Pharmacogenomics and Personalized Medicine

Dovepress

open access to scientific and medical research

Open Access Full Text Article

ORIGINAL RESEARCH

Genome-Wide Association Study on the Hematological Phenotypic Characteristics of the Han Population from Northwest China

Wei Yang[1–3], Xue He[1,2], Yuying Yao[2], Hongyan Lu[2], Yuliang Wang[2], Zhanhao Zhang[2], Yuhe Wang[1,2,4], Li Wang[1,2], Yongjun He[1,2], Dongya Yuan[1,2], Tianbo Jin[1,2]

[1]Key Laboratory of Molecular Mechanism and Intervention Research for Plateau Diseases of Tibet Autonomous Region, School of Medicine, Xizang Minzu University, Xianyang, Shaanxi, 712082, People's Republic of China; [2]School of Basic Medical Sciences, Xizang Minzu University, Xianyang, Shaanxi, 712082, People's Republic of China; [3]Department of Emergency, The Affiliated Hospital of Xizang Minzu University, Xianyang, Shaanxi, 712082, People's Republic of China; [4]Department of Clinical Laboratory, The Affiliated Hospital of Xizang Minzu University, Xianyang, Shaanxi, 712082, People's Republic of China

Correspondence: Tianbo Jin, Key Laboratory of Molecular Mechanism and Intervention Research for Plateau Diseases of Tibet Autonomous Region, School of Medicine, Xizang Minzu University, #6 East Wenhui Road, Xianyang, Shaanxi, 712082, People's Republic of China, Tel/Fax +86-29-88895902, Email tianbo__jin@163.com

Background: Hematological characteristics have positive reference value as clinical indicators in the evaluation of various diseases. The purpose of this study was to determine the gene loci associated with 20 hematological phenotypes in the Han population from northwest China.

Methods: A genome-wide association study (GWAS) was conducted on hematological indicators of 1005 Han people from northwest China. Genotyping was performed with a GeneTitan multichannel instrument and Axiom Analysis Suite 6.0. Using the 1000 Genomes Project (phase 3) as a reference, haplotype imputation was performed with IMPUTE2. SNVs (single nucleotide variants) significantly associated with hematological phenotypes were identified. The top SNV ($p < 5E-7$) was then selected for replication detection.

Results: Ninety genetic variations identified in the GWAS were significantly associated with hematological indicators. Among them, only rs35289401 (CCDC157) was significantly associated (genome-wide) with red blood cell distribution width (RDW) ($p = 4.21E-08$). The fourteen top SNVs were selected for replication verification and were significantly associated with hematological phenotypes. However, only HBS1 L-MYB rs1331309 was significantly associated with the mean hemoglobin content ($p = 6.42E-07$). We also found that the mean corpuscular hemoglobin (MCH) level in the rs1331309 GG/GT genotype was significantly higher than that in the TT genotype ($p = 0.023$).

Conclusion: The GWAS identified a total of 90 genetic variants significantly associated with hematological phenotypic indicators. In particular, rs1331309 (HBS1 L-MYB) is expected to be a biomarker for monitoring the dynamics of MCH levels. This study provides a reference for related studies on the genetic structure of hematological characteristics. It provides a valuable reference for the clinical diagnosis or prediction of a variety of diseases.

Keywords: hematological, GWAS, hematological phenotype, MCH, Han population from northwest China

Introduction

The detection of hematological components has positive reference value for many diseases in the clinic. Hematology characteristics mainly include three cell lineages: red blood cells (RBC), white blood cells (WBC) and platelets (PLT).[1] Phenotypic indicators related to these lineages are commonly used as clinical parameters and can be used to monitor immune function. In addition, they can also be used as biomarkers for monitoring the severity of a disease.[2,3] The deviation of hematological phenotypic indicators outside the normal range varies with the type of disease such as immune diseases, cancer, inflammation, and cardiovascular diseases.[3] Studies have shown that hematological characteristics are highly hereditary.[4,5]

Genome-wide association studies (GWASs) have become the main method to study complex diseases and their susceptibility genes because of their ability to encompass single nucleotide polymorphisms (SNPs) for the whole genome.[6] This type

Received: 22 February 2022
Accepted: 14 June 2022
Published: 3 August 2022

of study can efficiently find the gene loci associated with the occurrence and development of a disease. GWASs conduct a population-level statistical analysis of genotypes and phenotypes to determine the phenotypic changes associated with gene loci. Therefore, it is highly feasible to identify genetic polymorphisms associated with hematological phenotypic indicators through a GWAS, which will help us understand the genetic structure of hematological characteristics at a deeper level.

To date, GWASs of hematological phenotypic indicators have been performed among populations with diverse genetic backgrounds, including European,[3] Korean,[7] Caucasian and African American,[8,9] and Japanese populations.[10,11] However, the genetic structure of the hematological characteristics of the Han population from northwest China has not been described thus far.

Therefore, we performed a GWAS on 20 phenotypic indicators of hematological characteristics in the Han population from northwest China to identify the gene loci associated with hematological phenotypic indicators. Our study will supplement data on genetic variation associated with hematological characteristics, which will help to further explore the genetic structure of these hematological characteristics. Our study will provide a valuable reference for the clinical monitoring of human diseases.

Materials and Methods

Study Subjects and DNA Extraction

The research group consisted of 1005 participants (494 men and 511 women) from the health examination center of the Affiliated Hospital of Xizang Minzu University. The inclusion criteria of participants were as follows: healthy people without disease, no family history of disease, no medication (for at least two weeks), and no pregnancy. We orally informed each participant of the purpose of this study, and all participants signed informed consent forms. The content of the informed consent mainly includes the background, purpose, method, significance and privacy policy of this study. After obtaining written informed consent from all participants, whole blood was collected. Our study was conducted under the standard approved by the ethics committee of the Affiliated Hospital of Xizang Minzu University. We extracted whole genomic DNA according to the kit instructions (GoldMag, Xi'an). Subsequently, a GWAS was performed on 20 hematological phenotypic indicators. These 20 indicators were white blood cell count (WBC), percent lymphocytes (LYMPH%), percent mononuclear cells (MONO%), percent neutrophils (NEUT%), percent eosinophils (EO%), percent basophils (BASO%), absolute monocyte count (MONO), absolute neutrophil count (NEUT), eosinophil count (EO), absolute value of basophils (BASO), platelet count (PLT count), platelet distribution width (PDW), mean platelet volume (MPV), red blood cell count (RBC count), hemoglobin (HGB), mean corpuscular volume (MCV), mean corpuscular hemoglobin (MCH), mean corpuscular hemoglobin concentration (MCHC), percent red cell distribution width (RDW%), and red blood cell distribution width (RDW).

Genotyping and Quality Control

A Thermo Scientific Genotyping Chip (Applied Biosystems, Axiom Precision Medicine Diversity Array, PMDA) was used. We used a Gene Titan multichannel instrument and Axiom Analysis Suite 6.0 software for genotyping. Within the scope of our sample, we performed a full gene scan through Axiom, and a total of 874,190 loci were detected. Indels, copy number variation, sex chromosomes and duplicate sites were excluded, and then the necessary quality control was performed on the remaining sites (sample call rate > 0.95, maker call rate > 0.90, and HWE > 5×10^{-6}). Ultimately, 796,288 loci remained before imputation. After excluding indels, copy number variation, sex chromosomes and duplicate sites, the necessary quality control was performed on the remaining loci (sample call rate > 0.95, maker call rate > 0.90, HWE > 5×10^{-6}). In the end, 796,288 loci remained before imputation.

Imputation and Quality Control

We used IMPUTE2 software and the haplotype of the 1000 Genomes Project (phase 3) as a reference for imputation. After the imputation was completed, sites that met the following conditions were retained: sample call rate > 95%, marker call rate > 90%, HWE-control > 5×10^{-6}, and allele = 2. Ultimately, a total of 9,336,679 SNVs were used for the subsequent analysis.

Statistical Analysis

An automatic Triad hematology analyzer (Mindray; BC-2800) was used to measure the levels of 20 hematological phenotypic indicators. The levels of all hematological phenotypic indicators were expressed as the mean ± SD, and relevant statistical

analysis was conducted using SPSS 22.0 software (SPSS Inc., Chicago, IL, USA). Our study used the single-locus mixed model algorithm implemented in SNP & Variation Suite v 8.7 (Golden Helix Inc., Bozeman, MT) to perform the genome-wide association study.[12] Based on the SNP & Variation Suite manual (https://doc.goldenhelix.com/SVS/latest/svs_index.html), we used the mixed linear-additive genetic model and added the IBD matrix to the model to detect SNVs associated with the phenotypic indicators. To remove the influence of confounding factors, all results were adjusted by age and sex. Also, Manhattan plots and quantile figures related to each indicator were drawn. In this study, a p value < 5E-8 was considered to indicate a significant association between the SNV and the phenotypic indicator with genome-wide significance. When the p value < 5E-6, it suggested that the SNV may have a genome-wide significant association with the phenotype indicator.[13]

Verification of Replication

After the completion of the GWAS, SNVs with p value < 5×10^{-7} were selected as top SNVs for replication testing, which have been suggested to have genome-wide significance with phenotypic indicators. We recruited 2047 participants at the Affiliated Hospital of Xizang Minzu University for replication verification. In this study, the extreme measurement values (mean > ±3 SD) of each phenotypic indicator were excluded and then normalized with rankbaseINTs. The association analysis between the top SNVs and phenotypic indicators was then performed to select SNVs that were still significantly associated with phenotypic indicators in replication testing. Finally, SPSS software was used to draw a box plot of phenotypic indicator distribution under different genotypes of significant gene loci. In the replication test, a p value < 0.05 was considered significant.

Results

A total of 1005 Chinese Han people were recruited as the research subjects. The basic characteristics of the research subjects, the ratio of males to females and the average levels of various phenotypic indicators are summarized in Table 1.

Table 1 Sample Information of Different Hematological Phenotypic Indicators

Indicator	N	Mean ± SD	Age (Years) Mean ± SD	Gender	
				Male	Female
WBC	996	5.79 ± 1.47	43.41 ± 9.89	490 (49.20%)	506 (50.80%)
LYMPH%	1000	32.74 ± 7.03	43.39 ± 9.85	490 (49.00%)	510 (51.00%)
MONO%	997	6.87 ± 1.68	43.36 ± 9.89	492 (49.35%)	505 (50.65%)
NEUT%	1000	57.77 ± 7.55	43.38 ± 9.87	491 (49.10%)	509 (50.90%)
EO%	987	1.90 ± 1.26	43.32 ± 9.86	484 (49.04%)	503 (50.96%)
BASO%	989	0.47 ± 0.25	43.40 ± 9.89	488 (49.34%)	501 (50.66%)
MONO	996	0.40 ± 0.14	43.43 ± 9.88	488 (49.00%)	508 (51.00%)
NEUT	995	3.36 ± 1.07	43.41 ± 9.90	491 (49.35%)	504 (50.65%)
EO	990	0.11 ± 0.08	43.35 ± 9.91	482 (48.69%)	508 (51.31%)
BASO	1005	0.03 ± 0.11	43.38 ± 9.88	494 (49.15%)	511 (50.85%)
PLT	1003	228.68 ± 57.00	43.41 ± 9.87	493 (49.15%)	510 (50.85%)
PDW	966	13.99 ± 2.73	43.42 ± 9.85	480 (49.69%)	486 (50.31%)
MPV	965	11.06 ± 1.17	43.41 ± 9.85	479 (49.64%)	486 (50.36%)
RBC	965	0.25 ± 0.07	43.40 ± 9.87	481 (49.84%)	484 (50.16%)
HGB	997	146.26 ± 17.49	43.41 ± 9.87	493 (49.45%)	504 (50.55%)
MCV	1000	43.28 ± 4.60	43.36 ± 9.86	492 (49.20%)	508 (50.80%)
MCH	1001	89.52 ± 6.67	43.35 ± 9.89	491 (49.05%)	510 (50.95%)
MCHC	1003	30.12 ± 2.46	43.36 ± 9.88	494 (49.25%)	509 (50.75%)
RDW%	1004	335.59 ± 12.87	43.37 ± 9.88	494 (49.20%)	510 (50.80%)
RDW	1002	43.63 ± 3.11	43.41 ± 9.87	491 (49.00%)	511 (51.00%)

Abbreviations: WBC, white blood cell count; LYMPH%, lymphocytes percentage; MONO%, mononuclear cells percentage; NEUT%, neutrophil percentage; EO%, eosinophil percentage; BASO%, basophils percentage; MONO, absolute monocyte count; NEUT, absolute neutrophil count; EO, eosinophils count; BASO, absolute value of basophils; PLT, platelet count; PDW, platelet distribution width; MPV, mean platelet volume; RBC, red blood cell count; HGB, hemoglobin; MCV, mean corpuscular volume; MCH, mean corpuscular hemoglobin; MCHC, mean corpuscular hemoglobin concentration; RDW%, red cell distribution width; RDW, red blood cell distribution width; N, number.

Table 2 Genetic Loci Significantly Associated with 10 Phenotypic Indicators Related to White Blood Cell Identified by the GWAS After Imputation Analysis

Indicator	Gene	SNP ID	Chr	Position	Alleles	MA	MAF	Func	β	SE	*p*
WBC	SYT16	rs10132604	14	62094524	T /A	T	0.016	Intronic	0.87	0.20	1.67E-06
	INTS9	rs147409350	8	28828571	A /G	A	0.331	Intronic	0.22	0.05	3.71E-06
	LINC02101, PLK2	**rs2964173**	**5**	**58210943**	**C /T**	**C**	**0.013**	**Intergenic**	**0.91**	**0.20**	**1.33E-07**
	LINC01927, LINC01879	rs66640713	18	76679809	A /G	A	0.303	Intergenic	0.17	0.04	4.53E-06
	SAMD5, SASH1	rs75814375	6	148253461	T /A	T	0.135	Intergenic	0.29	0.06	3.63E-06
LYMPH%	LRP2	rs141807916	2	169338731	G /A	G	0.002	Intronic	2.39	0.49	1.48E-06
	RYR1	rs75146254	19	38505613	A /G	A	0.060	Intronic	0.45	0.10	4.85E-06
	LINC01734, LINC01370	rs149771501	20	39907448	C /A	C	0.002	Intergenic	2.29	0.49	3.98E-06
MONO%	ADGRL4, LINC01781	rs35658585	1	79582164	A /G	A	0.293	Intergenic	0.23	0.05	3.74E-06
	OLIG3, LOC102723649	rs17066693	6	137664353	T /C	T	0.184	Intergenic	0.26	0.06	3.57E-06
	PARP11, CCND2-AS1	rs12578463	12	3901098	C /T	C	0.202	Intergenic	0.26	0.06	4.26E-06
	LINC01441, CBLN4	rs79157887	20	55613955	T /C	T	0.011	Intergenic	1.02	0.21	1.14E-06
NEUT%	LRP2	rs141807916	2	169338731	G /A	G	0.002	Intronic	2.44	0.49	8.51E-07
	GNAO1	rs1478465	16	56260215	G /A	G	0.383	Intronic	0.21	0.05	3.55E-06
EO%	UBAC2, LINC01232	rs201566719	13	99431696	T /C	T	0.003	Intergenic	1.95	0.39	6.95E-07
BASO%	DAB1	rs77405092	1	57554207	A /G	A	0.064	Intronic	0.42	0.09	2.57E-06
	MR1	rs6667291	1	181042373	A /G	A	0.377	Intronic	0.21	0.05	4.27E-06
	SNTG2	rs60794971	2	1194997	A /G	A	0.181	Intronic	0.27	0.06	2.78E-06
	WDR41, OTP	rs4518368	5	77594997	A /G	A	0.295	Intergenic	0.22	0.05	3.47E-06
	TRIM39-RPP21, HLA-E	rs78139585	6	30422473	C /G	C	0.018	Intergenic	0.77	0.17	4.73E-06
	LOC349160	rs1023338786	7	137153033	A /T	A	0.026	ncRNA_intronic	0.64	0.14	3.69E-06
	CDH8, NONE	rs11860730	16	62275351	A /G	A	0.255	Intergenic	0.23	0.05	4.12E-06
MONO	DAB1	rs1949831	1	57534661	C /T	C	0.252	Intronic	0.24	0.05	5.15E-07
	STXBP6	rs7141943	14	24969044	G /A	G	0.456	Intronic	0.20	0.04	1.18E-06
NEUT	RBM47	rs113149044	4	40471619	A /G	A	0.019	Intronic	0.76	0.17	4.73E-06
	UNC5C	**rs78843681**	**4**	**95474708**	**A /G**	**A**	**0.029**	**Intronic**	**0.69**	**0.13**	**6.42E-08**
	JAK2	rs3780373	9	5098223	C /T	C	0.210	Intronic	0.26	0.05	1.17E-06
	LOC105376360	rs72765379	10	3535801	A /C	A	0.012	ncRNA_intronic	0.91	0.20	4.78E-06
	EMSY, LRRC32	rs17135049	11	76628603	T /C	T	0.018	Intergenic	0.74	0.16	3.37E-06
	XYLT1, NPIPA7	**rs7195345**	**16**	**17529946**	**G /C**	**G**	**0.046**	**Intergenic**	**0.54**	**0.11**	**3.58E-07**
EO	LINC01210, CLDN18	rs9871499	3	137809932	A /G	A	0.395	Intergenic	0.20	0.04	4.55E-06
	CCDC170	rs200727762	6	151572670	T /G	T	0.240	Intronic	0.24	0.05	3.77E-06
	C8orf87, LINC00535	rs16915831	8	93268246	T /C	T	0.301	Intergenic	0.23	0.05	1.60E-06
	UBAC2, LINC01232	rs201566719	13	99431696	T /C	T	0.003	Intergenic	1.78	0.38	4.05E-06
BASO	DAB1	rs77405092	1	57554207	A /G	A	0.065	Intronic	0.42	0.09	7.13E-07
	ARPP21, STAC	rs17240824	3	36117951	A /G	A	0.071	Intergenic	0.40	0.08	1.95E-06
	MIR4300HG	rs76522348	11	81977012	C /T	C	0.033	ncRNA_intronic	0.59	0.12	1.73E-06
	EMSY, LRRC32	**rs144735144**	**11**	**76597718**	**A /G**	**A**	**0.017**	**Intergenic**	**0.89**	**0.17**	**1.12E-07**
	LINC01029, SALL3	rs7244606	18	78624064	A /G	A	0.447	Intergenic	0.19	0.04	2.99E-06

Notes: *p* value < 5E-08 indicates that the candidate SNVs have genome-wide significance; *p* value < 5E-06 indicates that the candidate SNVs have suggestive genome-wide significance. Text bold represent SNVs with *p* value < 5E-07, which will be selected as the Top SNVs for replication test.

Abbreviations: WBC, white blood cell count; LYMPH%, lymphocytes percentage; MONO%, mononuclear cells percentage; NEUT%, neutrophil percentage; EO%, eosinophil percentage; BASO%, basophils percentage; MONO, absolute monocyte count; NEUT, absolute neutrophil count; EO, eosinophils count; BASO, absolute value of basophils; N, number.

GWAS Appraisal results

The GWAS results showed that a total of 90 gene loci were associated with the level of hematological phenotypic indicators investigated in this study (Tables 2–5). The association between 89 SNVs and hematological phenotypic indicator levels reached suggestive genome-wide significance (p value < 5E-06). Only the association between CCDC157-rs35289401 and the level of RDW was significant genome-wide (p value = 4.21E-08).

In this study, hematology characteristics were divided into four categories for analysis, including WBC, PLT, RBC and hemoglobin. We constructed Manhattan plots (Figures 1–4), in which the red line represents the suggestive cutoff value for genome-wide significance (5E-06). Quantile–quantile plots are shown in Figures 5–8 the x-coordinate represents the expected p value, and the y-coordinate represents the actual p value. The QQ plots showed that the distribution of p values for the association test had no systemic bias.

WBC

The GWAS results (Table 2) showed that a total of 39 gene loci were significantly associated with phenotypic indicators related to white blood cells (WBC, LYMPH%, MONO%, NEUT%, EO%, BASO%, MONO, NEUT, EO, and BASO). The results suggested that these significant associations may have genome-wide significance (p < 5E-06). Four SNVs were selected as top SNVs for subsequent replication verification: LINC02101-PLK2 rs2964173 (WBC count), UNC5C rs78843681 (NEUT), XYLT1-NPIPA7 rs7195345 (NEUT), and EMSY-LRRC32 rs144735144 (BASO).

PLT

The GWAS results (Table 3) showed that 5 gene loci were significantly associated with 3 phenotypic indicators related to platelets (PLT count, PDW, and MPV), which suggested that the associations may have genome-wide significance.

RBC

The GWAS results (Table 4) showed that there was a significant association between gene loci and RDW that reached genome-wide significance. This gene locus was CCDC157-rs35289401 (p = 4.21E-08). In addition, the significant association between 27 SNVs and 4 phenotypic indicators related to RBCs (RBC count, RDW%, RDW, and MCV) were suggested to have genome-wide significance. Finally, 5 gene loci were selected as top SNVs for replication verification: STAP1 rs191799779 (RBC), C15orf53-C15orf54 rs2912390 (RBC), LOC100506474-LINC00276 rs118103202 (RDW%), LINC00578 rs1875098 (RDW), and CCDC157 W rs35289401 (RDW).

Hemoglobin

The GWAS results (Table 5) showed that there may be a potentially significant genome-wide association between 18 SNVs and 3 phenotypic indicators related to hemoglobin (HGB, MCH, and MCHC). Four gene loci were selected as top SNVs for replication verification: ARHGAP25 rs10208669 (MCH), NRIP1-USP25 rs12482879 (MCH), HBS1 L-MYB rs1331309 (MCH), and CBLN1-C16orf78 rs148933121 (MCHC).

Table 3 Genetic Loci Significantly Associated with 3 Phenotypic Indicators Related to Platelets Identified by the GWAS After Imputation Analysis

Indicator	Gene	SNP ID	Chr	Position	Alleles	MA	MAF	Func	β	SE	*p*
PLT	FIGN, GRB14	rs10180568	2	164191260	G /A	G	0.001	Intergenic	3.05	0.69	2.56E-06
	OR2H1, MAS1L	rs116973845	6	29469134	T /C	T	0.262	Intergenic	0.16	0.04	2.89E-06
	FASTKD2, CPO	rs138214429	2	206904970	C /T	C	0.485	Intergenic	0.20	0.05	1.39E-06
PDW	CDC73	rs140402302	1	193147821	G /A	G	0.014	Intronic	0.95	0.20	2.05E-06
MPV	LINC01507, TLE1	rs1333934	9	80320182	C /A	C	0.487	Intergenic	0.01	0.00	3.39E-06

Notes: p value < 5E-08 indicates that the candidate SNVs have genome-wide significance; p value < 5E-06 indicates that the candidate SNVs have suggestive genome-wide significance.

Abbreviations: PLT, platelet count; PDW, platelet distribution width; MPV, mean platelet volume; N, number.

https://doi.org/10.2147/PGPM.S361809 DovePress

Table 4 Genetic Loci Significantly Associated with 4 Phenotypic Indicators Related to Red Blood Cells Identified by the GWAS After Imputation Analysis

Indicator	Gene	SNP ID	Chr	Position	Alleles	MA	MAF	Func	β	SE	*p*
RBC	TBX3, MED13L	rs11067555	12	115377103	T /C	T	0.022	Intergenic	0.68	0.15	2.47E-06
	AGO1	rs138026185	1	35878638	A /G	A	0.485	Intronic	0.20	0.05	3.18E-06
	RBMS3	rs191082924	3	29798074	A /G	A	0.278	Intronic	0.20	0.04	3.78E-06
	STAP1	**rs191799779**	**4**	**67608223**	**T /A**	**T**	**0.274**	**Downstream**	**0.20**	**0.04**	**2.19E-07**
	DCBLD2	rs278382	3	98808894	A /G	A	0.305	Intronic	0.16	0.04	3.59E-06
	C15orf53, C15orf54	**rs2912390**	**15**	**38803385**	**C /T**	**C**	**0.040**	**Intergenic**	**0.52**	**0.11**	**2.75E-07**
	SVEP1	rs77133716	9	110502789	T /G	T	0.302	Intronic	0.17	0.04	4.62E-06
	FMO9P, POGK	rs77634171	1	166643290	T /C	T	0.192	Intergenic	0.27	0.06	3.92E-06
	LINC02429, MIR548AG1	rs78117683	4	59074899	G /A	G	0.029	Intergenic	0.58	0.12	2.27E-06
	FREM2, STOML3	rs9532301	13	38902978	G /C	G	0.299	Intergenic	0.18	0.04	2.35E-06
RDW%	LOC100506474, LINC00276	**rs118103202**	**2**	**13902329**	**A /T**	**A**	**0.158**	**Intergenic**	**0.33**	**0.06**	**3.49E-07**
	EPHA3, NONE	rs6551413	3	89591200	G /C	G	0.189	Intergenic	0.27	0.06	2.35E-06
	NECTIN3, CD96	rs4682281	3	111513738	C /A	C	0.050	Intergenic	0.48	0.10	1.96E-06
	TLE1	rs78796872	9	81639243	G /T	G	0.038	Intronic	0.58	0.12	2.89E-06
	MALRD1	rs736242	10	19684333	G /C	G	0.027	Intronic	0.66	0.14	4.80E-06
	LLGL2	rs1661723	17	75566210	T /C	T	0.123	Intronic	0.33	0.07	1.14E-06
RDW	YTHDF2, OPRD1	rs10799121	1	28803420	C /T	C	0.100	Intergenic	0.38	0.07	7.78E-07
	FOXJ3	rs141684413	1	42299978	A /G	A	0.001	Intronic	3.24	0.68	2.25E-06
	LINC01677, LINC01661	rs10458515	1	106645258	T /A	T	0.493	Intergenic	0.20	0.04	3.03E-06
	LINC00578	**rs1875098**	**3**	**177596880**	**A /G**	**A**	**0.002**	**ncRNA_intronic**	**2.50**	**0.48**	**2.36E-07**
	FER1L6-AS2	rs6859250	5	124068266	G /A	G	0.022	ncRNA_intronic	0.70	0.15	4.10E-06
	LINC00703, LINC00704	rs7074882	10	4505038	G /T	G	0.038	Intergenic	0.59	0.12	5.99E-07
	CCDC157	**rs35289401**	**22**	**30368637**	**T /C**	**T**	**0.121**	**Intronic**	**0.37**	**0.07**	**4.21E-08**
MCV	TOX	rs10504266	8	58865140	A /C	A	0.362	Intronic	0.18	0.04	1.65E-06
	NRIP1, USP25	rs12482879	21	15085300	C /G	C	0.244	Intergenic	0.20	0.04	1.71E-06
	LINC02237, CSMD3	rs13261386	8	111579539	C /T	C	0.279	Intergenic	0.16	0.04	3.68E-06
	RASA3, CDC16	rs1810797	13	114163337	G /A	G	0.014	Intergenic	0.81	0.18	7.20E-07
	NRXN1	rs34533417	2	49923506	G /A	G	0.194	Intronic	0.26	0.06	2.92E-06

Notes: *p* value < 5E-08 indicates that the candidate SNVs have genome-wide significance; *p* value < 5E-06 indicates that the candidate SNVs have suggestive genome-wide significance. Text bold represent SNVs with *p* value < 5E-07, which will be selected as the top SNVs for replication test.

Abbreviations: RBC, red blood cell count; MCV, mean corpuscular volume; RDW%, red cell distribution width%; RDW, red blood cell distribution width; N, number.

Table 5 Genetic Loci Significantly Associated with 3 Phenotypic Indicators Related to Hemoglobin Identified by the GWAS After Imputation Analysis

Indicator	Gene	SNP ID	Chr	Position	Alleles	MA	MAF	Func	β	SE	*p*
HGB	LINC01036	rs112550911	1	187162697	G /A	G	0.170	ncRNA_intronic	0.04	0.05	4.09E-06
	AJAP1/MIR4417	rs12119802	1	5547928	C /T	C	0.243	Intergenic	0.20	0.04	1.88E-06
	PTCH1/LINC00476	rs184266331	9	95620832	A /C	A	0.014	Intergenic	0.81	0.18	4.75E-06
	HPGDS/PDLIM5	rs35642552	4	94397195	T /A	T	0.248	Intergenic	0.24	0.05	9.77E-07
MCH	ARHGAP25	**rs10208669**	**2**	**68791655**	**A /G**	**A**	**0.008**	**Intronic**	**1.13**	**0.25**	**3.13E-07**
	C16orf72/LINC02177	rs112628002	16	9183925	T /G	T	0.073	Intergenic	0.38	0.08	3.38E-06
	LRFN2	rs115628240	6	40542274	A /G	A	0.299	Intronic	0.16	0.04	1.14E-06
	NRIP1/USP25	**rs12482879**	**21**	**16558085**	**C /G**	**C**	**0.243**	**Intergenic**	**0.20**	**0.04**	**1.18E-07**
	HBS1L/MYB	**rs1331309**	**6**	**135085040**	**G /T**	**G**	**0.406**	**Intergenic**	**0.15**	**0.03**	**1.10E-07**
	LINC00934	rs2555300	12	119297157	A /G	A	0.091	ncRNA_intronic	0.35	0.08	2.83E-06
	GSAP	rs35429354	7	77321611	A /T	A	0.258	Intronic	0.17	0.04	4.64E-06
	PTPRN2	rs7455229	7	158039751	G /A	G	0.008	Intronic	0.92	0.19	1.21E-06
MCHC	LRP1B	rs148177700	2	141623782	T /C	T	0.442	Intronic	0.21	0.05	6.72E-07
	CBLN1/C16orf78	**rs148933121**	**16**	**49370122**	**C /G**	**C**	**0.484**	**Intergenic**	**0.20**	**0.05**	**4.55E-07**
	SOX5	rs16927632	12	24440888	A /G	A	0.484	Intronic	0.20	0.05	2.07E-06
	KCNS3/RDH14	rs2881044	2	18041124	T /C	T	0.300	Intergenic	0.16	0.04	4.22E-06
	LINC00929/GABRB3	rs74006238	15	26417707	C /T	C	0.469	Intergenic	0.22	0.05	5.01E-06
	LINC02202	rs7448500	5	159102484	G /C	G	0.017	ncRNA_intronic	0.79	0.17	4.25E-06

Notes: *p* value < 5E-08 indicates that the candidate SNVs have genome-wide significance; *p* value < 5E-06 indicates that the candidate SNVs have suggestive genome-wide significance. Text bold represent SNVs with *p* value < 5E-07, which will be selected as the Top SNVs for replication test.

Abbreviations: HGB, hemoglobin; MCH, mean corpuscular hemoglobin; MCHC, mean corpuscular hemoglobin concentration; N, number.

https://doi.org/10.2147/PGPM.S361809

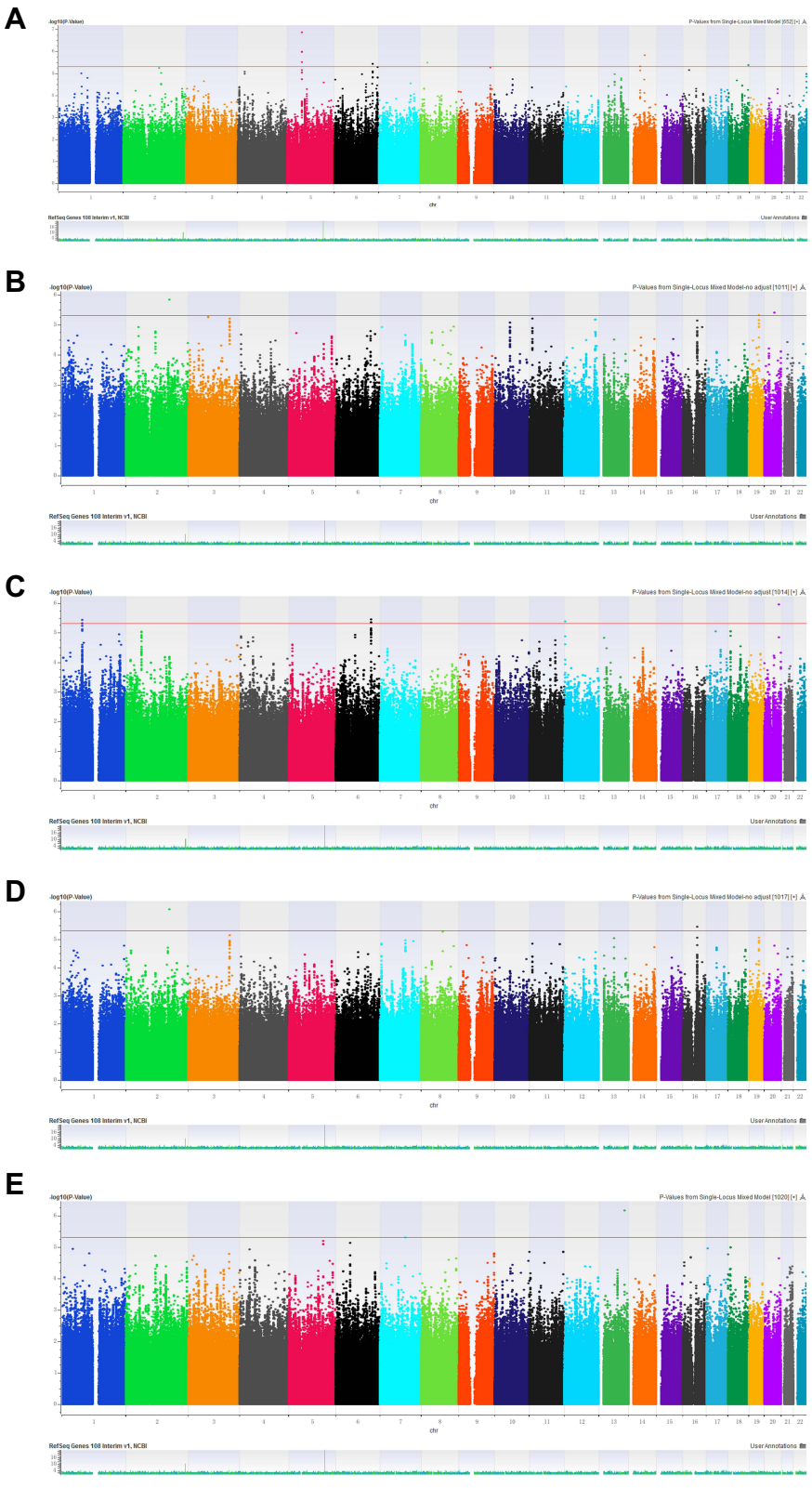

Figure 1 Continued.

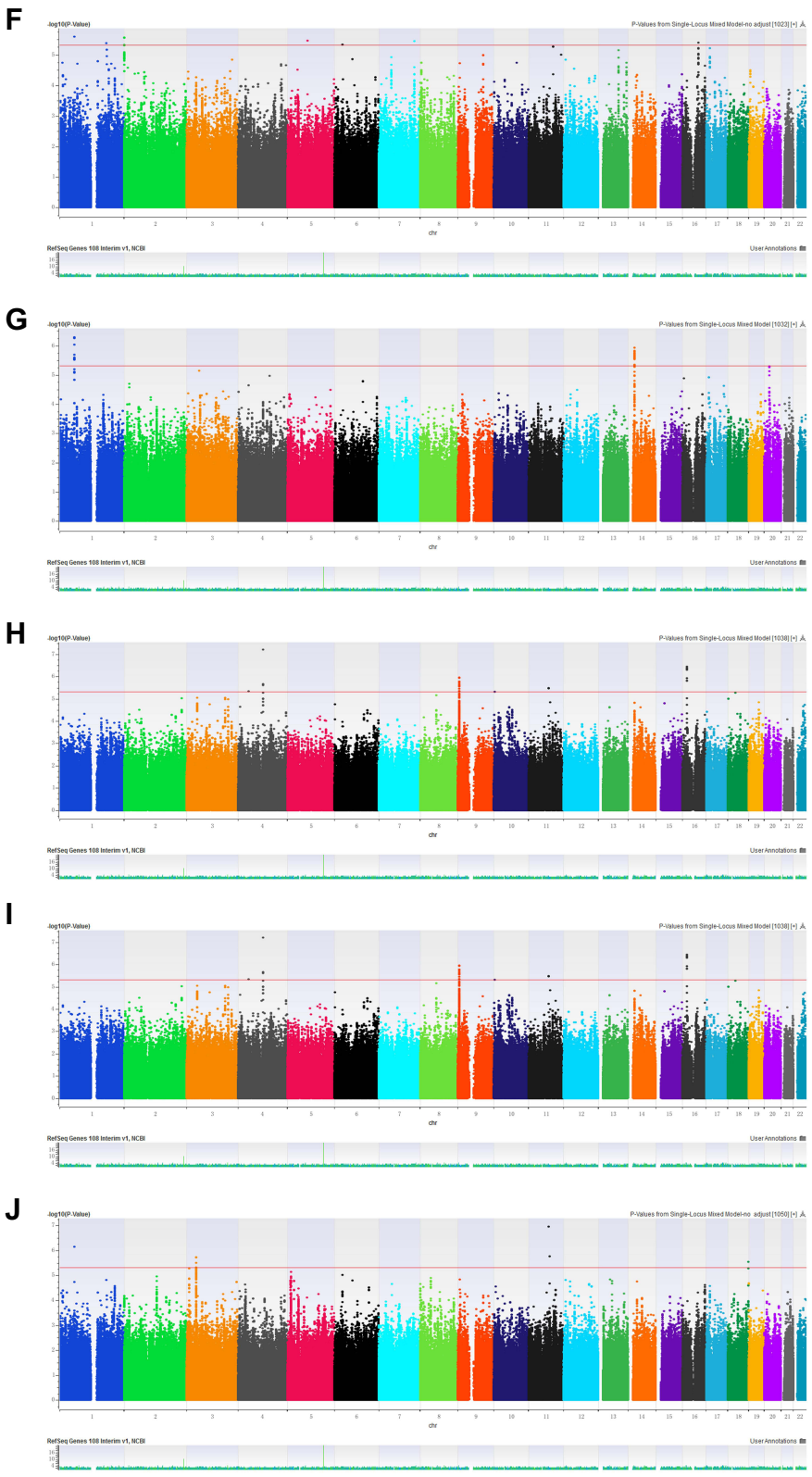

Figure 1 Manhattan plot of the results of the genome-wide association study (10 phenotypic indicators related to white blood cells). The phenotypic indicators from top to bottom in the figure are as follows: (**A**) WBC, (**B**) LYMPH%, (**C**) MONO%, (**D**) NEUT%, (**E**) EO%, (**F**) BASO%, (**G**) MONO, (**H**) NEUT, (**I**) EO, and (**J**) BASO. The x-axis represents chromosomes, whereas the y-axis represents the $-\log_{10}$ of the *p* value. The red line represents the suggested cutoff value for genome-wide significance (5.0×10^{-6}).

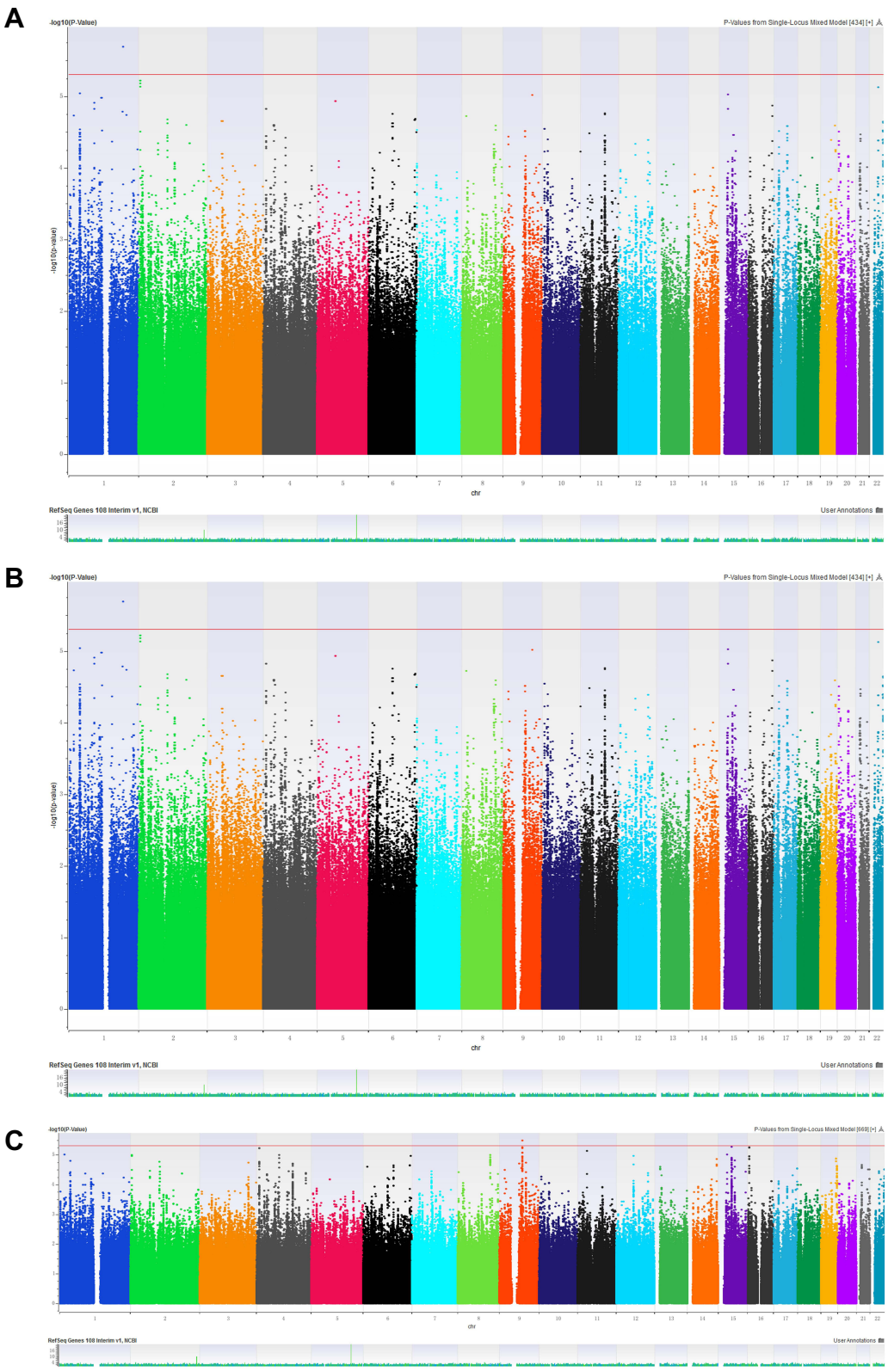

Figure 2 Manhattan plot of the results of the genome-wide association study (3 phenotypic indicators related to platelets). The phenotypic indicators from top to bottom in the figure are as follows: (**A**) PLT, (**B**) PDW, and (**C**) MPV. The x-axis represents chromosomes, whereas the y-axis represents the $-\log_{10}$ of the *p* value. The red line represents the suggested cutoff value for genome-wide significance (5.0×10^{-6}).

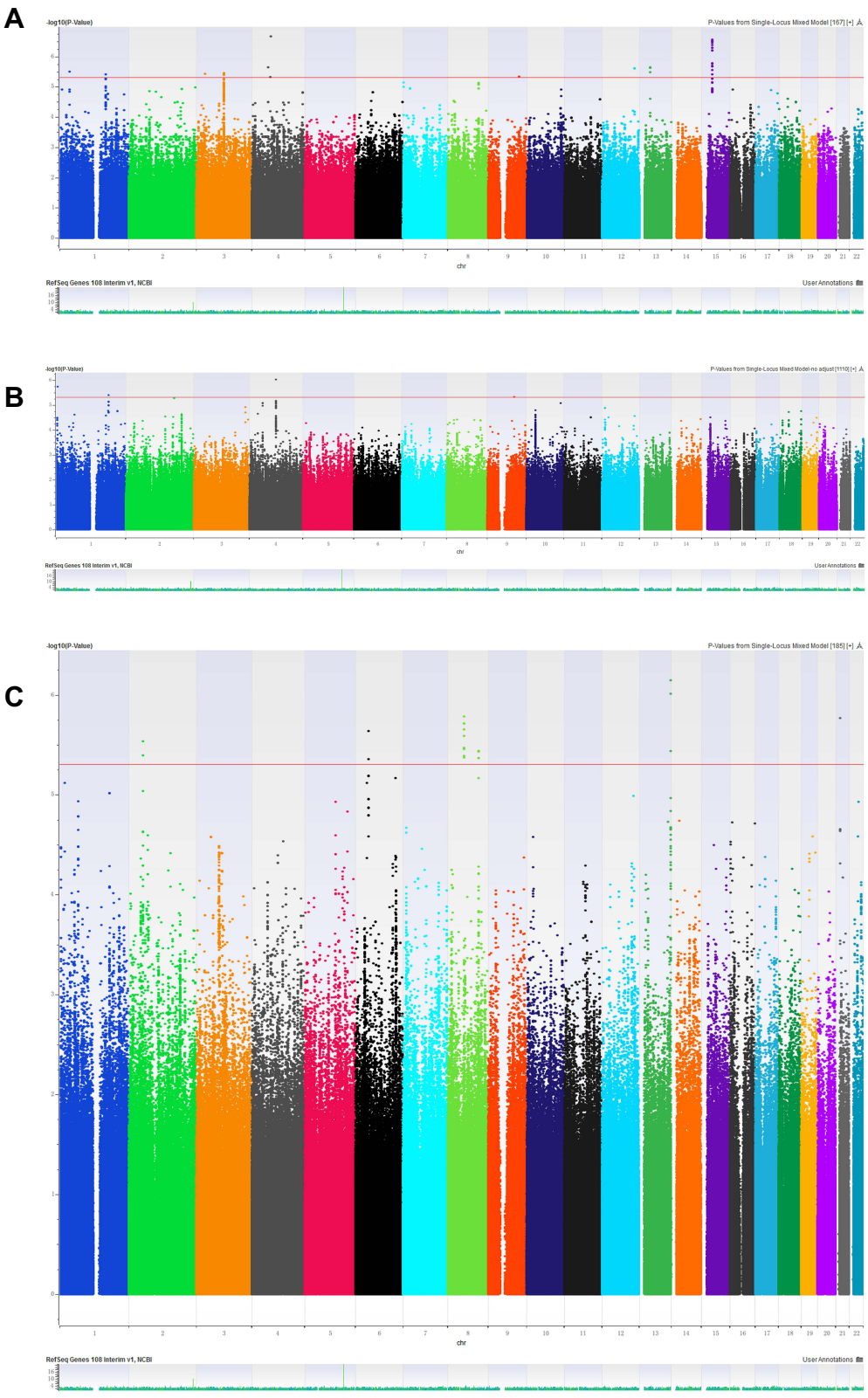

Figure 3 Continued.

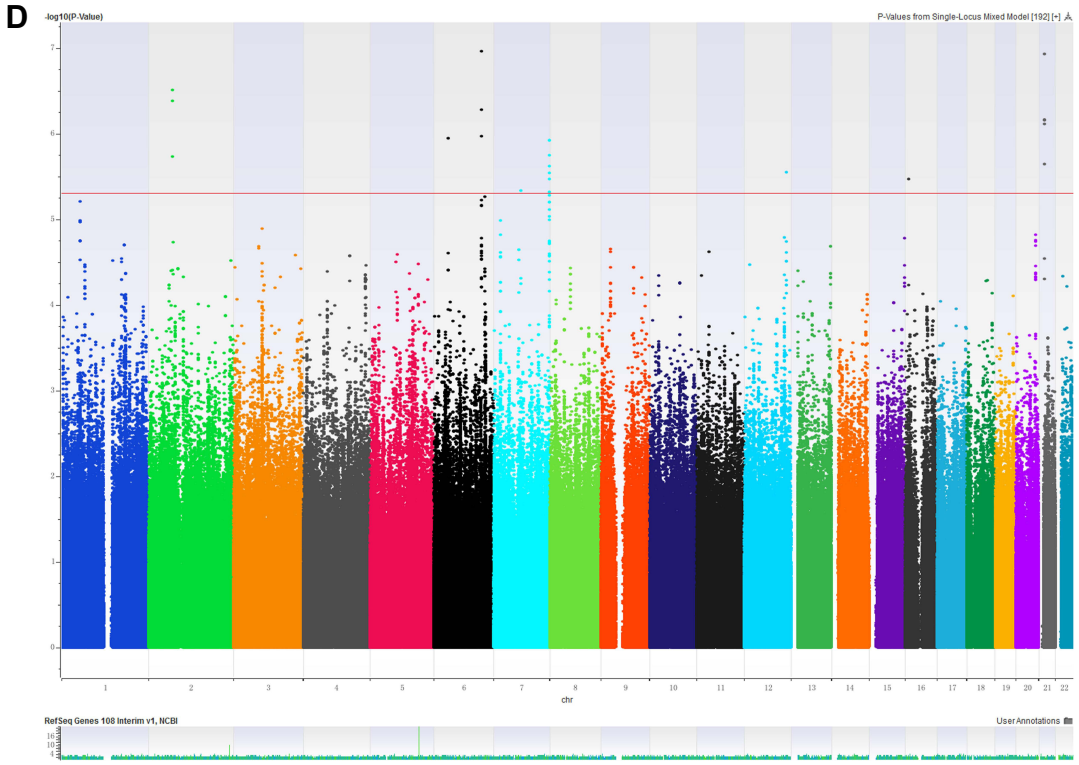

Figure 3 Manhattan plot of the results of the genome-wide association study (4 phenotypic indicators related to red blood cells). The phenotypic indicators from top to bottom in the figure are as follows: (**A**) RBC, (**B**) RDW%, (**C**) RDW, and (**D**) MCV. The x-axis represents chromosomes, whereas the y-axis represents the $-\log_{10}$ of the *p* value. The red line represents the suggested cutoff value for genome-wide significance (5.0×10^{-6}).

Replication Verification

In this study, a total of 13 genetic variants were selected as top SNVs for replication verification. Supplemental Figures 1–3 show a map of the regions associated with each hematological phenotypic indicator on different chromosomes. The replication verification results showed (Table 6) that only HBS1 L-MYB rs1331309 (p = 6.42E-07) was still significantly associated with the level of MCH in participants different from the GWAS subjects. In addition, the results (Table 7) showed that the level of MCH under genotype GG/GT of HBS1 L-MYB rs1331309 was significantly higher than that under genotype TT. Figure 9 shows the distribution box plot of MCH under different genotypes of rs1331309.

Discussion

Hematological characteristics are very important for the diagnosis of health status and diseases. Studies have reported that hematological characteristics have a degree of heritability, and genetic factors play a very important role in the variation in hematological characteristics among individuals.[1] Although some identified genetic variants associated with hematological phenotypic indicators can be shared among different ethnic groups, a large number of studies have confirmed the existence of racial differences.[8–10,14,15] To date, there are relatively few GWASs on the hematological characteristics of the Han population from northwest China, and the genetic structure of these hematological characteristics is still unclear. Our study conducted a more comprehensive genome-wide association study of hematological characteristics in the Han population from northwest China, rather than focusing on phenotypic indicators.

A total of 90 genetic variants were identified that were significantly associated with hematological phenotypic indicators. We found that the significant association between CCDC157 rs35289401 and RDW reached genome-wide significance. Our results suggest that the remaining 89 genes may have genome-wide significance. In addition, the results

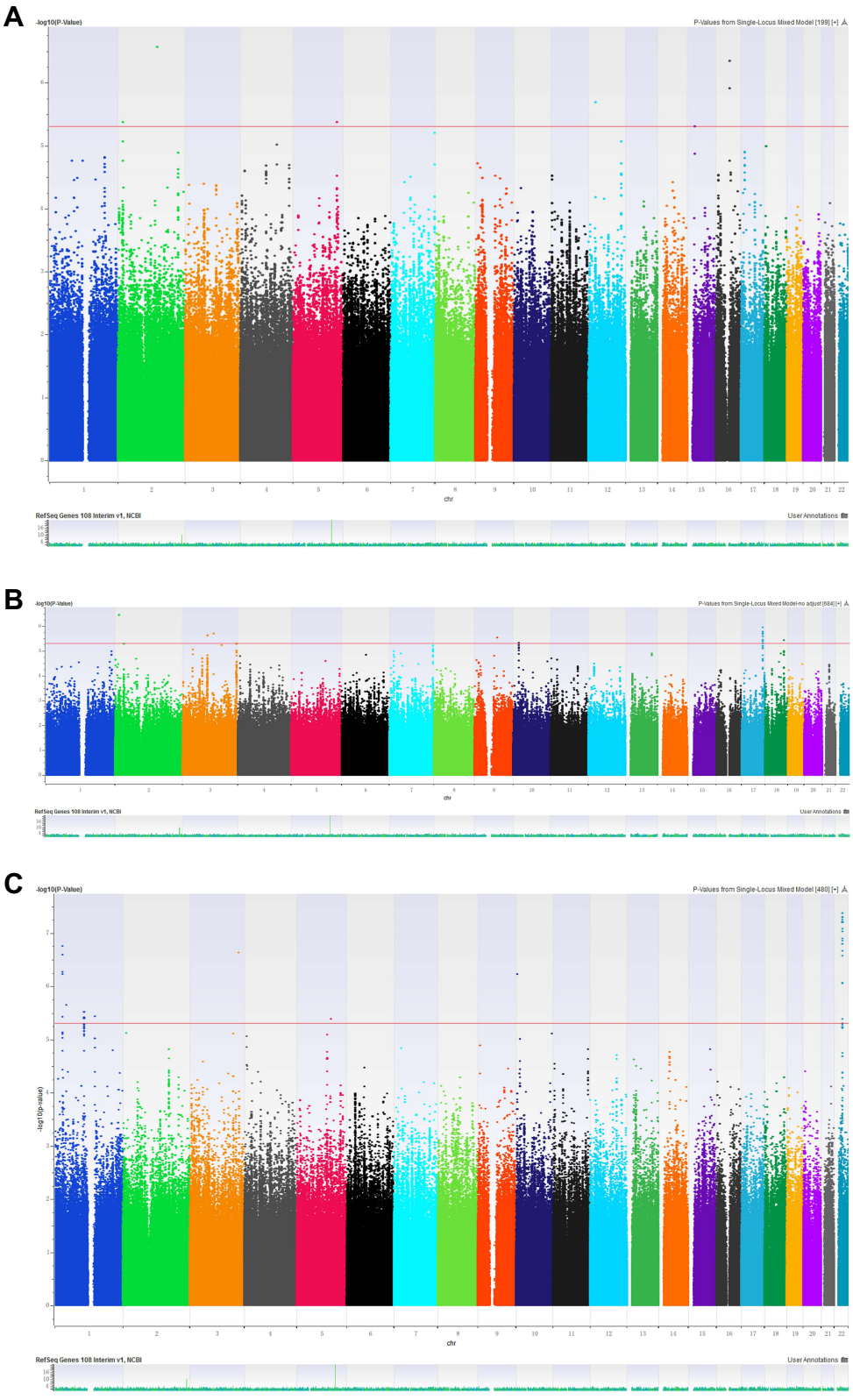

Figure 4 Manhattan plot of the results of the genome-wide association study (3 phenotypic indicators related to hemoglobin). The phenotypic indicators from top to bottom in the figure are as follows: (**A**) HGB, (**B**) MCH, and (**C**) MCHC. The x-axis represents chromosomes, whereas the y-axis represents the $-\log_{10}$ of the *p* value. The red line represents the suggested cutoff value for genome-wide significance (5.0×10^{-6}).

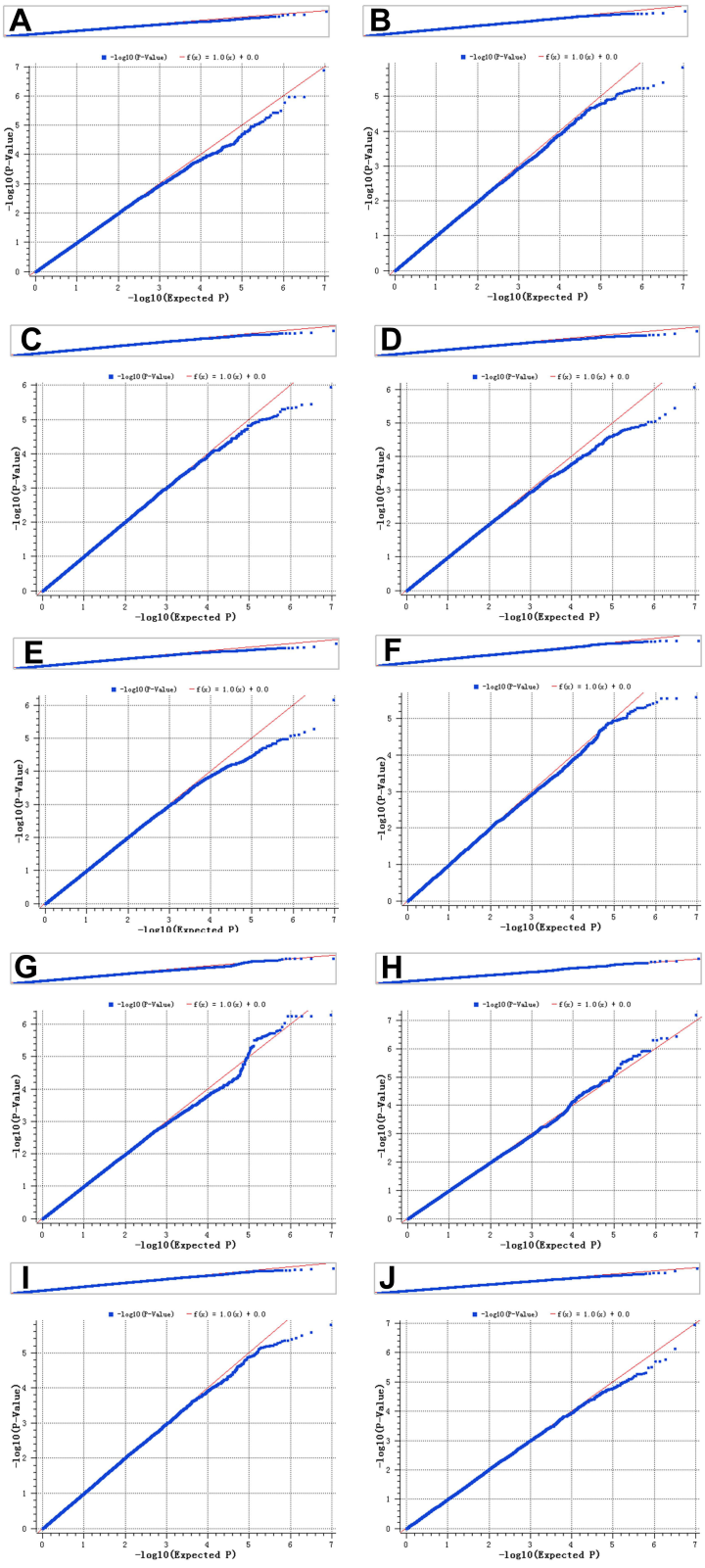

Figure 5 Quantile–quantile plots of the results of the GWAS (10 phenotypic indicators related to white blood cells). (**A**) WBC, (**B**) LYMPH%, (**C**) MONO%, (**D**) NEUT%, (**E**) EO%, (**F**) BASO%, (**G**) MONO, (**H**) NEUT, (**I**) EO, and (**J**) BASO. The x-coordinate represents the expected p value, and the y-coordinate represents the actual p value.

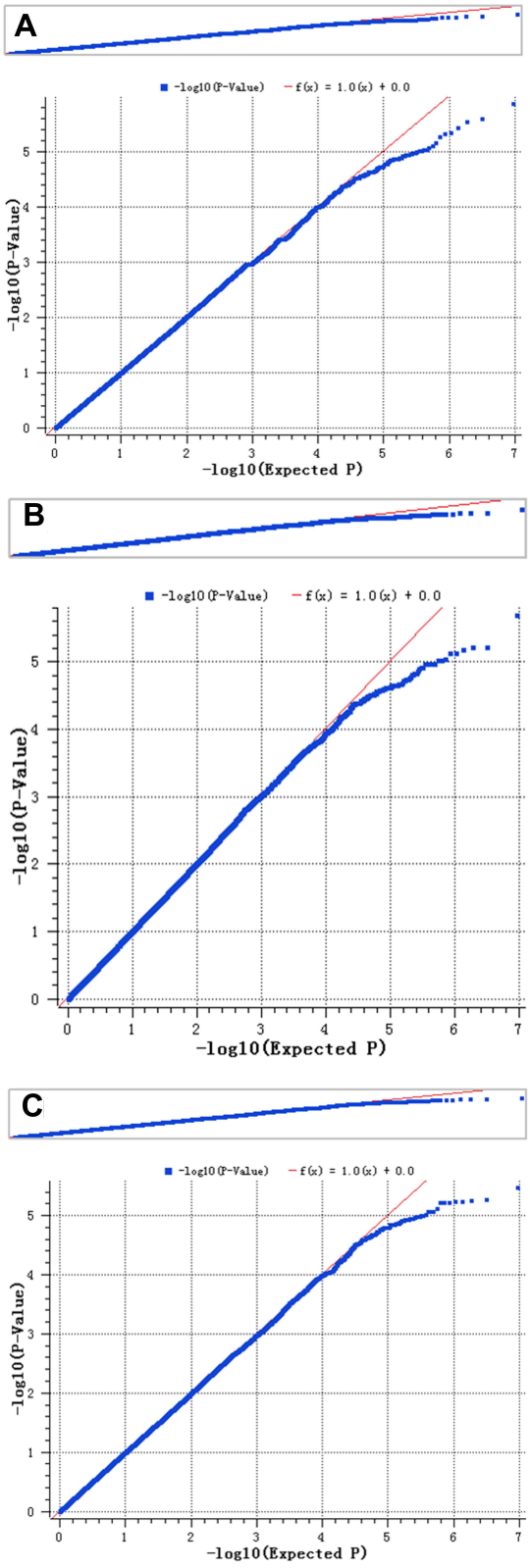

Figure 6 Quantile–quantile plots of the results of the GWAS (3 phenotypic indicators related to platelets). (**A**) PLT, (**B**) PDW, and (**C**) MPV. The x-coordinate represents the expected p value, and the y-coordinate represents the actual p value.

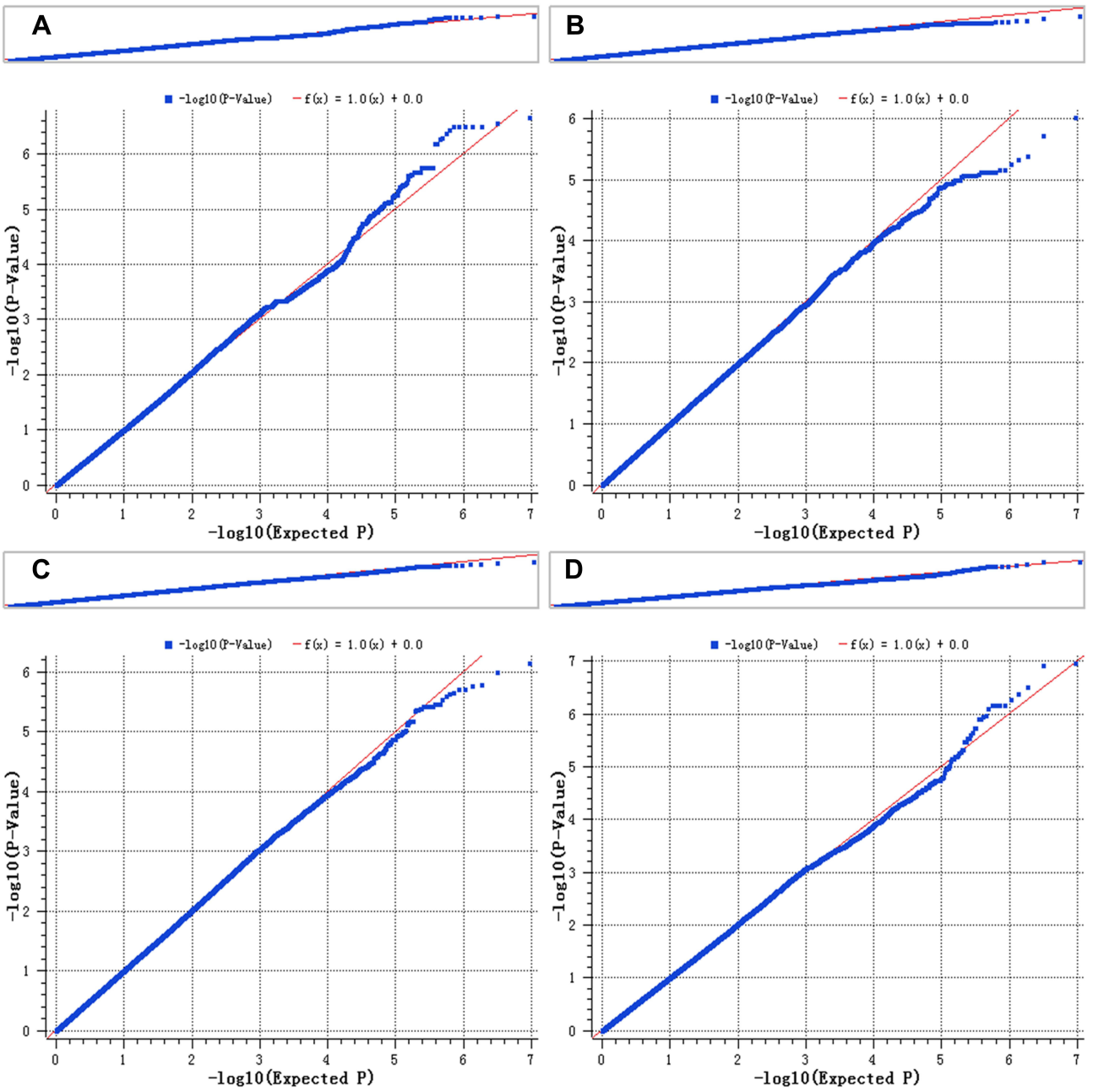

Figure 7 Quantile–quantile plots of the results of the GWAS (4 phenotypic indicators related to red blood cells). (**A**) RBC, (**B**) RDW%, (**C**) RDW, and (**D**) MCV. The x-coordinate represents the expected p value, and the y-coordinate represents the actual p value.

of the replication test showed that HBS1 L-MYB rs1331309 was still significantly associated with MCH in participants who differed from the GWAS subjects.

RDW can be used to determine the degree of red blood cell heterogeneity and is widely used in the clinical diagnosis of blood system diseases or anemia.[16,17] Red blood cell distribution width is a common phenotypic indicator in clinical practice. In recent years, a number of studies have found that elevated RDW levels can help to clinically diagnose a variety of diseases and predict the severity of diseases such as cardiovascular disease,[18] ischemic stroke, carotid atherosclerosis,[19] and hepatitis B virus-related liver disease.[20] We found that there is a significant genome-wide correlation between CCDC157 rs35289401 and RDW. To our knowledge, we are the first to report a significant genome-wide association between CCDC157 rs35289401 and RDW. However, it is worth

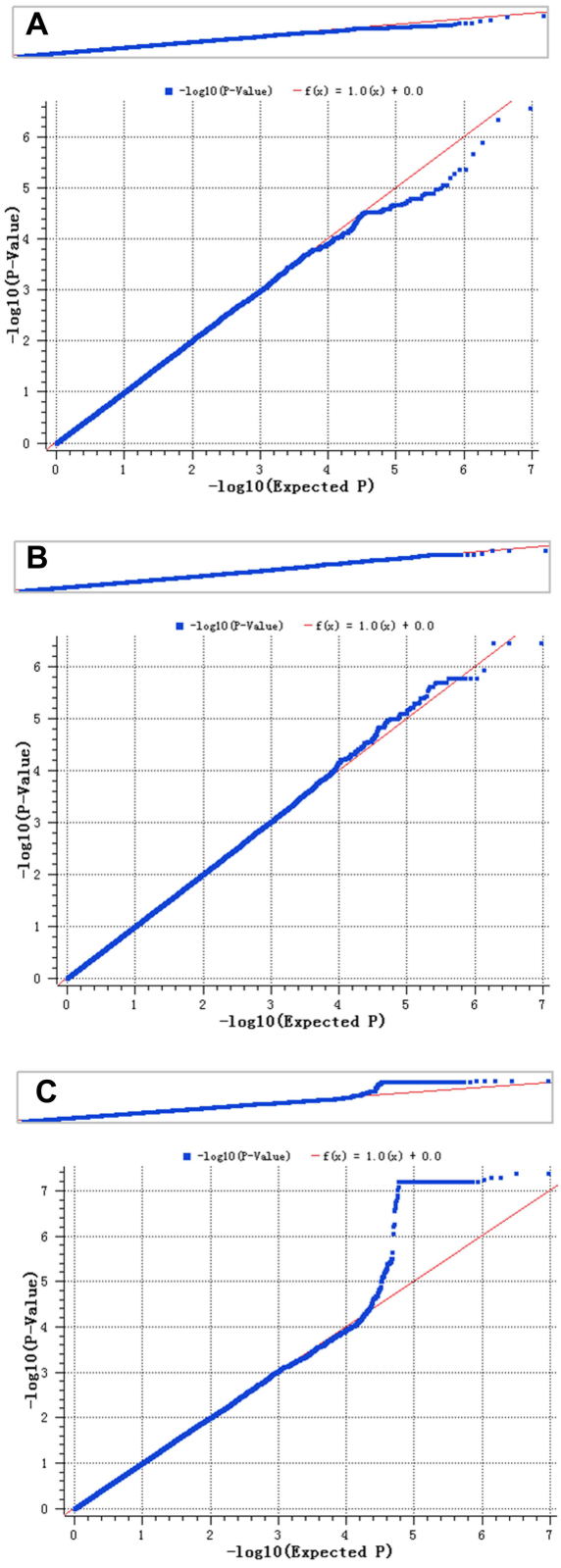

Figure 8 Quantile–quantile plots of the results of the GWAS (3 phenotypic indicators related to hemoglobin). (**A**) HGB, (**B**) MCH, and (**C**) MCHC. The x-coordinate represents the expected p value, and the y-coordinate represents the actual p value.

https://doi.org/10.2147/PGPM.S361809 DovePress

Table 6 According to the Results of GWAS SNVs with *p*-value < 5E-07 Were Selected for Replication Testing in Another Population

Indicator	Indicator	Gene	SNP ID	Alleles	MA	MAF	Func	β	SE	*p* value of GWAS	*p*
Hemocyte	WBC	LINC02101, PLK2	rs2964173	T /C	T	0.045	Intergenic	−0.005	0.072	1.33E-07	0.941
	NEUT	UNC5C	rs78843681	A /G	A	0.021	Intronic	−0.171	0.106	6.42E-08	0.105
		XYLT1/NPIPA7	rs7195345	G /C	G	0.085	Intergenic	0.014	0.056	3.58E-07	0.801
	BASO	EMSY/LRRC32	rs144735144	A /G	A	0.041	Intergenic	0.0001	0.077	1.12E-07	0.997
Red blood cells	RBC	STAP1	rs191799779	T /A	T	0.013	Downstream	−0.041	0.113	2.19E-07	0.715
		C15orf53/C15orf54	rs2912390	T /C	T	0.321	Intergenic	0.006	0.037	2.75E-07	0.871
	RDW%	LOC100506474, LINC00276	rs118103202	A /T	A	0.208	Intergenic	−0.018	0.040	3.49E-07	0.656
	RDW	LINC00578	rs1875098	A /G	A	0.023	ncRNA_intronic	0.005	0.105	2.36E-07	0.963
		CCDC157	rs35289401	T /C	T	0.126	Intronic	0.007	0.047	4.21E-08	0.878
Hemoglobin	MCH	ARHGAP25	rs10208669	A /G	A	0.362	Intronic	−0.008	0.046	3.13E-07	0.860
		NRIP1/USP25	rs12482879	C /G	C	0.376	Intergenic	0.024	0.045	1.18E-07	0.598
		HBS1L/MYB	rs1331309	G /T	G	0.328	Intergenic	0.231	0.046	1.10E-07	**6.42E-07***
	MCHC	CBLN1/C16orf78	rs148933121	C /G	C	0.045	Intergenic	0.016	0.057	4.55E-07	0.785

Notes: *p* value < 0.05 indicates statistical significance; "*" and bolding indicates statistical significance.

Abbreviations: WBC, white blood cell count; NEUT%, neutrophil percentage; EO%, eosinophil percentage; BASO%, basophils percentage; BASO, absolute value of basophils; RBC, red blood cell count; MCH, mean corpuscular hemoglobin; MCHC, mean corpuscular hemoglobin concentration; RDW%, red cell distribution width CV; RDW, red blood cell distribution width SD; N, number.

Table 7 The Level of MCH Under Different Genotypes of rs1331309

Indicator	Gene	SNP ID	Genotype	N	Mean	Std. Deviation	*p*
MCH	HBS1L/MYB	rs1331309	GG	221 (11.09%)	31.32	1.60	**0.023***
			GT	863 (43.32%)	29.74	1.56	
			TT	908 (45.58%)	28.52	1.50	
			Total	1992 (100.00%)	29.86	1.56	

Notes: *p* value < 0.05 indicates statistical significance; "*" and bolding indicates statistical significance.
Abbreviations: MCH, mean corpuscular hemoglobin; N, number.

noting that in our replication test, CCDC157 rs35289401 was not associated with RDW. We speculate that the reason for this discrepancy may be the small sample size. If we directly identify CCDC157 rs35289401 as a genetic signal of RDW among the Han population from northwest China, it appears that the evidence is insufficient. It is necessary to expand the sample size and perform multiple verification tests. Regardless, our findings provide a valuable reference for a follow-up study of the genetic structure of RDW in the Han population from northwest China. It will also provide supplemental data for the GWAS of hematological characteristics.

In addition, the significant association between the HBS1 L-MYB intergenic region rs1331309 and MCH passed the replication verification. The level of MCH in the rs1331309 GG/GT genotype was significantly higher than that in the wild genotype TT. The results of our study are similar to those of previous studies. A genome-wide sequencing and interpolation study conducted by Southam et al in isolated populations from the island of Crete identified a significant association between rs1331309 and MCH levels (*p* value = 2.00E-9).[21] In addition, Chen et al found a significant association between rs1331309 and white blood cell count in a cross-racial GWAS on blood cells (*p* value = 1.00E-34).[22] The above evidence suggests rs1331309 is expected to become a kind of monitoring the level of MCH biomarkers. Studies have reported that low levels of MCH are important phenotypic indicators that can be used to diagnose many diseases.[23,24] Combined with the results of our study, we speculate that rs1331309 genetic variants can be used to monitor changes in MCH levels and facilitate diagnosis or predict the occurrence and development of many diseases. Our study lays the foundation for the study of the genetic structure of hematological characteristics in the Han population from northwest China. It also provides a valuable reference for the clinical diagnosis or prediction of a variety of diseases.

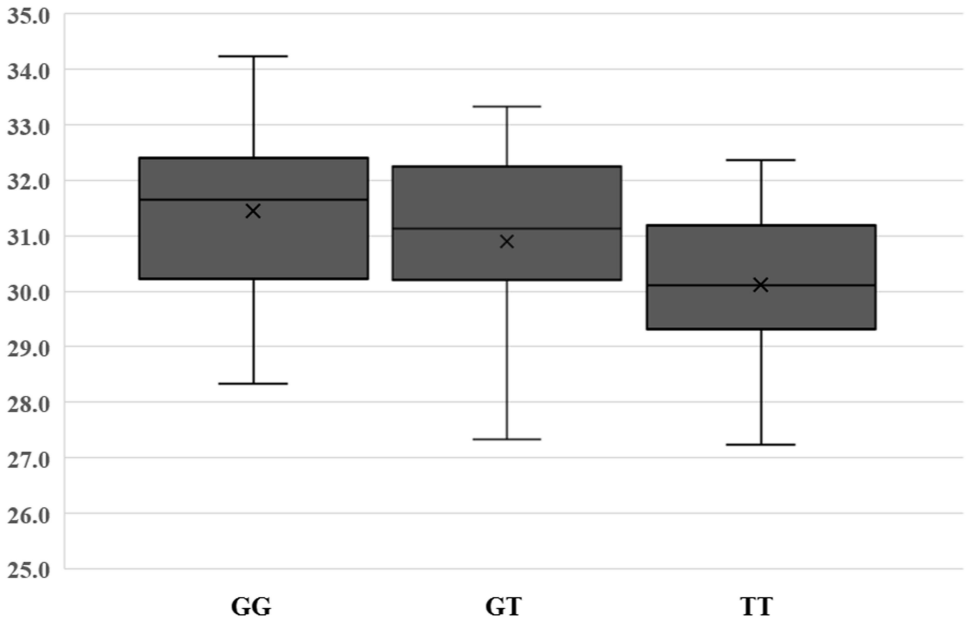

Figure 9 Box plot of MCH levels under different genotypes of HBS1L-MYB rs1331309.

Our study has certain limitations: the association between hematological phenotypic indicators in the Han population from northwest China and CCDC157 rs35289401/HBS1 L-MYB rs1331309 has never been reported. Therefore, further functional analysis is necessary to verify the relationship between these two genetic variants and hematological-related phenotypic indicators. In addition, a larger sample size is necessary to confirm the results and allow new discoveries.

Conclusions

In summary, 90 genetic variants were significantly associated with hematological phenotypic indicators among the Han population from northwest China. In particular, rs1331309 (HBS1 L-MYB) was significantly associated with MCH levels and passed the replication test and is expected to be a biomarker for monitoring the dynamics of MCH levels. This study provides a reference for the study of the genetic structure of hematological characteristics. It also provides a valuable reference for the clinical diagnosis/prediction of a variety of diseases.

Data Sharing Statement

The datasets used and analyzed in the current study are available from the corresponding author on reasonable request.

Ethics Approval and Consent to Participate

This study was conducted under the standard approved by the ethics committee of the Affiliated Hospital of Xizang Minzu University. The study conformed to the ethical principles for medical research involving humans of the World Medical Association Declaration of Helsinki. All participants provided written informed consent before participating in this study.

Acknowledgments

We thank all the authors for their contributions and support. In addition, we also thank Hongyan Lu, Yuliang Wang and Zhanhao Zhang for their assistance in this study.

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

This study was supported by the Key R&D Program of Xizang (Tibet) Autonomous Region (XZ202101ZY0018G), the Natural Science Foundation of Tibet Autonomous Region (XZ 2019 ZR G-42(Z)), a Talent Development Supporting Project entitled "Tibet-Shaanxi Himalaya of Xizang Minzu University" (2020 Plateau Scholar), and the Innovation and Entrepreneurship Program for College Students of Tibet Nationalities University (MD202010695093).

Disclosure

The authors declare that they have no conflicts of interest in this work.

References

1. Okada Y, Kamatani Y. Common genetic factors for hematological traits in humans. *J Hum Genet*. 2012;57(3):161–169. doi:10.1038/jhg.2012.2
2. Zhang Z, Hong Y, Gao J, et al. Genome-wide association study reveals constant and specific loci for hematological traits at three time stages in a White Duroc × Erhualian F2 resource population. *PLoS One*. 2013;8(5):e63665. doi:10.1371/journal.pone.0063665
3. Soranzo N, Spector TD, Mangino M, et al. A genome-wide meta-analysis identifies 22 loci associated with eight hematological parameters in the HaemGen consortium. *Nat Genet*. 2009;41(11):1182–1190. doi:10.1038/ng.467
4. Evans DM, Frazer IH, Martin NG. Genetic and environmental causes of variation in basal levels of blood cells. *Twin Res*. 1999;2(4):250–257. doi:10.1375/twin.2.4.250
5. Garner C, Tatu T, Reittie JE, et al. Genetic influences on F cells and other hematologic variables: a twin heritability study. *Blood*. 2000;95(1):342–346. doi:10.1182/blood.V95.1.342

6. Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet*. 2005;6(2):95–108. doi:10.1038/nrg1521
7. Kim YK, Oh JH, Kim YJ, 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. doi:10.1155/2015/914965
8. Li J, Glessner JT, Zhang H, 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(7):1457–1464. doi:10.1093/hmg/dds534
9. Lo KS, Wilson JG, Lange LA, et al. Genetic association analysis highlights new loci that modulate hematological trait variation in Caucasians and African Americans. *Hum Genet*. 2011;129(3):307–317. doi:10.1007/s00439-010-0925-1
10. Kamatani Y, Matsuda K, Okada Y, et al. Genome-wide association study of hematological and biochemical traits in a Japanese population. *Nat Genet*. 2010;42(3):210–215. doi:10.1038/ng.531
11. Seiki T, Naito M, Hishida A, 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
12. Mucha S, Mrode R, Coffey M, Kizilaslan M, Desire S, Conington J. Genome-wide association study of conformation and milk yield in mixed-breed dairy goats. *J Dairy Sci*. 2018;101(3):2213–2225. doi:10.3168/jds.2017-12919
13. Son HY, Hwangbo Y, Yoo SK, et al. Genome-wide association and expression quantitative trait loci studies identify multiple susceptibility loci for thyroid cancer. *Nat Commun*. 2017;8:15966. doi:10.1038/ncomms15966
14. Yasukochi Y, Sakuma J, Takeuchi I, et al. Identification of nine novel loci related to hematological traits in a Japanese population. *Physiol Genomics*. 2018;50(9):758–769. doi:10.1152/physiolgenomics.00088.2017
15. van Rooij FJA, Qayyum R, Smith AV, et al. Genome-wide trans-ethnic meta-analysis identifies seven genetic loci influencing erythrocyte traits and a role for RBPMS in erythropoiesis. *Am J Hum Genet*. 2017;100(1):51–63. doi:10.1016/j.ajhg.2016.11.016
16. Fava C, Cattazzo F, Hu ZD, Lippi G, Montagnana M. The role of red blood cell distribution width (RDW) in cardiovascular risk assessment: useful or hype? *Ann Transl Med*. 2019;7(20):581. doi:10.21037/atm.2019.09.58
17. Salvagno GL, Sanchis-Gomar F, Picanza A, Lippi G. Red blood cell distribution width: a simple parameter with multiple clinical applications. *Crit Rev Clin Lab Sci*. 2015;52(2):86–105. doi:10.3109/10408363.2014.992064
18. Li N, Zhou H, Tang Q. Red blood cell distribution width: a novel predictive indicator for cardiovascular and cerebrovascular diseases. *Dis Markers*. 2017;2017:7089493. doi:10.1155/2017/7089493
19. Feng GH, Li HP, Li QL, Fu Y, Huang RB. Red blood cell distribution width and ischaemic stroke. *Stroke Vasc Neurol*. 2017;2(3):172–175. doi:10.1136/svn-2017-000071
20. Fan X, Deng H, Wang X, et al. Association of red blood cell distribution width with severity of hepatitis B virus-related liver diseases. *Clin Chim Acta*. 2018;482:155–160. doi:10.1016/j.cca.2018.04.002
21. Southam L, Gilly A, Süveges D, et al. Whole genome sequencing and imputation in isolated populations identify genetic associations with medically-relevant complex traits. *Nat Commun*. 2017;8:15606. doi:10.1038/ncomms15606
22. Chen MH, Raffield LM, Mousas A, et al. Trans-ethnic and ancestry-specific blood-cell genetics in 746,667 Individuals from 5 global populations. *Cell*. 2020;182(5):1198–213.e14. doi:10.1016/j.cell.2020.06.045
23. Elkhalifa AME, Abdul-Ghani R. Hematological indices and abnormalities among patients with uncomplicated falciparum malaria in Kosti city of the White Nile state. *Sudan*. 2021;21(1):507.
24. Beyan C, Kaptan K, Beyan E, Turan M. The platelet count/mean corpuscular hemoglobin ratio distinguishes combined iron and vitamin B12 deficiency from uncomplicated iron deficiency. *Int J Hematol*. 2005;81(4):301–303. doi:10.1532/IJH97.E0311