

# A Genome-wide Association Study of Copy Number Variation in Hematological Parameters in the Korean Population

Ka-Kyung Kim<sup>1,2</sup>, Yoon Shin Cho<sup>3</sup>, Nam H. Cho<sup>4</sup>, Chol Shin<sup>5</sup> and Jong-Won Kim<sup>2\*</sup>

<sup>1</sup>Samsung Biomedical Research Institute, Samsung Medical Center, Seoul 135-710, Korea, <sup>2</sup>Department of Laboratory Medicine and Genetics, Sungkyunkwan University School of Medicine, Samsung Medical Center, Seoul 135-710, Korea, <sup>3</sup>Center for Genome Science, National Institute of Health, Seoul 122-701, Korea, <sup>4</sup>Department of Preventive Medicine, Ajou University School of Medicine, Suwon 443-749, Korea, <sup>5</sup>Division of Pulmonary and Critical Care Medicine, Department of Internal Medicine, Korea University Ansan Hospital, Ansan 425-707, Korea

## Abstract

Abnormal hematological values are associated with various disorders including cancer and cardiovascular, metabolic, infectious, and immune diseases. We report the copy number variations (CNVs) in clinically relevant hematological parameters, including hemoglobin level, red and white blood cell counts, platelet counts, and red blood cell (RBC) volume. We describe CNVs in several loci associated with these hematological parameters in 8,842 samples from Korean population-based studies. The data that we evaluated included four RBC parameters, one platelet parameter, and one associated with total white blood cell (WBC) count, exceeding the genome-wide significance. We show that CNVs in hematological parameters are associated with some loci, different from previously associated loci reported in single nucleotide polymorphism (SNP) association studies.

**Keywords:** hematological parameters, blood cell count, erythrocytes, genome-wide association study, copy number variation

## Introduction

The hematopoietic system has various functions including oxygen transport, immunity, blood wall surveillance, homeostasis, and wound curing. Hematological param-

eters include not only the number of white blood cells (WBC) and red blood cells (RBC), but also the number of platelets. The hemoglobin (Hb), the hematocrit (Hct), and the mean corpuscular hemoglobin concentration (MCHC) were also evaluated. All of these parameters are heritable, vary among individuals and are used as an index for clinical treatment (Ferreira *et al.*, 2009). Abnormal values of hemoglobin concentration are the main indicator of common genetic diseases such as cancer, infection, and immune disease.

The RBCs are in charge of blood transport to tissues. The MCHC value is an index of anemia, and Hct, the proportion of blood volume that is occupied by RBCs, is the index that defines anemia. An increased number of WBCs is reported to be a risk factor for coronary artery disease (CAD), myocardial infarction (MI), and leukemia (Soranzo *et al.*, 2009). Malfunctions in RBC production such as anemia or erythrocytosis are related to hypertension, cardiovascular diseases, asthma, and atherosclerosis, the heritability is known to be 40~90% (Ganesh *et al.*, 2009). Platelets have important roles in the clotting system, and increased platelet count is related to cardiovascular diseases and cancer (Soranzo *et al.*, 2009). A recent genome-wide association study (GWAS) of single nucleotide polymorphism (SNP) has shown a significant association among hematological traits (Ganesh *et al.*, 2009; Kamatani *et al.*, 2010; Soranzo *et al.*, 2009).

The copy number variations (CNV) representing structural variants were recently reported to be related to disease susceptibility such as HIV infection, autoimmune disease, asthma, and nervous disease. However, the association of CNV to hematological traits has not been reported.

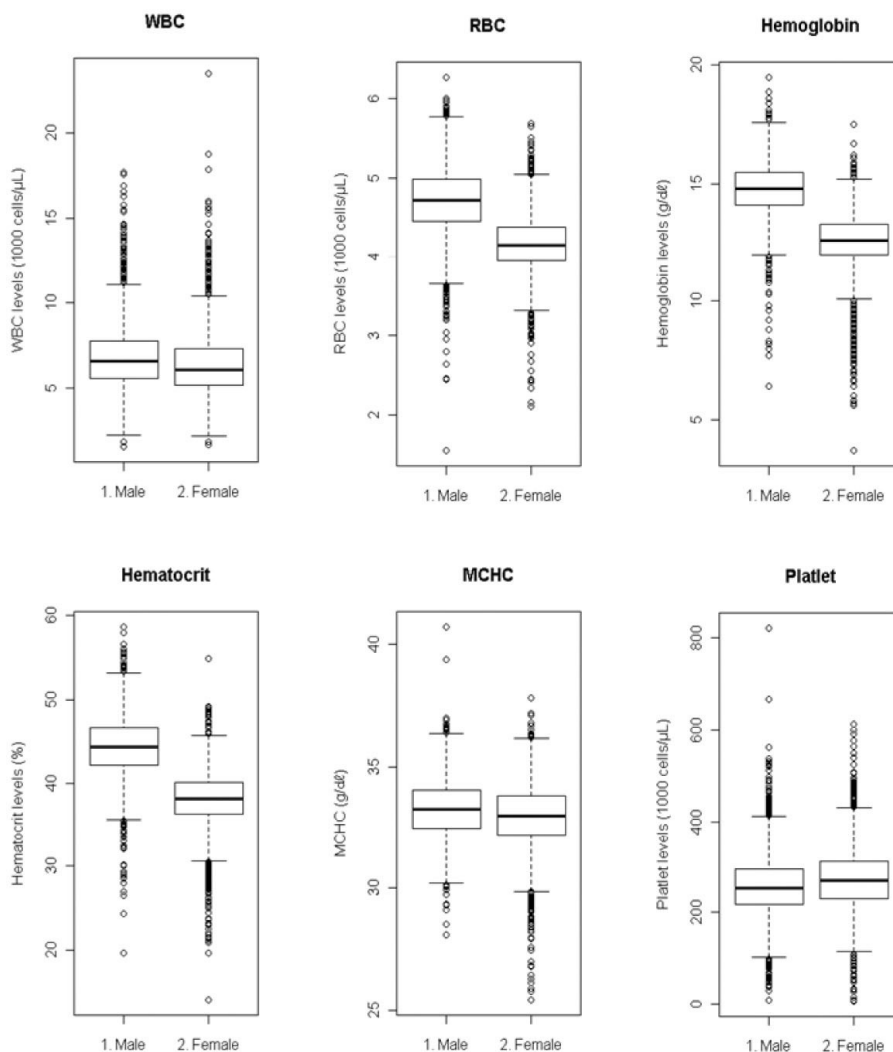
The purpose of this study was to determine a genetic relationship between SNPs and CNVs of hematological traits in the general Korean population. We carried out GWAS of CNV for the chosen traits and identified several loci associated with these traits. The loci identified in our study are different from those reported in a previous GWAS of SNP.

## Methods

### Study samples

We collected data from 10,004 individuals in two populations from rural Ansong and urban Ansan cohorts as

\*Corresponding author: E-mail kimjw@skku.edu  
Tel +82-2-3410-2705, Fax +82-2-3410-2719  
Accepted 15 September 2010



**Fig. 1.** Box plot for hematological traits.

part of the Korean Genome Epidemiology Study (KoGES). The initial samples included 5,018 and 5,020 respective participants aged 40 to 69. Detailed information was shown in a previous study (Cho *et al.*, 2009). The characteristics of the hematological traits are summarized in Fig. 1 and Table 1.

### Genotyping and quality control

The majority of genomic DNA genotyped on the Affymetrix Genome-Wide Human SNP array 5.0 was isolated from peripheral blood drawn from cohort participants. Samples were genotyped using 500 ng of genomic DNA and the Bayesian Robust Linear Modeling using the Mahalanobis Distance (BRLMM) Genotyping Algorithm for genotype determination of 500,568 SNPs. Samples were excluded due to high missing genotype call rate, high heterozygosity, and gender inconsistencies. Also, any samples that had been obtained from in-

dividuals who had developed any kind of cancer or high pairwise identity-by-state (IBS) were also excluded, leaving a total of 8,842 individuals.

### CNV calling

We used PennCNV to perform CNV calling which simultaneously incorporates the log R ratio and B allele frequency based on a hidden Markov model (HMM) framework. We adopted the detail state definition for more precise modeling of CNV events (Colella *et al.*, 2007). The procedure for CNV calling is described in a previous report (Wang *et al.*, 2007).

### Statistical analysis

Most association analyses were conducted using PLINK (version 2.6) and R programs (version 2.9). For each hematological trait, we set control where the value was

**Table 1.** Summary of hematological traits

Traits	Sex	N	Mean	Std	Max	Q3	Median	Q1	Min	Range
WBC	Male	4182	6.8	1.9	17.7	7.8	6.6	5.6	1.6	16.1
	Female	4659	6.4	1.8	23.5	7.3	6.1	5.2	1.7	21.8
	All	8841	6.6	1.8	23.5	7.6	6.3	5.3	1.6	21.9
RBC	Male	4182	4.7	0.4	6.3	5.0	4.7	4.5	1.6	4.7
	Female	4659	4.2	0.3	5.7	4.4	4.2	3.9	2.1	3.6
	All	8841	4.4	0.5	6.3	4.7	4.4	4.1	1.6	4.7
Hemoglobin	Male	4182	14.7	1.1	19.5	15.5	14.8	14.1	6.4	13.1
	Female	4659	12.5	1.2	17.5	13.3	12.6	12.0	3.7	13.8
	All	8841	13.6	1.6	19.5	14.7	13.5	12.5	3.7	15.8
Hematocrit	Male	4182	44.3	3.4	58.7	46.6	44.4	42.2	19.6	39.1
	Female	4659	38.0	3.3	54.9	40.1	38.2	36.3	14.0	40.9
	All	8841	41.0	4.6	58.7	44.3	40.9	37.9	14.0	44.7
Platelet	Male	4182	258.0	65.0	821.0	295.0	253.0	217.0	9.0	812.0
	Female	4659	273.8	64.6	612.0	311.0	269.0	231.0	7.0	605.0
	All	8841	266.3	65.3	821.0	304.0	261.0	224.0	7.0	814.0
MCHC	Male	4182	33.3	1.2	40.7	34.0	33.3	32.5	28.1	12.6
	Female	4659	33.0	1.6	100.0	33.8	33.0	32.2	25.4	74.6
	All	8841	33.1	1.4	100.0	33.9	33.1	32.3	25.4	74.6

Std, standard deviation; WBC, white blood cell; RBC, Red blood cell; MCHC, mean corpuscular hemoglobin concentration. The numbers were rounded off to the nearest tenths.

**Table 2.** Normal values for hematological traits. Square indicates the normal values of the control group, and Italics indicates the lower values for the case group *Italic: case, Square: control*

Hematological traits	Normal value	Low	Normal	High	Missing
White Blood Cell (WBC)	4-11 × 10 <sup>3</sup> cells/ul	<i>339</i>	8309	193	1
Red Blood Cell (RBC)	♂ 4.4 ~ 6 × 10 <sup>3</sup> /ul ♀ 4 ~ 5.4 × 10 <sup>3</sup> /ul	<i>4830</i>	3918	93	1
Hemoglobin (HB)	♂ 14 ~ 18 g/dl ♀ 12 ~ 16 g/dl	<i>758</i>	8073	10	1
Hematocrit (Hct)	♂ 40 ~ 54% ♀ 36 ~ 47%	<i>1342</i>	7474	25	1
Mean Corpuscular Hemoglobin Concentration (MCHC=Hb/Hct)	32 ~ 36 g/dl	<i>1485</i>	7295	61	1
Platelet	150 × 10 <sup>3</sup> ~ 450 × 10 <sup>3</sup> cells/ul	<i>216</i>	8540	85	1

within normal value and case where the values were lower than normal (Table 2). The case-control status of hematological traits was tested for association using Fisher's exact test.

## Results

The GWAS results with Manhattan plots are shown in Fig. 2 and Supplementary Fig. 1-6.

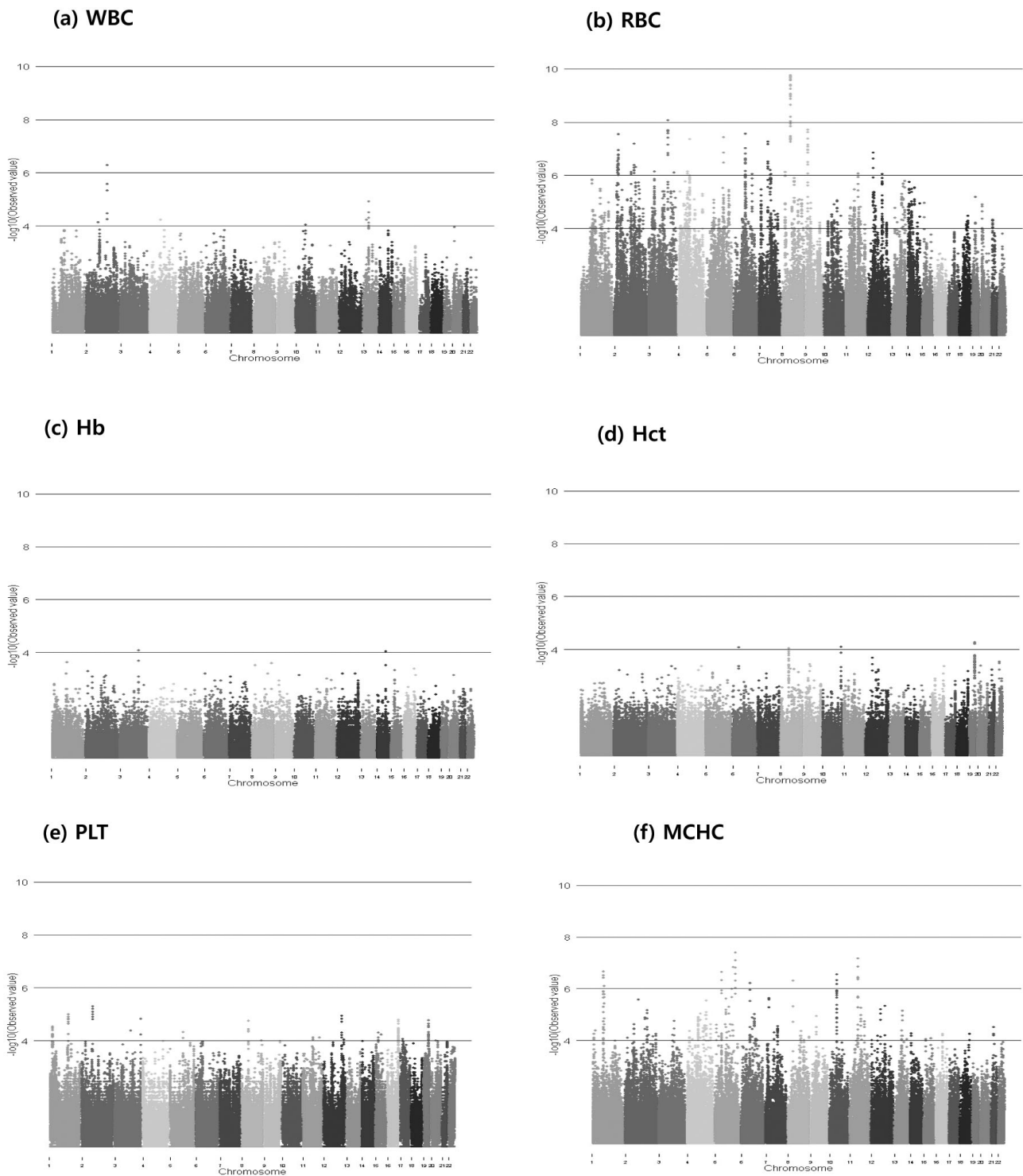
### CNV detection

The average number of CNVs per genome was 41.2 kb, larger than those of the CNVs previously reported using lower resolution platforms. However, these results are

similar to those of a previous report of healthy, unrelated Korean people provided by the Korea National Institute of Health (KNIH). The median copy number gain was 54.4 kb (range 0.02-8788.5 kb), and that of copy number loss was 47.6 kb (range 0.03-20075.4 kb). Copy number losses were 2.8 times more frequent and smaller than were copy number gains. Our range of CNV size was wider and had less frequent ratio losses to gains compared to those in previous reports (Yim *et al.*, 2010).

### White blood cell count

GWAS for WBC count revealed few newly associated loci. Association of CNVs with WBC was found in loci



**Fig. 2.** Genome-wide scan for hematological traits. The  $-\log_{10}$  of Fisher's exact test p-value for CNVs for each hematological trait.

on 2q22 ( $p=5.2 \times 10^{-7}$ ). The nearest gene, nebulin (NEB) encodes the giant nebulin protein component of the cytoskeletal matrix that coexists with the myosin and actin

filaments within the sarcomeres of skeletal muscle and is expressed in most tissues.

### Red blood cell traits

We performed GWAS for four erythrocyte-related traits, red blood cell (RBC) count, hemoglobin (Hb) concentration, hematocrit (Hct), and mean corpuscular hemoglobin concentration (MCHC).

Clustered CNV gains on 8p11.11 ( $p=1.8 \times 10^{-10} \sim 5.5 \times 10^{-8}$ ), a nearby centromere of chromosome 8, are associated with RBCs (Supplementary Fig. 2). The nearest gene was a POTE ankyrin domain family, member A (POTEA), the protein of which is known to be expressed in pancreas, ovary, testis and placenta. The CNV losses showed genome-wide significance for association with RBC on 3q22 ( $p=8.5 \times 10^{-9}$ ). The neighbor gene, COL6A5 (collagen alpha-5(VI) chain; collagen, type XXIX, alpha 1), has been linked to atopic dermatitis and is expressed in lymph, lung, ovary, brain, retina, bladder, and skin.

We observed a locus associated with Hb in CPNE4 on 3q22.1 ( $p=2.6 \times 10^{-5}$ ). CPNE4 is calcium-dependent and regulates molecular events at the interface of the cell membrane and cytoplasm.

The most significant association with Hb was found in CPNE4 on 3q22.1 ( $p=8.3 \times 10^{-5}$ ), and this gene encodes calcium-dependent membrane-binding proteins regulating molecular events at the interface of the cell membrane and cytoplasm.

We identified a Hct-associated loci on 6p21.2 ( $p=8.1 \times 10^{-5}$ ) and the nearest gene, hypothetical LOC401260 (FLJ41649), that had no functional importance. Variation\_4715 reported by Wong *et al.* and Variation\_29871 reported by Jakobsson *et al.* are located in this region. Additional signals were collectively found in FAM53B, a family with sequence similarity 53, member B, on 10q26.13 ( $p=7.9 \times 10^{-5}$ ), pregnancy specific beta-1-glycoprotein 11 (PSG11), and pregnancy specific beta-1-glycoprotein 6 (PSG6) on 19q13.31. PSGs are mainly produced by the placental syncytiotrophoblasts during pregnancy and belong to the immunoglobulin superfamily.

We identified CNV losses associated with MCHC on 6p21.2 ( $p=8.1 \times 10^{-5}$ ). The nearest gene, DEAH (Asp-Glu-Ala-His) box polypeptide 9 pseudogene (AB231715), had no functional importance yet but Variation\_3851 and Variation\_2925 were reported by Redon *et al.* in the region.

### Platelet count

The most significant association with PLT was found on 2p13.1 ( $p=5.0 \times 10^{-6}$ ) and spanned many genes, such as DEAQ box RNA-dependent ATPase 1 (DQX1), ancient ubiquitous protein 1 (AUP1), HtrA serine peptidase 2 (HTRA2), and lysyl oxidase-like 3 (LOXL3). Among those

genes, AUP1 is involved in binding ubiquitin-conjugating enzymes. AUP1 also binds to the cytoplasmic tail of integrin alpha (IIb) subunits, which play a crucial role in the integrin alpha (IIb) beta (3) inside-out signaling in platelets and megakaryocytes that leads to platelet aggregation and thrombus formation. HTRA2 encodes a serine protease that induces apoptosis. LOXL3 is essential to the biogenesis of connective tissue having roles in developmental regulation, senescence, tumor suppression, cell growth control, and chemotaxis to each member of the family.

### Discussion

Thousands of structural variants, consisting of 5-12% of the human genome, are not only revealing, but they are also expected in gene expression. CNVs representing structural variants are reported to be related to disease susceptibility. The frequency of CNVs is different between populations, and more than 60% of detected CNV are known to be located outside of genes (Conrad *et al.*, 2010). Most common CNV (MAF > 5%) is well tagged on SNP arrays, and this suggests that a possible contribution of common CNVs to disease analyzing the disease association of CNV-tagging SNPs (McCarroll *et al.*, 2008).

It is becoming increasingly clear that genetic studies of complex diseases and quantitative traits must pay closer attention to the contributions of CNVs. Recent developments in genome-wide analysis have helped to identify various loci related to quantitative traits, and complex diseases, hemostasis, thrombosis, and cardiovascular disease have been studied broadly for their relationships to these traits (Ganesh *et al.*, 2009). In contrast to the well-developed resources available for SNP association studies, we are still in the early phases of incorporating structural genetic variation in genome-wide association studies. There was no strong genome-wide association of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls in WTCCC consortium though common CNVs are well tagged by SNP (Craddock *et al.*, 2010). In contrast to the well-developed resources available for SNP association studies, we are still in the early phases of incorporating structural genetic variation in genome-wide association studies. There was no strong genome-wide association of CNVs in 16,000 cases of eight common diseases, and 3,000 shared controls in the WTCCC consortium though common CNVs are well tagged according to SNP (Ganesh *et al.*, 2009; Soranzo *et al.*, 2009). The associations of these hematological traits in the Japanese were replicated in many of the loci, in addition to some novel loci. Most of the known loci associated with the

hematological traits in other populations were replicated in a Korean in our previous study (data not shown).

In this study, we report several CNV loci associated with hematological traits. We found genome-wide significant associations with RBC and MCHC, and other hematological traits showed several peak signals. CNV gains on 8p11.11 ( $p=1.8 \times 10^{-10} \sim 5.5 \times 10^{-8}$ ) for association with RBC and CNV losses for association with RBC on 3q22 ( $p=8.5 \times 10^{-9}$ ) showed genome-wide significance, although the neighbor genes were not functionally related to hematological traits.

We observed WBC-associated loci in NEB on 2q22 ( $p=5.2 \times 10^{-7}$ ), which impacts the function of the cytoskeletal matrix, and 2p13.1 ( $p=5.0 \times 10^{-6}$ ), which includes DQX1, AUP1, HTRA2 and LOXL3. AUP1 plays a crucial role in the signaling of platelets and megakaryocytes, leading to platelet aggregation and thrombus formation. HTRA2 is related to apoptosis, and LOXL3 is essential to the biogenesis of connective tissue, although no relation with a hematological trait was reported.

We also found loci associated with Hb on 3q22.1 ( $p=2.6 \times 10^{-5}$ ), with Hct on 6p21.2 ( $p=8.1 \times 10^{-5}$ ) and with MCHC on 6p21.2 ( $p=8.1 \times 10^{-5}$ ). CPNE4 on 3q22.1 impacts the interaction between the cell membrane and the cytoplasm. No other neighbor genes have this function.

Previously reported Hct- and MCHC-associated CNV loci were detected, but no functional importance has been assigned to any of these loci (Iafate *et al.*, 2004). In the future, curate-associated loci should be identified using various algorithms, since CNV detection is algorithm-dependent. Furthermore, in silico results should be verified using quantitative PCR (qPCR) and functional analysis.

To the best of our knowledge, this is the first genome-wide association study of CNV with hematological traits. In conclusion, our study offers understanding of hematological disease susceptibility as well as personalized prediction and prevention through combined SNP and CNV GWAS.

## Acknowledgement

This study was supported by a grant of Korea Health 21 R&D Project, Ministry of Health and Welfare, Republic of Korea (0405-BC02-0604-0004).

## References

Cho, Y.S., Go, M.J., Kim, Y.J., Heo, J.Y., Oh, J.H., Ban, H.J., Yoon, D., Lee, M.H., Kim, D.J., Park, M., Cha, S.H., Kim, J.W., Han, B.G., Min, H., Ahn, Y., Park, M.S., Han, H.R., Jang, H.Y., Cho, E.Y., Lee, J.E., Cho, N.H., Shin,

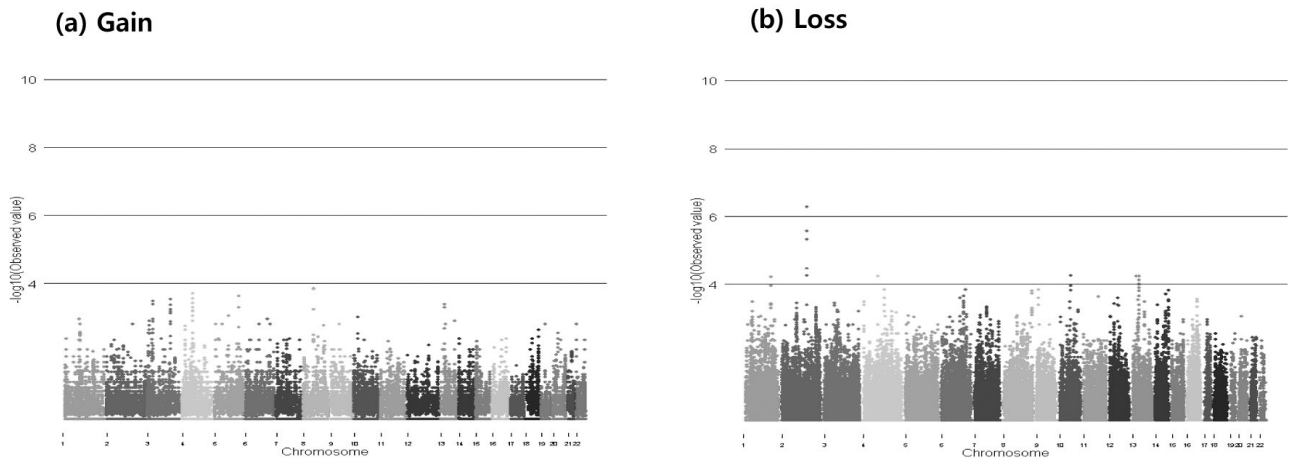
C., Park, T., Park, J.W., Lee, J.K., Cardon, L., Clarke, G., McCarthy, M.I., Lee, J.Y., Lee, J.K., Oh, B., and Kim, H.L. (2009). A large-scale genome-wide association study of Asian populations uncovers genetic factors influencing eight quantitative traits. *Nat. Genet.* 41, 527-534.

Colella, S., Yau, C., Taylor, J.M., Mirza, G., Butler, H., Clouston, P., Bassett, A.S., Seller, A., Holmes, C.C., and Ragoussis, J. (2007). QuantiSNP: an Objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data. *Nucl. Acids Res.* 35, 2013-2025.

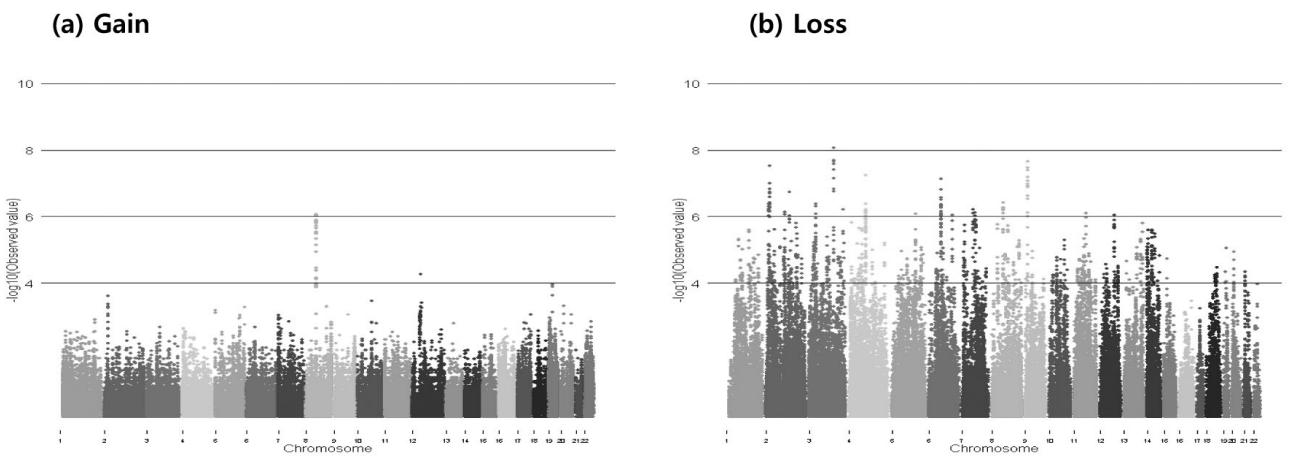
Conrad, D.F., Pinto, D., Redon, R., Feuk, L., Gokcumen, O., Zhang, Y., Aerts, J., Andrews, T.D., Barnes, C., Campbell, P., Fitzgerald, T., Hu, M., Ihm, C.H., Kristiansson, K., Macarthur, D.G., Macdonald, J.R., Onyiah, I., Pang, A.W., Robson, S., Stirrups, K., Valsesia, A., Walter, K., Wei, J., Tyler-Smith, C., Carter, N.P., Lee, C., Scherer, S.W., and Hurles, M.E. (2010). Origins and functional impact of copy number variation in the human genome. *Nature* 464, 704-712.

Craddock, N., Hurles, M.E., Cardin, N., Pearson, R.D., Plagnol, V., Robson, S., Vukcevic, D., Barnes, C., Conrad, D.F., Giannoulatou, E., Holmes, C., Marchini, J.L., Stirrups, K., Tobin, M.D., Wain, L.V., Yau, C., Aerts, J., Ahmad, T., Andrews, T.D., Arbury, H., Attwood, A., Auton, A., Ball, S.G., Balmforth, A.J., Barrett, J.C., Barroso, I., Barton, A., Bennett, A.J., Bhaskar, S., Blaszczyk, K., Bowes, J., Brand, O.J., Braund, P.S., Bredin, F., Breen, G., Brown, M.J., Bruce, I.N., Bull, J., Burren, O.S., Burton, J., Byrnes, J., Caesar, S., Clee, C.M., Coffey, A.J., Connell, J.M., Cooper, J.D., Dominiczak, A.F., Downes, K., Drummond, H.E., Dudakia, D., Dunham, A., Ebbs, B., Eccles, D., Edkins, S., Edwards, C., Elliot, A., Emery, P., Evans, D.M., Evans, G., Eyre, S., Farmer, A., Ferrier, I.N., Feuk, L., Fitzgerald, T., Flynn, E., Forbes, A., Forty, L., Franklyn, J.A., Freathy, R.M., Gibbs, P., Gilbert, P., Gokumen, O., Gordon-Smith, K., Gray, E., Green, E., Groves, C.J., Grozeva, D., Gwilliam, R., Hall, A., Hammond, N., Hardy, M., Harrison, P., Hassanali, N., Hebaishi, H., Hines, S., Hinks, A., Hitman, G.A., Hocking, L., Howard, E., Howard, P., Howson, J.M., Hughes, D., Hunt, S., Isaacs, J.D., Jain, M., Jewell, D.P., Johnson, T., Jolley, J.D., Jones, I.R., Jones, L.A., Kirov, G., Langford, C.F., Lango-Allen, H., Lathrop, G.M., Lee, J., Lee, K.L., Lees, C., Lewis, K., Lindgren, C.M., Maisuria-Armer, M., Maller, J., Mansfield, J., Martin, P., Massey, D.C., McArdle, W.L., McGuffin, P., McLay, K.E., Mentzer, A., Mimmack, M.L., Morgan, A.E., Morris, A.P., Mowat, C., Myers, S., Newman, W., Nimmo, E.R., O'Donovan, M.C., Onipinla, A., Onyiah, I., Ovington, N.R., Owen, M.J., Palin, K., Parnell, K., Pernet, D., Perry, J.R., Phillips, A., Pinto, D., Prescott, N.J., Prokopenko, I., Quail, M.A., Rafelt, S., Rayner, N.W., Redon, R., Reid, D.M., Renwick, Ring, S.M., Robertson, N., Russell, E., St Clair, D., Sambrook, J.G., Sanderson, J.D., Schuilenburg, H., Scott, C.E., Scott, R., Seal, S., Shaw-Hawkins, S., Shields, B.M., Simmonds, M.J., Smyth, D.J., Somaskantharajah, E., Spanova, K., Steer, S., Stephens, J., Stevens, H.E.,

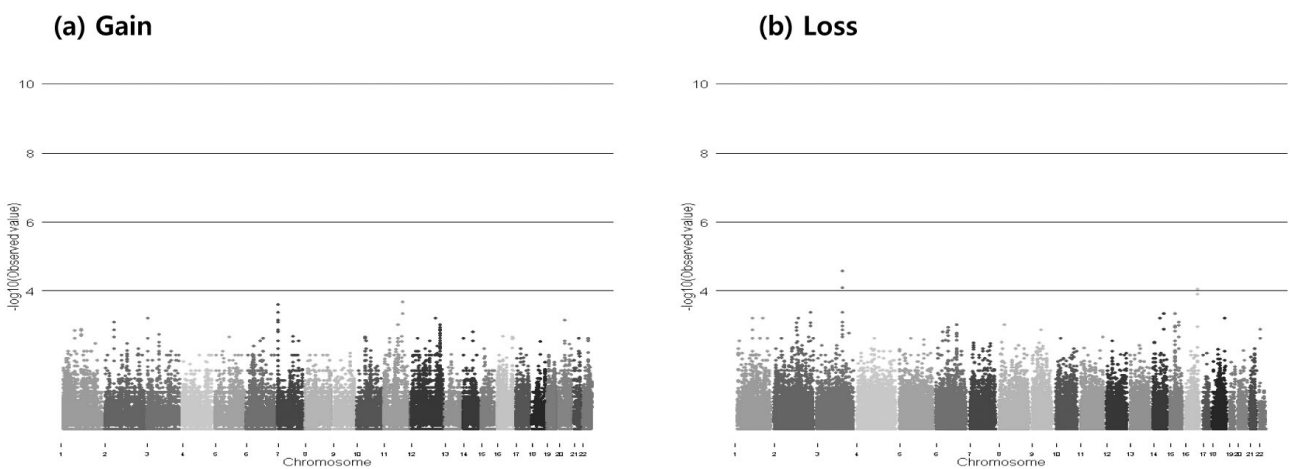
- Stone, M.A., Su, Z., Symmons, D.P., Thompson, J.R., Thomson, W., Travers, M.E., Turnbull, C., Valsesia, A., Walker, M., Walker, N.M., Wallace, C., Warren-Perry, M., Watkins, N.A., Webster, J., Weedon, M.N., Wilson, A.G., Woodburn, M., Wordsworth, B.P., Young, A.H., Zeggini, E., Carter, N.P., Frayling, T.M., Lee, C., McVean, G., Munroe, P.B., Palotie, A., Sawcer, S.J., Scherer, S.W., Strachan, D.P., Tyler-Smith, C., Brown, M.A., Burton, P.R., Caulfield, M.J., Compston, A., Farrall, M., Gough, S.C., Hall, A.S., Hattersley, A.T., Hill, A.V., Mathew, C.G., Pembrey, M., Satsangi, J., Stratton, M.R., Worthington, J., Deloukas, P., Duncanson, A., Kwiatkowski, D.P., McCarthy, M.J., Ouwehand, W., Parkes, M., Rahman, N., Todd, J.A., Samani, N.J., and Donnelly, P. (2010). Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature* 464, 713-720.
- Ferreira, M.A., Hottenga, J.J., Warrington, N.M., Medland, S.E., Willemsen, G., Lawrence, R.W., Gordon, S., de Geus, E.J., Henders, A.K., Smit, J.H., Campbell, M.J., Wallace, L., Evans, D.M., Wright, M.J., Nyholt, D.R., James, A.L., Beilby, J.P., Penninx, B.W., Palmer, L.J., Frazer, I.H., Montgomery, G.W., Martin, N.G., and Boomsma, D.I. (2009). Sequence variants in three loci influence monocyte counts and erythrocyte volume. *Am. J. Hum. Genet.* 85, 745-749.
- Ganesh, S.K., Zakai, N.A., van Rooij, F.J., Soranzo, N., Smith, A.V., Nalls, M.A., Chen, M.H., Kottgen, A., Glazer, N.L., Dehghan, A., Kuhnel, B., Aspelund, T., Yang, Q., Tanaka, T., Jaffe, A., Bis, J.C., Verwoert, G.C., Teumer, A., Fox, C.S., Guralnik, J.M., Ehret, G.B., Rice, K., Felix, J.F., Rendon, A., Eiriksdottir, G., Levy, D., Patel, K.V., Boerwinkle, E., Rotter, J.I., Hofman, A., Sambrook, J.G., Hernandez, D.G., Zheng, G., Bandinelli, S., Singleton, A.B., Coresh, J., Lumley, T., Uitterlinden, A.G., Vangils, J.M., Launer, L.J., Cupples, L.A., Oostra, B.A., Zwaginga, J.J., Ouwehand, W.H., Thein, S.L., Meisinger, C., Deloukas, P., Nauck, M., Spector, T.D., Gieger, C., Gudnason, V., van Duijn, C.M., Psaty, B.M., Ferrucci, L., Chakravarti, A., Greinacher, A., O'Donnell, C.J., Witteman, J.C., Furth, S., Cushman, M., Harris, T.B., and Lin, J.P. (2009). Multiple loci influence erythrocyte phenotypes in the CHARGE Consortium. *Nat. Genet.* 41, 1191-1198.
- Iafraite, A.J., Feuk, L., Rivera, M.N., Listewnik, M.L., Donahoe, P.K., Qi, Y., Scherer, S.W., and Lee, C. (2004). Detection of large-scale variation in the human genome. *Nat. Genet.* 36, 949-951.
- Kamatani, Y., Matsuda, K., Okada, Y., Kubo, M., Hosono, N., Daigo, Y., Nakamura, Y., and Kamatani, N. (2010). Genome-wide association study of hematological and biochemical traits in a Japanese population. *Nat. Genet.* 42, 210-215.
- McCarroll, S.A., Kuruvilla, F.G., Korn, J.M., Cawley, S., Nemes, J., Wysoker, A., Shaper, M.H., de Bakker, P.I., Maller, J.B., Kirby, A., Elliott, A.L., Parkin, M., Hubbell, E., Webster, T., Mei, R., Veitch, J., Collins, P.J., Handsaker, R., Lincoln, S., Nizzari, M., Blume, J., Jones, K.W., Rava, R., Daly, M.J., Gabriel, S.B., and Altshuler, D. (2008). Integrated detection and population-genetic analysis of SNPs and copy number variation. *Nat. Genet.* 40, 1166-1174.
- Soranzo, N., Spector, T.D., Mangino, M., Kuhnel, B., Rendon, A., Teumer, A., Willenborg, C., Wright, B., Chen, L., Li, M., Salo, P., Voight, B.F., Burns, P., Laskowski, R.A., Xue, Y., Menzel, S., Altshuler, D., Bradley, J.R., Bumpstead, S., Burnett, M.S., Devaney, J., Doring, A., Elosua, R., Epstein, S.E., Erber, W., Falchi, M., Garner, S.F., Ghorri, M.J., Goodall, A.H., Gwilliam, R., Hakonarson, H.H., Hall, A.S., Hammond, N., Hengstenberg, C., Illig, T., König, I.R., Knouff, C.W., McPherson, R., Melander, O., Mooser, V., Nauck, M., Nieminen, M.S., O'Donnell, C.J., Peltonen, L., Potter, S.C., Prokisch, H., Rader, D.J., Rice, C.M., Roberts, R., Salomaa, V., Sambrook, J., Schreiber, S., Schunkert, H., Schwartz, S.M., Serbanovic-Canic, J., Sinisalo, J., Siscovick, D.S., Stark, K., Surakka, I., Stephens, J., Thompson, J.R., Volker, U., Volzke, H., Watkins, N.A., Wells, G.A., Wichmann, H.E., Van Heel, D.A., Tyler-Smith, C., Thein, S.L., Kathiresan, S., Perola, M., Reilly, M.P., Stewart, A.F., Erdmann, J., Samani, N.J., Meisinger, C., Greinacher, A., Deloukas, P., Ouwehand, W.H., and Gieger, C. (2009). A genome-wide meta-analysis identifies 22 loci associated with eight hematological parameters in the HaemGen consortium. *Nat. Genet.* 41, 1182-1190.
- Wang, K., Li, M., Hadley, D., Liu, R., Glessner, J., Grant, S.F., Hakonarson, H., and Bucan, M. (2007). PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res.* 17, 1665-1674.
- Yim, S.H., Kim, T.M., Hu, H.J., Kim, J.H., Kim, B.J., Lee, J.Y., Han, B.G., Shin, S.H., Jung, S.H., and Chung, Y.J. (2010). Copy number variations in East-Asian population and their evolutionary and functional implications. *Hum. Mol. Genet.* 19, 1001-1008.



**Supplementary Fig. 1.** Genome-wide scan for WBC. The  $-\log_{10}$  of Fisher's exact test p-value for CNVs for WBC.

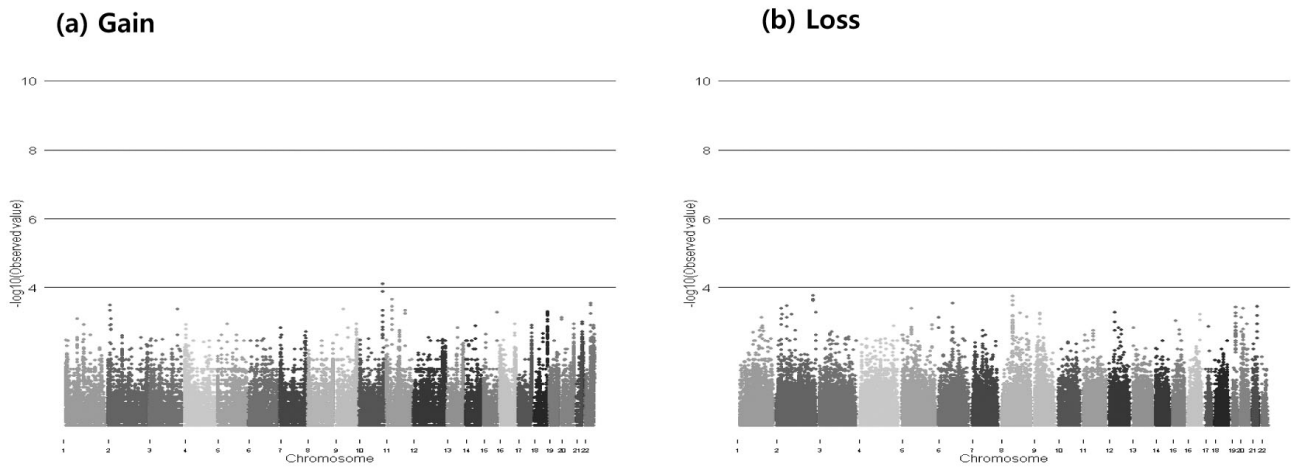


**Supplementary Fig. 2.** Genome-wide scan for RBC. The  $-\log_{10}$  of Fisher's exact test p-value for CNVs for RBC.

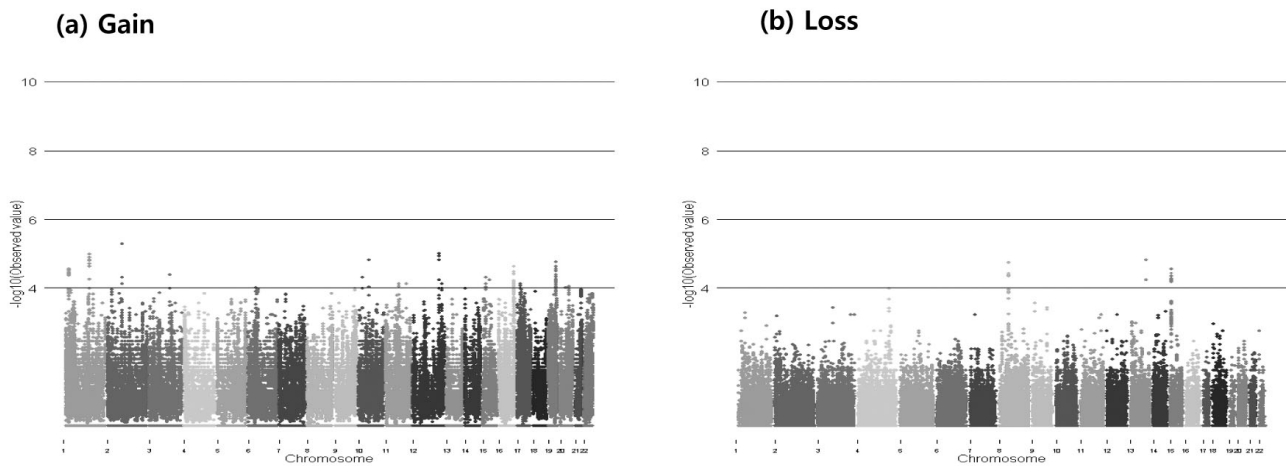


**Supplementary Fig. 3.** Genome-wide scan for Hb. The  $-\log_{10}$  of Fisher's exact test p-value for CNVs for Hb.

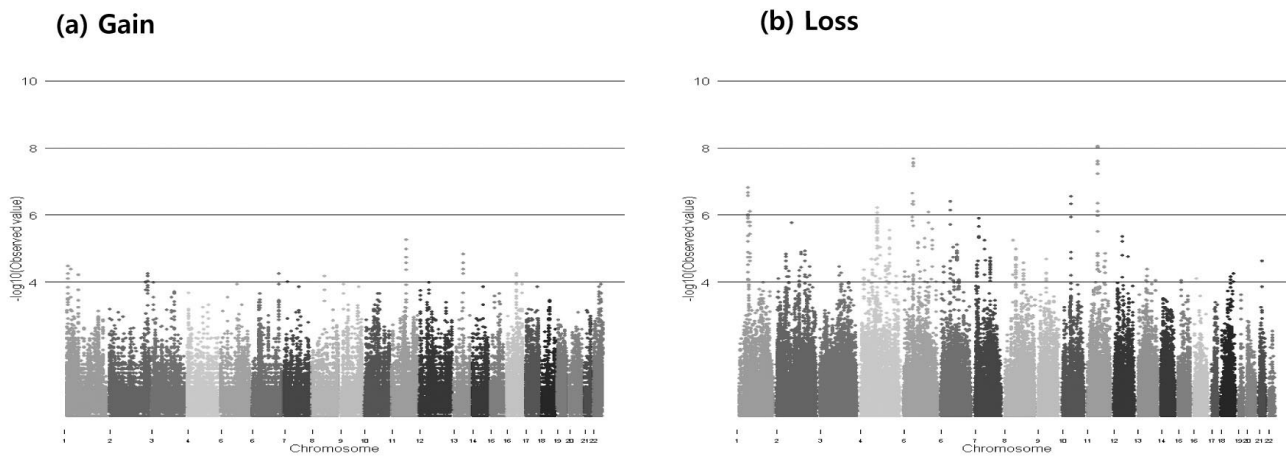




Supplementary Fig. 4. Genome-wide scan for Hct. The  $-\log_{10}$  of Fisher's exact test p-value for CNVs for Hct.



Supplementary Fig. 5. Genome-wide scan for PLT. The  $-\log_{10}$  of Fisher's exact test p-value for CNVs for PLT.



Supplementary Fig. 6. Genome-wide scan for MCHC. The  $-\log_{10}$  of Fisher's exact test p-value for CNVs for MCHC.