($p<1x10^{-6}$). Incidence of malaise and fatigue was also found to be a potential predictor of MPV in our cohort.

Associations with MCV included hemoglobinopathies and hemolytic anemias ($p<1x10^{-35}$), as well as iron deficient anemias ($p<1x10^{-20}$). Again, association with (morbid) obesity was evident ($p<1x10^{-20}$). Alcoholism and related liver diseases were associated with MCV at a significance level of $p<1x10^{-8}$; abnormal glucose and diabetes were also linked to MCV at $p<1x10^{-5}$. We identified a strong association in our cohort between platelet counts and thrombocytopenia and purpura ($p<1x10^{-30}$). Associations with other hematologic phenotypes such as various anemias and pancytopenia also reached significance ($p<1x10^{-8}$). Additionally, (morbid) obesity and cirrhosis were statistically significantly associated with PC with $p<1x10^{-8}$ significance level. These three PheWAS results are

231 shown in Supplementary Fig S4(A-C). As an example for the reader, we include the PheWAS results
232 for MCV in Fig 2.
233

234 # Discussion

235 In this study, we first performed three independent GWASs of 4,700 Healthy Nevada Project
236 participants with 500,000 genotypes against the RBC components: platelet count, mean platelet volume
237 and mean corpuscular volume. We followed these with two independent PheWASs for each component
238 to identify additional phenotypic associations with each blood component-significant SNP, and
239 phenotypic associations with measures of each blood component.
240

241 Our genome-wide association analysis identified ten different chromosomal cytoband regions
242 associated with at least one RBC component. Nine of those regions were previously associated to RBC
243 components in other studies; the region 22q13.33 represents a novel region in our study
244 [17,18,20,25,27,28,30,32,49,59,60]. Nine genes lie in the cytoband regions: their functions are outlined
245 in Table 3.

246 **Table 3. Table Presenting Gene Functions**

Gene	Gene Description	Region	RBC	Function	Reference
ARHGEF3	Rho Guanine Nucleotide Exchange Factor 3	p14.3	MPV	Increases activity of Rho GTPases by catalyzing the release of bound GDP; may have a role in megakaryocytes maturation	[61]
JMJD1C	Histone Demethylase	q21.3	MPV	Possible hormone-dependent transcriptional activation	[17]
REEP3	Receptor Accessory Protein 3	q21.3	MPV	Membrane protein	[18]
WDR66	WD Repeat Domain 66	q24.31	MPV	May create and alter platelet volumes	[24]
RHOF	Ras Homolog Family Member F	q24.31	MPV	May regulate platelet filopodia formation	[62]
CD226	Cluster of Differentiation 226	q22.22	MPV	Catalyzes binding of activated platelets to endothelial cells; may have a role in in platelet signal transduction	[63]
DOK6	Docking Protein 6	q22.2	MPV	Protein scaffolding	[64]

TMPRSS6	Transmembrane Protease, Serine 6	q12.13	MCV	Acts by cleaving hemojuvelin	[27,31,38]
RCL1	RNA Terminal Phosphate Cyclase Like 1	p24.1	PC	rRNA processing	[28]

This table presents functions of genes associated to all SNPs found significantly associated to one or more RBC components in the GWASs.

Our GWAS results were very similar to previous MPV GWAS associations. The most significant genetic association with MPV (rs1354034, $p = 2.39\text{x}10^{-13}$) is found in an intronic region within *ARHGEF3* on chromosome 3p14.3. The gene *ARHGEF3* codes for a Rho guanine nucleotide exchange factor 3 protein and was associated to MPV in previous studies [13,17,18,28,33,61], further demonstrating that our study was able to replicate associations with RBC components in prior single-cohort studies. The mechanism by which rs1354034 affects MPV values is still ambiguous. As it lies in a DNase I hypersensitive region within open chromatin, it could directly affect *ARHGEF3* expression in human megakaryocytes maturation [61]. Our second most significant association (rs7961894, $p= 2.68\text{x}10^{-11}$) was also previously linked with MPV [13,18,24,28]. This SNP lies in intron 3 of *WDR66* on chromosome 12q24.31. Expression levels of *WDR66* have been directly tied to MPV, possibly indicating that *WDR66* is involved in the establishment of platelet volumes. SNP rs7961894 is not directly correlated with *WDR66* expression levels, implying an indirect role possibly through other regulatory mechanisms [24].

We also identified several SNPs on chromosome 10q21.3 to be associated with MPV in our cohort that were linked to sex hormone levels in previous studies [53]. This may imply a possible relationship between sex hormone levels and MPV. These SNPs almost exclusively lie in *JMJD1C*, a gene that encodes as a probable histone demethylase, and may have a function in hormone-dependent transcriptional activation [17]. This could indicate that the transcription of certain hematopoietic target genes may be enhanced or repressed when specific sex hormones are present; however, the exact targets and mechanisms have yet to be studied and clinical evidence for such association is scant.

Further, the chromosomal region 18q22.2 was shown to be associated with MPV [13], although the significant SNPs in this region have not been linked to MPV in previous studies. Three out of the four SNPs in this region are intronic to *CD226*, while one is in an untranslated region of *DOK6*. *CD226* codes for a protein, which mediates the binding of activated platelets to endothelial cells and may participate in platelet signal transduction [63]. Soranzo et al. also identified this gene as having a possible role in megakaryocyte (MK) development, thus these SNPs in *CD226* may influence platelet development [17]. *DOK6* encodes a docking protein, necessary for protein scaffolding, but to our knowledge has no known relation to platelet function; therefore, the functional relevance of a SNP in this gene is ambiguous. The mechanism by which these SNPS within 18q22.2 affect *CD226*, *DOK6* or MPV is also currently unknown.

The majority of SNPs associated with MCV and PC are in non-coding regions, and most were previously associated with these components in previous studies [17,27,32,45]. Our two strongest associations with MCV (rs855791, $p = 5.23\text{x}10^{-12}$) and (rs4820268, $p = 2.65\text{x}10^{-11}$) are in the gene *TMPRSS6* and could cause an altered or loss of function for the matriptase-2 protein. Altered function of the protein will likely influence iron status within the body, demonstrating why these SNPS are highly associated with anemia caused by iron deficiency [31,38]. PC was associated with only a single gene in our GWAS. This gene, *RCL1*, which encodes an RNA terminal phosphate cyclase-like 1 protein, was previously associated with PC [28]. The SNP associated to PC in this gene (rs10974808, $p= 3.53\text{x}10^{-09}$) in our cohort has not been linked to PC by other studies to the best of our knowledge. Our strongest association (rs385893, $p = 8.04\times10^{-10}$) was previously found to affect *JAK2*, a gene 400 kb downstream of the locus and a key regulator of megakaryocyte maturation, illustrating that these SNPs may influence changes over large genetic regions [17]. This also highlights the difficulty

295 determining the exact mechanisms by which these SNPS alter components, such as RBC, given their
296 large theoretical range of influence.
297
298 We present here two comprehensive PheWAS analyses of RBC components. The first examines
299 whether additional phenotypic associations exist between SNPs associated to an RBC component in
300 our cohort. The second groups extensive EHR phenotypic data from the Healthy Nevada Project
301 clinical database into 1,488 different phenotype groups and examines the association (predictive value)
302 between their incidence rate with continuous RBC component values. To the best of our knowledge,
303 this is the first PheWAS targeted at RBC components. Not surprisingly, many of our strongest
304 associations were with hematopoietic phenotypes, indicating that the incidence of having one (or more)
305 abnormal hematopoietic characteristics is a potential predictor of RBC component levels. Interestingly,
306 the incidence of having vitamin D deficiency may be linked to MPV levels and requires further study,
307 as incident solar radiation in the Northern Nevadan location of the study is high. Also of interest is that
308 MCV and PC levels could be associated to the occurrence of (morbid) obesity, alcoholism and cirrhosis
309 which are linked to poor vitamin D synthesis [65].
310
311 The identified associations between the RBD indices and hematopoietic findings and pathologies are
312 mostly expected due to their known physiologic association and reconfirm previously reported
313 findings. Iron deficiency anemia is often microcytic and characterized by reduced MCV [66]. Iron
314 deficiency also affects megakaryocytes and may induce changes in megakaryocyte differentiation as
315 well as increased platelet counts and volume [67]. As noted earlier, one of the strongest associations
316 reported here is in the vicinity of JAK2, a known regulator of megakaryocytes maturation [68].
317
318 While thrombocytopenias are clearly synonymous with reduced PC, associated platelet volume and
319 size changes can be used to differentiate between inherited macrothrombocytopenias and idiopathic

thrombocytopenic purpura (ITP) [69], thus establishing an association with MPV that may be positive or inverse. While this study demonstrated a strong negative association between PC and purpura, and a positive association with MPV, it is important to note that not all purpuras are necessarily caused by platelet deficiency. However, phenotypic groupings were not specific enough to identify associations with respect to specific etiologies (See Supplementary Table S3).

Vitamin D, independently, and in association with platelet activity and increased platelet indices, has been associated with cardiovascular disease [70]. The positive association between vitamin D deficiency and MPV levels is intriguing and follows other findings. Cumhur et al. [71] observed an inverse correlation between vitamin D levels and MPV and hypothesized that this may be due to increased release of proinflammatory cytokines present with vitamin D deficiency. Park et al. also reported an inverse association between PC and MPV and vitamin D levels in adults [72].

Platelet activation, as evidenced by platelet indices, is a recognized phenomenon in metabolic syndrome [73,74]. This study resulted in a positive association between PC and morbid obesity, and a negative association between MCV and obesity and morbid obesity. While previous evidence [75] does not necessarily support all-gender association between obesity and increased platelet counts, our finding may reflect an association between the central obesity of metabolic syndrome and the associated platelet activation of metabolic syndrome. However, the phenotype groups were not specific enough to allow for specific differentiation between obesity types (See Supplementary Table S3).

Thrombocytopenia is often observed in chronic liver disease and cirrhosis and platelet activation may play a role in liver regeneration [76,77]. Alcoholism is also associated with thrombocytopenia [78]. However, evidence of an association between liver disease or alcoholism and platelet activation indices is lacking. Moreover, evidence points to platelet function defects in chronic alcoholism [79]. Thus, the

345 negative effect of PC on cirrhosis and positive effect of MCV on cirrhosis, alcoholism, and alcohol-
346 related disorders found in this study is intriguing and merits further confirmation and research.
347

348 # Materials and Methods

349 *The Renown EHR Database*

350 The Renown Health EHR system was instated in 2007 on the EPIC system (EPIC System Corporation,
351 Verona, Wisconsin, USA), and currently contains lab results, diagnosis codes (ICD9 and ICD10) and
352 demographics of more than 1 million patients seen in the hospital system since 2005.
353

354 *Sample Collection*

355 Saliva as a source of DNA was collected from 10,000 adults in Northern Nevada as the first phase of
356 the Healthy Nevada Project to contribute to comprehensive population health studies in Nevada. The
357 personal genetics company 23andMe® was used to genotype these individuals. using the Orogene®
358 DX OGD-500.001 saliva kit [DNA Genotek, Ontario, Canada]. Genotypes are based on the Illumina
359 Human OmniExpress-24 BeadChip platform [San Diego, CA, USA] including approximately 570,000
360 SNPs.
361

362 *IRB and Ethics Statement*

363 The study was approved by our local Institutional Review Board (IRB, project 956068-12). Participants
364 in the Healthy Nevada Project consent to having genetic information associated with electronic health
365 information in a de-identified manner. Neither researchers nor participants have access to the complete
366 EHR data and cannot map participants to patient identifiers. These data are not incorporated into the
367 HER; rather, EHR and genetic data are linked in a separate environment via a unique identifier as
368 approved by the IRB.

Processing of EHR data

Most cohort participants had multiple RBC recordings across thirteen years; in these cases, the mean age of each participant across those records was computed and later used as a covariate for each component in GWAS and PheWAS analyses. Normalization of test values was necessary as lab tests were updated across the 13 years of data collection. Many of the participants had lab results (for the same RBC component) recorded across different tests with different healthy reference ranges. For example, the 4,700 participants had measurements for MCV with respect to one or more of ten different MCV lab tests and corresponding healthy reference ranges. Many participants had records across several of these ten different tests. Only those tests/reference ranges having records for more than one individual were used in analyses. To standardize the RBC values across different normal reference ranges, a simple linear transform was computed using each test's reference range and the most recent test's range. All component measures within each separate test were then transformed into ranges of the most recent via each range's specific linear transform. The most recent healthy normal reference range for each component is listed in Table 1. Distributions of raw and transformed laboratory test values can be found in Supplementary Fig S5(A-C).

Genotyping and Quality Control

Genotyping was performed by 23andMe using the Illumina Infimum® DNA Human OmniExpress-24 BeadChip V4. This genotyping platform (Illumina, San Diego, CA) consists of approximately 570,000 SNPs. DNA extraction and genotyping were performed on saliva samples by the National Genetics Institute (NG1), a CLIA licensed clinical laboratory and a subsidiary of the Laboratory Corporation of America.

Raw genotype data were processed through a standard quality control process [46,47,80-82]. SNPs with a minor allele frequency (MAF) less than 0.01 were removed. SNPS that were out of HWE (*p*-value < $1x10^{-6}$) were also excluded. Any SNP with call rate less than 95% was removed; any individual with a call rate less than 95% was also excluded from further study. Two pairs of participants were excluded due to high IBS (Identical by State) in all three cohorts). Additionally, twelve people were excluded due to high autosomal heterozygosity (FDR < 1%). This left 498,916 high-quality SNPS and 4,699 participants in the MCV cohort with mean autosomal heterozygosity of 0.321. The same process yielded 4,591 participants for MPV with the same mean autosomal heterozygosity of 0.321. Similarly, the PC cohort consisted of 4,700 participants with same mean autosomal heterozygosity.

Additionally, a principal component analysis (PCA) was performed to identify principal components to correct for population substructure. Genotype data was pruned to exclude SNPs with high linkage disequilibrium using *PLINK* and standard pruning parameters of 50 SNPs per sliding window; window size of five SNPs; r^2=0.5 [80]. Regression models were adjusted by the first four components, decreasing the genomic inflation factor of all RBC components to $\lambda \leq 1.04$, well within standard ranges [17,27,83].

GWAS

Using *PLINK* v1.9 [84], we performed a simple linear regression analysis with an assumed additive model (number of copies of the minor allele) including age, gender and the first four principal components as covariates to correct for any bias generated by these variables. Standardized values of all three components followed approximate normal distributions (Supplementary Fig S5(A-C) (row 2). Total phenotypic variance explained by the SNPs was calculated by first producing a genetic relationship matrix of all SNPs on autosomal chromosomes in *PLINK*. Subsequently, a restricted

maximum likelihood analysis was conducted using GTCA on the relationship matrix to estimate the variance explained by the SNPS.

A simple one-way ANOVA was performed on the mean RBC component values across the three genotypes. The raw *p*-values associated to the F-test statistic are included in Supplementary Table S2. QUANTO [85] was used to calculate power in our study. We found that for every combination of a SNP's effect size and its MAF, power was greater than 90% based on the approximately 4,500 participants used for each SNP's analysis. Specifically, effect sizes ranged between [-0.62 11.49] (note that greater effect sizes are associated with greater mean component values) and MAFs ranged between [.073, 0.497]. These values are included in Table 2.

PheWAS

The **R** package PheWAS [86] was used to perform two independent PheWAS analyses. The first examined associations between statistically significant SNPs identified in an RBC GWAS and EHR phenotypes based on ICD9 codes. The second identified associations between RBC levels in our cohort and ICD9-based diagnoses only. ICD9 and ICD10 codes for each individual in the cohort recorded in the Renown EHR were aggregated via a mapping from the Center for Medicare and Medicaid services (https://www.cms.gov/Medicare/Coding/ICD10/2018-ICD-10-CM-and-GEMs.html). A total of 34,555 individual diagnoses mapped to 6,632 documented ICD9 codes. ICD9 codes were aggregated and converted into 1,814 individual phenotype groups ("phecodes") using the PheWAS package as described in Carroll and Denny [86,87]. Of these, only the phecodes that included at least 20 cases were used for downstream analyses, following Carroll's protocol [86]: there were 1,488 phecodes with more than 20 cases in each PheWAS. Age, gender, and ethnicity were included in all PheWAS models. The first PheWAS detected associations between statistically significant SNPs ($p<5x10^{-8}$) identified in each of the three GWASs above and case/control status of EHR phenotypes represented by

ICD9 codes. Specifically, a logistic regression between the incidence (number of cases) of each phenotype group (phecode) and the additive genotypes of each statistically significant SNP was performed, using age and gender as covariates. Possible associations of 1,488 phecodes with each previously detected SNP were assessed. The level of statistical significance was computed as a Bonferroni correction for all possible associations per component: p=0.05/ N_p /N_s, where N_p is the number of phecodes tested and N_s is the number of SNPs examined in the specific blood component. This significance level is represented by a red line in Supplementary Fig S3(A-C).

A second PheWAS, as outlined in Carroll et al. (2014) [86], was performed to examine associations between each of the three quantitative RBC components and the phecode categories. Specifically, a linear regression between the RBC measure and the case/control status of a phecode was performed (with age and gender as covariates) for each of 1,488 phecodes. This analysis resulted in a number of hematologic phenotypes that associated with RBC component levels (Table 4). A single-SNP Bonferroni correction $3.4\text{x}10^{-5}$=0.05/N_p (with N_p=1,488) was used to compute the level of statistical significance. Phecodes with association levels $p<3.4\text{x}10^{-5}$ are highlighted in Supplementary Fig S4 (A-C). Note that two associations with MPV at slightly higher p-values ($p=5.43\text{x}10^{-5}$ and $p=6.55\text{x}10^{-5}$) are also included; these are presented in the Discussion.

Table 4. PheWAS Results for MPV, MCV and PC.

Phecode	Description	Group	RBC	β	SE	p	N
287.3	Thrombocytopenia	hematopoietic	MPV	0.75	0.11	$9.06\text{x}10^{-12}$	4104
287	Purpura and other hemorrhagic conditions	hematopoietic	MPV	0.62	0.10	$2.86\text{x}10^{-10}$	4124
286.3	Coagulation defects complicating pregnancy or postpartum	hematopoietic	MPV	2.47	0.39	$3.66\text{x}10^{-10}$	4029
655	Known or suspected fetal abnormality affecting mother	pregnancy complications	MPV	0.56	0.11	$8.85\text{x}10^{-7}$	4455
798	Malaise and fatigue	symptoms	MPV	0.16	0.03	$5.21\text{x}10^{-6}$	4162
261	Vitamin deficiency	endocrine/metabolic	MPV	0.14	0.04	$5.43\text{x}10^{-5}$	4049
61.4	Vitamin D deficiency	endocrine/metabolic	MPV	0.14	0.04	$6.55\text{x}10^{-5}$	3992
282.8	Other hemoglobinopathies	hematopoietic	MCV	-12.18	0.87	$1.97\text{x}10^{-43}$	3751

282	Hereditary hemolytic anemias	hematopoietic	MCV	-10.33	0.82	$8.62x10^{-36}$	3754
280	Iron deficiency anemias	hematopoietic	MCV	-3.76	0.36	$1.30x10^{-25}$	3854
278	Overweight, obesity and other hyperalimentation	endocrine/metabolic	MCV	-1.48	0.14	$3.56x10^{-25}$	4365
278.1	Obesity	endocrine/metabolic	MCV	-1.71	0.17	$2.01x10^{-23}$	3874
280.1	Iron deficiency anemias unspecified or not due to blood loss	hematopoietic	MCV	-3.81	0.38	$5.03x10^{-23}$	3837
278.11	Morbid obesity	endocrine/metabolic	MCV	-2.04	0.21	$6.83x10^{-22}$	3540
281.9	Deficiency anemias	hematopoietic	MCV	8.73	1.10	$2.60x10^{-15}$	3743
289.9	Abnormality of red blood cells	hematopoietic	MCV	-7.30	1.08	$1.94x10^{-11}$	3742
289	Other diseases of blood and blood-forming organs	hematopoietic	MCV	2.84	0.43	$7.37x10^{-11}$	3827
317.11	Alcoholic liver damage	mental disorders	MCV	8.46	1.31	$1.38x10^{-10}$	4009
281	Other deficiency anemia	hematopoietic	MCV	4.55	0.73	$4.33x10^{-10}$	3759
317	Alcohol-related disorders	mental disorders	MCV	4.15	0.67	$5.96x10^{-10}$	4041
317.1	Alcoholism	mental disorders	MCV	5.59	0.91	$9.92x10^{-10}$	4021
571.8	Liver abscess and sequelae of chronic liver disease	digestive	MCV	8.01	1.50	$9.85x10^{-8}$	3855
571.51	Cirrhosis of liver without mention of alcohol	digestive	MCV	7.22	1.42	$3.61x10^{-7}$	3856
342	Hemiplegia	neurological	MCV	-14.80	3.09	$1.67x10^{-6}$	4060
573.2	Liver replaced by transplant	digestive	MCV	14.01	2.99	$2.97x10^{-6}$	3849
571.81	Portal hypertension	digestive	MCV	9.80	2.12	$4.00x10^{-6}$	3851
250.4	Abnormal glucose	endocrine/metabolic	MCV	-0.88	0.19	$6.30x10^{-6}$	3946
250	Diabetes mellitus	endocrine/metabolic	MCV	-1.03	0.23	$9.77x10^{-6}$	3697
250.21	Type 2 diabetes with ketoacidosis	endocrine/metabolic	MCV	12.99	3.04	$1.94x10^{-5}$	3296
530.2	Esophageal bleeding (varices/hemorrhage)	digestive	MCV	8.18	1.91	$1.96x10^{-5}$	3176
70.2	Viral hepatitis B	infectious diseases	MCV	-17.81	4.31	$3.63x10^{-5}$	4266
539	Bariatric surgery	digestive	MCV	-1.98	0.48	$3.71x10^{-5}$	4550
287.3	Thrombocytopenia	hematopoietic	PC	-85.15	6.56	$9.17x10^{-38}$	4104
287	Purpura and other hemorrhagic conditions	hematopoietic	PC	-71.06	5.92	$1.11x10^{-32}$	4124
278	Overweight, obesity and other hyperalimentation	endocrine/metabolic	PC	14.23	1.97	$6.12x10^{-13}$	4366
284	Aplastic anemia	hematopoietic	PC	-98.52	14.12	$3.54\ x10^{-12}$	3749
284.1	Pancytopenia	hematopoietic	PC	-100.90	15.03	$2.16x10^{-11}$	3747
278.1	Obesity	endocrine/metabolic	PC	14.80	2.35	$3.18x10^{-10}$	3875
278.11	Morbid obesity	endocrine/metabolic	PC	17.70	2.92	$1.56\ x10^{-9}$	3541
571.51	Cirrhosis of liver without mention of alcohol	digestive	PC	-120.03	19.85	$1.61x10^{-9}$	3857

571.8	Liver abscess and sequelae of chronic liver disease	digestive	PC	-121.58	21.03	$7.96x10^{-9}$	3856
288.1	Decreased white blood cell count	hematopoietic	PC	-32.91	5.82	$1.68x10^{-8}$	3832
287.31	Primary thrombocytopenia	hematopoietic	PC	-134.74	25.98	$2.24x10^{-7}$	4028
286.3	Coagulation defects complicating pregnancy or postpartum	hematopoietic	PC	-120.29	23.71	$4.10x10^{-7}$	4029
655	Known or suspected fetal abnormality affecting mother	pregnancy complications	PC	-35.93	7.09	$4.27x10^{-7}$	4455
571.81	Portal hypertension	digestive	PC	-143.14	29.73	$1.53x10^{-6}$	3852
288.2	Elevated white blood cell count	hematopoietic	PC	22.82	4.99	$4.88x10^{-6}$	3875
395.2	Nonrheumatic aortic valve disorders	circulatory system	PC	-29.71	6.50	$5.03x10^{-6}$	4019
280.1	Iron deficiency anemias unspecified or not due to blood loss	hematopoietic	PC	24.29	5.74	$2.36x10^{-5}$	3838
555	Inflammatory bowel disease and other gastroenteritis and colitis	digestive	PC	33.57	7.99	$2.70x10^{-5}$	3504

Table of phenotype groups (phecodes) reaching statistical significance ($p<3.4x10^{-5}$) when associated to continuous MPV, MCV and PC component values. Phecodes and their description, effect sizes (β) of the regression, standard error (SE), and *p*-values are included. Each phecode group contains at least 20 cases.

Data Availability Statement

EHR Data

EHR data for the Healthy Nevada cohort are subject to HIPAA and other privacy and compliance restrictions. Mean standardized RBC component values for each individual are available in Supplementary Table S1.

GWAS Results

To reduce the possibility of a privacy breach, 23andMe requires that the statistics for only 10,000 SNPs be made publicly available. This is the amount of data considered by 23andMe to be insufficient to enable a re-identification attack. The statistical summary results of the top 10,000 SNPs for the

23andMe data are available here: https://www.dri.edu/grzymski2020. All column definitions are listed in Table 5.

Table 5. Column Identifiers for GWAS Results.

Column name	Definition
CHR	Chromosome
SNP	Individual SNP identifier
BP	Location of SNP on relative chromosome
A1	Alternative Allele
TEST	Selected statistical test – ADD represents the additive effect
NMISS	Indicates the number of observations – non-missing genotypes
BETA	The effect size for this variant, defined per copy of the A1 allele
SE	The standard error of the effect size
LE	Lower end of the 95% confidence interval for the effect size
UE	Upper end of the 95% confidence interval for the effect size
STAT	The value of the test statistic
P	The p-value for the association test

Table describing the column headers for the results file of our genome wide associations. This summary results file only lists the top 10,000 SNPs in order to prevent a re-identification attack.

PheWAS Results

Summarized counts of each ICD9 classification and phenotype group (phecode) are presented in Supplementary Table 3.

For more information please contact joeg@dri.edu.

Acknowledgements

We thank Michele Henderson, Toni Curreri and all the ambassadors of the Healthy Nevada Project. We also thank Iva Neveux for her helpful discussions with phenotypic data. We thank Renown Health and DRI marketing and all the folks at 23andMe who helped launch the project. Research support was provided by the Governor's Office of Economic Development Knowledge Fund. Support for the Healthy Nevada Project and personal genetics was provided by the Renown Health Foundation.

References

1. Letcher RL, Chien S, Pickering TG, Laragh JH. Elevated blood viscosity in patients with borderline essential hypertension. Hypertension. 1983;5: 757–762. doi:10.1161/01.hyp.5.5.757

2. Sharp DS, Curb JD, Schatz IJ, Meiselman HJ, Fisher TC, Burchfiel CM, et al. Mean red cell volume as a correlate of blood pressure. Circulation. 1996;93: 1677–1684. doi:10.1161/01.cir.93.9.1677

3. Sarnak MJ, Tighiouart H, Manjunath G, MacLeod B, Griffith J, Salem D, et al. Anemia as a risk factor for cardiovascular disease in the atherosclerosis risk in communities (aric) study. J Am Coll Cardiol. 2002;40: 27–33. doi:10.1016/s0735-1097(02)01938-1

4. Simone G de, Devereux RB, Chinali M, Best LG, Lee ET, Welty TK. Association of Blood Pressure With Blood Viscosity in American Indians The Strong Heart Study. Hypertension. 2005;45: 625–630. doi:10.1161/01.hyp.0000157526.07977.ec

5. Chen Z, Tang H, Qayyum R, Schick UM, Nalls MA, Handsaker R, et al. Genome-wide association analysis of red blood cell traits in African Americans: the COGENT Network. Hum Mol Genet. 2013;22: 2529–2538. doi:10.1093/hmg/ddt087

6. Honn KV, Tang DG, Crissman JD. Platelets and cancer metastasis: a causal relationship? Cancer Metastasis Rev. 1992;11: 325–351.

7. Zoppo GJD. The role of platelets in ischemic stroke. Neurology. 1998;51: S9–S14. doi:10.1212/wnl.51.3_suppl_3.s9

8. Pain A, Ferguson DJP, Kai O, Urban BC, Lowe B, Marsh K, et al. Platelet-mediated clumping of Plasmodium falciparum-infected erythrocytes is a common adhesive phenotype and is associated with severe malaria. Proc Natl Acad Sci U S A. 2001;98: 1805–1810. doi:10.1073/pnas.98.4.1805

9. Willoughby S, Holmes A, Loscalzo J. Platelets and cardiovascular disease. Eur J Cardiovasc Nurs. 3rd ed. 2002;1: 273–288. doi:10.1016/S1474-51510200038-5

10. McBane RD, Karnicki K, Miller RS, Owen WG. The impact of peripheral arterial disease on circulating platelets. Thromb Res. 2004;113: 137–145. doi:10.1016/j.thromres.2004.02.007

11. Weber C. Platelets and chemokines in atherosclerosis: partners in crime. Circ Res. 2005;96: 612–616. doi:10.1161/01.RES.0000160077.17427.57

12. Jain S, Harris J, Ware J. Platelets: linking hemostasis and cancer. Arterioscler Thromb Vasc Biol. 2010;30: 2362–2367. doi:10.1161/ATVBAHA.110.207514

13. 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. doi:10.1182/blood-2008-10-184234

14. Alper AT, Sevimli S, Hasdemir H, Nurkalem Z, Güvenç TS, Akyol A, et al. Effects of high altitude and sea level on mean platelet volume and platelet count in patients with acute coronary syndrome. J Thromb Thrombolysis. 3rd ed. Springer US; 2009;27: 130–134. doi:10.1007/s11239-007-0159-9

15. Hudson JG, Bowen AL, Navia P, Rios-Dalenz J, Pollard AJ, Williams D, et al. The effect of high altitude on platelet counts, thrombopoietin and erythropoietin levels in young Bolivian airmen visiting the Andes. Int J Biometeorol. Springer-Verlag; 1999;43: 85–90. doi:10.1007/s004840050120

16. Shrivastava A, Goyal A, (null) KN. Effect of high altitude on haematological parameters. Indian J Prev Soc Med. 2010;41: 2.

17. Soranzo N, Spector TD, Mangino M, Kühnel 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. doi:10.1038/ng.467

18. Shameer K, Denny JC, Ding K, Jouni H, Crosslin DR, Andrade M de, 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. doi:10.1007/s00439-013-1355-7

19. Lin J-P, O'Donnell CJ, Jin L, Fox C, Yang Q, Cupples LA. Evidence for linkage of red blood cell size and count: genome-wide scans in the Framingham Heart Study. Am J Hematol. 2007;82: 605–610. doi:10.1002/ajh.20868

20. 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. Proc Natl Acad Sci U S A. 2007;104: 11346–11351. doi:10.1073/pnas.0611393104

21. Lettre G, Sankaran VG, Bezerra MAC, Araújo 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. Proc Natl Acad Sci U S A. 2008;105: 11869–11874. doi:10.1073/pnas.0804799105

22. Ferreira MAR, Hottenga J-J, Warrington NM, Medland SE, Willemsen G, Lawrence RW, et al. Sequence variants in three loci influence monocyte counts and erythrocyte volume. Am J Hum Genet. 2009;85: 745–749. doi:10.1016/j.ajhg.2009.10.005

23. 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: 1191–1198. doi:10.1038/ng.466

24. Meisinger C, Prokisch H, Gieger C, Soranzo N, Mehta D, Rosskopf D, et al. A Genome-wide Association Study Identifies Three Loci Associated with Mean Platelet Volume. Am J Hum Genet. 2009;84: 66–71. doi:10.1016/j.ajhg.2008.11.015

25. Daly ME. Determinants of platelet count in humans. Haematologica. 2010;96: 10–13. doi:10.3324/haematol.2010.035287

26. 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. Nature Publishing Group; 2010;42: 210–215. doi:10.1038/ng.531

27. Kullo IJ, Ding K, Jouni H, Smith CY, Chute CG. A Genome-Wide Association Study of Red Blood Cell Traits Using the Electronic Medical Record. PLoS ONE. 2010;5: e13011. doi:10.1371/journal.pone.0013011

28. 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. doi:10.1038/nature10659

29. 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. doi:10.1371/journal.pgen.1002067

30. Paul DS, Nisbet JP, Yang T-P, Meacham S, Rendon A, Hautaviita K, et al. Maps of Open Chromatin Guide the Functional Follow-Up of Genome-Wide Association Signals: Application to Hematological Traits. PLoS Genet. 2011;7: e1002139. doi:10.1371/journal.pgen.1002139

31. An P, Wu Q, Wang H, Guan Y, Mu M, Liao Y, et al. TMPRSS6, but not TF, TFR2 or BMP2 variants are associated with increased risk of iron-deficiency anemia. Hum Mol Genet. 2012;21: 2124–2131. doi:10.1093/hmg/dds028

32. Ding K, Shameer K, Jouni H, Masys DR, Jarvik GP, Kho AN, et al. Genetic Loci implicated in erythroid differentiation and cell cycle regulation are associated with red blood cell traits. Mayo Clin Proc. 2012;87: 461–474. doi:10.1016/j.mayocp.2012.01.016

33. 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. doi:10.1093/hmg/dds534

34. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, et al. Systematic Localization of Common Disease-Associated Variation in Regulatory DNA. Science. 2012;337: 1190–1195. doi:10.1126/science.1222794

35. 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. doi:10.1371/journal.pgen.1002491

36. van der Harst P, Zhang W, Mateo Leach I, Rendon A, Verweij N, Sehmi J, et al. Seventy-five genetic loci influencing the human red blood cell. Nature. 2012;492: 369–375. doi:10.1038/nature11677

37. Cardoso GL, Diniz IG, Silva ANLMD, Cunha DA, Silva Junior JSD, Uchôa CTC, et al. DNA polymorphisms at BCL11A, HBS1L-MYB and Xmn1-HBG2 site loci associated with fetal hemoglobin levels in sickle cell anemia patients from Northern Brazil. Blood Cells Mol Dis. 2014;53: 176–179. doi:10.1016/j.bcmd.2014.07.006

38. Pei S-N, Ma M-C, You H-L, Fu H-C, Kuo C-Y, Rau K-M, et al. TMPRSS6 rs855791 Polymorphism Influences the Susceptibility to Iron Deficiency Anemia in Women at Reproductive Age. Int J Med Sci. 2014;11: 614–619. doi:10.7150/ijms.8582

39. Grote Beverborg N, Verweij N, Klip IT, van der Wal HH, Voors AA, van Veldhuisen DJ, et al. Erythropoietin in the general population: reference ranges and clinical, biochemical and genetic correlates. PLoS ONE. 2015;10: e0125215. doi:10.1371/journal.pone.0125215

40. Mtatiro SN, Mgaya J, Singh T, Mariki H, Rooks H, Soka D, et al. Genetic association of fetal-hemoglobin levels in individuals with sickle cell disease in Tanzania maps to conserved regulatory elements within the MYB core enhancer. BMC Med Genet. 2nd ed. 2015;16: 4. doi:10.1186/s12881-015-0148-3

41. 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. Nature Publishing Group; 2015;6: 1–11. doi:10.1038/ncomms7691

42. Lai Y, Chen Y, Chen B, Zheng H, Yi S, Li G, et al. Genetic Variants at BCL11A and HBS1L-MYB loci Influence Hb F Levels in Chinese Zhuang β-Thalassemia Intermedia Patients. Hemoglobin. 2016;40: 405–410. doi:10.1080/03630269.2016.1253586

43. Maharry SE, Walker CJ, Liyanarachchi S, Mehta S, Patel M, Bainazar MA, et al. Dissection of the Major Hematopoietic Quantitative Trait Locus in Chromosome 6q23.3 Identifies miR-3662 as a Player in Hematopoiesis and Acute Myeloid Leukemia. Cancer Discovery. 2016;6: 1036–1051. doi:10.1158/2159-8290.CD-16-0023

44. Mikobi TM, Tshilobo Lukusa P, Aloni MN, Lumaka AZ, Kaba DK, Devriendt K, et al. Protective BCL11A and HBS1L-MYB polymorphisms in a cohort of 102 Congolese patients suffering from sickle cell anemia. J Clin Lab Anal. 2018;32. doi:10.1002/jcla.22207

45. Seiki T, Naito M, Hishida A, Takagi S, Matsunaga T, Sasakabe T, et al. Association of genetic polymorphisms with erythrocyte traits: Verification of SNPs reported in a previous GWAS in a Japanese population. Gene. 2018;642: 172–177. doi:10.1016/j.gene.2017.11.031

46. Verma A, Basile AO, Bradford Y, Kuivaniemi H, Tromp G, Carey D, et al. Phenome-Wide Association Study to Explore Relationships between Immune System Related Genetic Loci and Complex Traits and Diseases. PLoS ONE. 2016;11: e0160573. doi:10.1371/journal.pone.0160573

47. Verma A, Lucas A, Verma SS, Zhang Y, Josyula N, Khan A, et al. PheWAS and Beyond: The Landscape of Associations with Medical Diagnoses and Clinical Measures across 38,662

Individuals from Geisinger. Am J Hum Genet. 2018;102: 592–608. doi:10.1016/j.ajhg.2018.02.017

48. 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. doi:10.1016/j.ajhg.2010.11.011

49. Yuan X, Waterworth D, Perry JRB, Lim N, Song K, Chambers JC, et al. Population-Based Genome-wide Association Studies Reveal Six Loci Influencing Plasma Levels of Liver Enzymes. Am J Hum Genet. 2008;83: 520–528. doi:10.1016/j.ajhg.2008.09.012

50. Panova-Noeva M, Schulz A, Hermanns MI, Grossmann V, Pefani E, Spronk HMH, et al. Sex-specific differences in genetic and nongenetic determinants of mean platelet volume: results from the Gutenberg Health Study. Blood. 2016;127: 251–259. doi:10.1182/blood-2015-07-660308

51. Chasman DI, Paré G, Mora S, Hopewell JC, Peloso G, Clarke R, et al. Forty-three loci associated with plasma lipoprotein size, concentration, and cholesterol content in genome-wide analysis. PLoS Genet. 2009;5: e1000730. doi:10.1371/journal.pgen.1000730

52. Chambers JC, Zhang W, Sehmi J, Li X, Wass MN, van der Harst P, et al. Genome-wide association study identifies loci influencing concentrations of liver enzymes in plasma. Nat Genet. 2011;43: 1131–1138. doi:10.1038/ng.970

53. Jin G, Sun J, Kim S-T, Feng J, Wang Z, Tao S, et al. Genome-wide association study identifies a new locus JMJD1C at 10q21 that may influence serum androgen levels in men. Hum Mol Genet. 2012;21: 5222–5228. doi:10.1093/hmg/dds361

54. Coviello AD, Haring R, Wellons M, Vaidya D, Lehtimäki T, Keildson S, et al. A genome-wide association meta-analysis of circulating sex hormone-binding globulin reveals multiple Loci implicated in sex steroid hormone regulation. PLoS Genet. 2012;8: e1002805. doi:10.1371/journal.pgen.1002805

55. Grigorova M, Punab M, Poolamets O, Adler M, Vihljajev V, Laan M. Genetics of Sex Hormone-Binding Globulin and Testosterone Levels in Fertile and Infertile Men of Reproductive Age. J Endocr Soc. 2017;1: 560–576. doi:10.1210/js.2017-00050

56. Tajuddin SM, Schick UM, Eicher JD, Chami N, Giri A, Brody JA, et al. Large-Scale Exome-wide Association Analysis Identifies Loci for White Blood Cell Traits and Pleiotropy with Immune-Mediated Diseases. Am J Hum Genet. 2016;99: 22–39. doi:10.1016/j.ajhg.2016.05.003

57. Smyth DJ, Plagnol V, Walker NM, Cooper JD, Downes K, Yang JHM, et al. Shared and distinct genetic variants in type 1 diabetes and celiac disease. N Engl J Med. 2008;359: 2767–2777. doi:10.1056/NEJMoa0807917

58. de Boer YS, van Gerven NMF, Zwiers A, Verwer BJ, van Hoek B, van Erpecum KJ, et al. Genome-wide association study identifies variants associated with autoimmune hepatitis type 1. Gastroenterology. 2014;147: 443–52.e5. doi:10.1053/j.gastro.2014.04.022

59. Giusti B, Marcucci R, Saracini C, Gori AM, Valenti R, Parodi G, et al. Mean platelet volume and platelet count in acute coronary syndrome patients: role of a genetic variants on chr7q22.3

and chr3p13-p21. Eur Heart J. 2013;34: P4879–P4879. doi:10.1093/eurheartj/eht310.p4879

60. Johnson AD. The genetics of common variation affecting platelet development, function and pharmaceutical targeting. J Thromb Haemost. 2011;9 Suppl 1: 246–257. doi:10.1111/j.1538-7836.2011.04359.x

61. Zou S, Teixeira AM, Kostadima M, Astle WJ, Radhakrishnan A, Simon LM, et al. SNP in human ARHGEF3 promoter is associated with DNase hypersensitivity, transcript level and platelet function, and Arhgef3 KO mice have increased mean platelet volume. PLoS ONE. 2017;12: e0178095. doi:10.1371/journal.pone.0178095

62. Goggs R, Williams CM, Mellor H, Poole AW. Platelet Rho GTPases-a focus on novel players, roles and relationships. Biochem J. 2015;466: 431–442. doi:10.1042/BJ20141404

63. Kojima H, Kanada H, Shimizu S, Kasama E, Shibuya K, Nakauchi H, et al. CD226 Mediates Platelet and Megakaryocytic Cell Adhesion to Vascular Endothelial Cells. J Biol Chem. 2003;278: 36748–36753. doi:10.1074/jbc.M300702200

64. Crowder RJ, Enomoto H, Yang M, Johnson EM, Milbrandt J. Dok-6, a Novel p62 Dok Family Member, Promotes Ret-mediated Neurite Outgrowth. J Biol Chem. 2004;279: 42072–42081. doi:10.1074/jbc.M403726200

65. Konstantakis C, Tselekouni P, Kalafateli M, Triantos C. Vitamin D deficiency in patients with liver cirrhosis. Ann Gastroenterol. 2016;29: 297–306. doi:10.20524/aog.2016.0037

66. Massey AC. Microcytic anemia. Differential diagnosis and management of iron deficiency anemia. Med Clin North Am. 1992;76: 549–566.

67. Evstatiev R, Bukaty A, Jimenez K, Kulnigg Dabsch S, Surman L, Schmid W, et al. Iron deficiency alters megakaryopoiesis and platelet phenotype independent of thrombopoietin. Am J Hematol. Wiley Online Library; 2014;89: 524–529. doi:10.1002/ajh.23682

68. Besancenot R, Roos-Weil D, Tonetti C, Abdelouahab H, Lacout C, Pasquier F, et al. JAK2 and MPL protein levels determine TPO-induced megakaryocyte proliferation vs differentiation. Blood. 2014;124: 2104–2115. doi:10.1182/blood-2014-03-559815

69. Noris P, Klersy C, Gresele P, Giona F, Giordano P, Minuz P, et al. Platelet size for distinguishing between inherited thrombocytopenias and immune thrombocytopenia: a multicentric, real life study. Br J Haematol. 2013;162: 112–119. doi:10.1111/bjh.12349

70. Mozos I, Marginean O. Links between Vitamin D Deficiency and Cardiovascular Diseases. Biomed Res Int. Hindawi; 2015;2015: 109275–12. doi:10.1155/2015/109275

71. Cumhur Cure M, Cure E, Yuce S, Yazici T, Karakoyun I, Efe H. Mean platelet volume and vitamin D level. Ann Lab Med. 2014;34: 98–103. doi:10.3343/alm.2014.34.2.98

72. Park YC, Kim J, Seo MS, Hong SW, Cho ES, Kim J-K. Inverse relationship between vitamin D levels and platelet indices in Korean adults. Hematology. 2017;22: 1–7. doi:10.1080/10245332.2017.1318334

73. Gaspar RS, Trostchansky A, Paes AM de A. Potential Role of Protein Disulfide Isomerase in

Metabolic Syndrome-Derived Platelet Hyperactivity. Oxid Med Cell Longev. Hindawi; 2016;2016: 2423547–10. doi:10.1155/2016/2423547

74. Vaidya D, Yanek LR, Faraday N, Moy TF, Becker LC, Becker DM. Native platelet aggregation and response to aspirin in persons with the metabolic syndrome and its components. Metab Syndr Relat Disord. 2009;7: 289–296. doi:10.1089/met.2008.0083

75. Samocha-Bonet D, Justo D, Rogowski O, Saar N, Abu-Abeid S, Shenkerman G, et al. Platelet counts and platelet activation markers in obese subjects. Mediators Inflamm. 2008;2008: 834153. doi:10.1155/2008/834153

76. Kurokawa T, Ohkohchi N. Platelets in liver disease, cancer and regeneration. World J Gastroenterol. 2017;23: 3228–3239. doi:10.3748/wjg.v23.i18.3228

77. Chauhan A, Adams DH, Watson SP, Lalor PF. Platelets: No longer bystanders in liver disease. Hepatology. Wiley-Blackwell; 2016;64: 1774–1784. doi:10.1002/hep.28526

78. Míguez-Burbano MJ, Nair M, Lewis JE, Fishman J. The role of alcohol on platelets, thymus and cognitive performance among HIV-infected subjects: are they related? Platelets. 2009;20: 260–267. doi:10.1080/09537100902964759

79. Mikhailidis DP, Jenkins WJ, Barradas MA, Jeremy JY, Dandona P. Platelet function defects in chronic alcoholism. Br Med J (Clin Res Ed). British Medical Journal Publishing Group; 1986;293: 715–718. doi:10.1136/bmj.293.6549.715

80. Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. Data quality control in genetic case-control association studies. Nature Protocols. Nature Publishing Group; 2010;5: 1564–1573. doi:10.1038/nprot.2010.116

81. Schlauch KA, Khaiboullina SF, De Meirleir KL, Rawat S, Petereit J, Rizvanov AA, et al. Genome-wide association analysis identifies genetic variations in subjects with myalgic encephalomyelitis/chronic fatigue syndrome. Transl Psychiatry. 2016;6: e730–e730. doi:10.1038/tp.2015.208

82. Schlauch KA, Kulick D, Subramanian K, De Meirleir KL, Palotás A, Lombardi VC. Single-nucleotide polymorphisms in a cohort of significantly obese women without cardiometabolic diseases. Int J Obes (Lond). Nature Publishing Group; 2018;106: 1656. doi:10.1038/s41366-018-0181-3

83. Winkler TW, Day FR, Croteau-Chonka DC, Wood AR, Locke AE, Mägi R, et al. Quality control and conduct of genome-wide association meta-analyses. Nature Protocols. 2014;9: 1192–1212. doi:10.1038/nprot.2014.071

84. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, 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. doi:10.1086/519795

85. Gauderman WJ. Sample size requirements for matched case-control studies of gene–environment interaction. Stat Med. Wiley Online Library; 2002;21: 35–50. doi:10.1002/sim.973

86. Carroll RJ, Bastarache L, Denny JC. R PheWAS: data analysis and plotting tools for phenome-wide association studies in the R environment. Bioinformatics. 2014;30: 2375–2376. doi:10.1093/bioinformatics/btu197

87. Denny JC, Bastarache L, Ritchie MD, Carroll RJ, Zink R, Mosley JD, et al. Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. Nat Biotechnol. 2013;31: 1102–1110. doi:10.1038/nbt.2749

Figure Legends

Fig 1: MCV GWAS Manhattan plot

Genome-wide association study results for MCV. The x-axis represents the genomic position of 498,916 SNPs. The y-axis represents -$\log_{10}$-transformed raw *p*-values of each genotypic association. The red horizontal line indicates the threshold of significance $p=5x10^{-8}$.

Fig 2: MCV PheWAS plot

This figure illustrates the results of individual linear regression between incidence of phenotype groups (phecodes) and continuous MCV component measures. The model includes age, gender and ethnicity as covariates. Each point represents the *p*-value of the association between one of 1,488 phecodes with at least 20 cases assigned to it, and the MCV component measure. The horizontal red line represents the significance level $p=3.4x10^{-5}$.

Supplemental Figure and Table Legends

Supplementary Table 1: Mean standardized RBC component values

This table includes mean standardized RBC component values for each individual along with age and gender. Due to the length of this table it can be found online at https://www.dri.edu/grzymski2020

Supplementary Table 2: General SNP table for MPV, MCV and PC

This table lists the 38 statistically significant SNPs associated to MPV, MCV and PC in our cohort. General information about the SNP such as chromosome location, GWAS *p*-value, power, genotype, cytoband, ANOVA, and references of associations identified in previous studies are listed.

Supplementary Table 3: Counts of each phecode group

This table presents the mapping between ICD9 codes and phecodes as presented in Carroll and the **R** package PheWAS [86] tested in our study, and the number of incidences from the RBC cohort in each phecode group.

Supplementary Fig S1 (A, B, C): GWAS results for RBC components MPV, MCV and PC

Genome-wide association study results for the three RBC components. The x-axis represents the genomic position of 498,916 SNPs. The y-axis represents $-\log_{10}$-transformed raw *p*-values of each genotypic association. The red horizontal line indicates the threshold of significance $p=5\text{x}10^{-8}$.

Supplementary Fig 2: ANOVA results of SNP rs7961894

This figure shows the box and whisker diagram for standardized values of MPV of all members in the cohort based on genotype. Mean and standard deviation values for each genotype are CC: 10.54 ± 0.97; CT: 10.74 ± 1.0; TT: 11.21 ± 0.87. The *p*-value for this ANOVA analysis is $p=8.7\text{x}10^{-12}$.

Supplementary Fig S3 (A, B, C): PheWAS results between RBC component-significant SNPs and phecodes

These three figures show the results of individual logistic regressions between incidence of phenotype groups (phecodes) and SNP genotypes, based on the additive model. Models include age, gender and ethnicity as covariates. Each point represents the *p*-value of one SNP and one of 1,488 phecodes with at least 20 cases assigned to it. The horizontal red line in each represents the significance level $p=1.60\text{x}10^{-6}$ for MPV, $p=2.40\text{x}10^{-6}$ for MCV, and $p=1.12\text{x}10^{-5}$ for PC.

Supplementary Fig S4 (A, B, C): PheWAS results between RBC component and phecodes

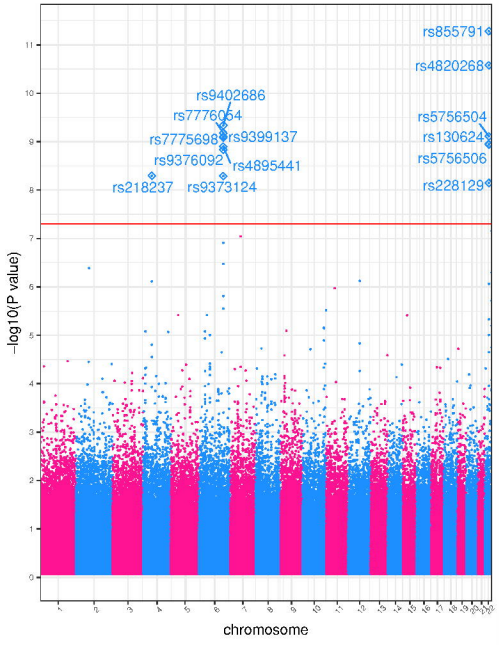
These three figures show the results of individual linear regressions between incidence of phenotype groups (phecodes) and continuous RBC component measures. Models include age, gender and ethnicity as covariates. Each point represents the *p*-value of the association between one of 1,488 phecodes with at least 20 cases assigned to it, and the RBC component measure. The horizontal red line in each represents the significance level $p=1.60\text{x}10^{-6}$ for MPV, $p=2.40\text{x}10^{-6}$ for MCV, and $p=1.12\text{x}10^{-5}$ for PC.

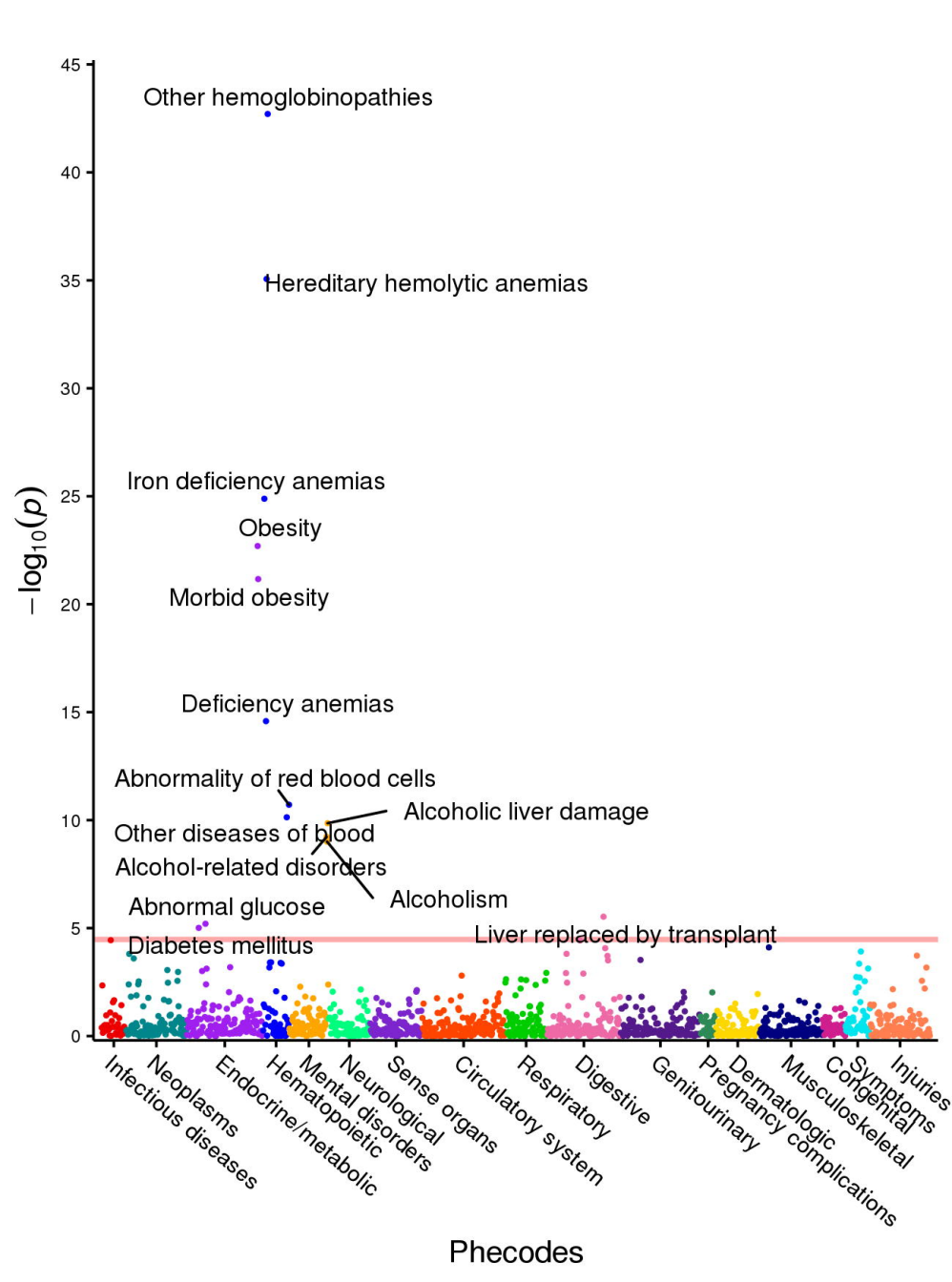
Supplementary Fig S5 (A, B, C): Raw and Standardized RBC Component Lab Measures

Distribution of raw RBC component values are presented in the first row; distribution of component values upon standardization to the most recent lab test are shown in the second row; the QQ-plot of the standardized values is pictured in the third row.

Author Contributions

KAS and RWR conducted genetic and clinical data analysis and wrote the manuscript. GE, ADS and KAS contributed to the clinical discussion. WJM extracted participants and their clinical health data from the Renown EHR. RA and the 23andMe research team provided participant genotype data and edited the manuscript. JJG conceived of and obtained funds to conduct this experiment. All authors reviewed, edited and approved the final version of the manuscript.


rs855791
rs4820268
rs9402686
rs7776054
rs5756504
rs7775698
rs9399137
rs130624
rs9376092
rs5756506
rs4895441
rs218237
rs9373124
rs228129
−log10(P value)
chromosome


Other hemoglobinopathies
Hereditary hemolytic anemias
Iron deficiency anemias
Obesity
Morbid obesity
Deficiency anemias
Abnormality of red blood cells
Alcoholic liver damage
Other diseases of blood
Alcohol-related disorders
Alcoholism
Abnormal glucose
Diabetes mellitus
Liver replaced by transplant
45
40
35
30
25
20
15
10
5
0
$-\log_{10}(p)$
Infectious diseases
Neoplasms
Endocrine/metabolic
Hematopoietic
Mental disorders
Neurological
Sense organs
Circulatory system
Respiratory
Digestive
Genitourinary
Pregnancy complications
Dermatologic
Musculoskeletal
Congenital
Symptoms
Injuries
Phecodes