



Published in final edited form as:

Mamm Genome. 2013 April ; 24(0): 105–118. doi:10.1007/s00335-013-9448-0.

Genome-wide Association Mapping of Blood Cell Traits in Mice

Atila van Nas^{1,2,*}, Richard C. Davis^{2,*}, Brian Bennett², Luz Orozco^{1,2}, Calvin Pan¹,
Christoph D. Rau³, Eleazar Eskin^{1,5}, and Aldons J. Lusis^{1,2,3,4}

¹Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, California

²Department of Medicine/Division of Cardiology, David Geffen School of Medicine, University of California, Los Angeles, California

³Department of Microbiology, Immunology & Molecular Genetics, David Geffen School of Medicine, University of California, Los Angeles, California

⁴Molecular Biology Institute, David Geffen School of Medicine, University of California, Los Angeles, California

⁵Department of Computer Science, University of California, Los Angeles, California

Abstract

Genetic variations in blood cell parameters can impact clinical traits. We report here the mapping of blood cell traits in a panel of 100 inbred strains of mice of the Hybrid Mouse Diversity Panel (HMDP) using genome-wide association (GWA). We replicated a locus previously identified using linkage analysis in several genetic crosses for mean corpuscular volume ¹ and a number of other red blood cell traits on distal chromosome 7. Our peak for SNP association to MCV occurred in a linkage disequilibrium (LD) block spanning from 109.38 to 111.75 Mb that includes *Hbb-b1*, the likely causal gene. Altogether, we identified 5 loci controlling red blood cell traits (on chromosomes 1, 7, 11, 12, and 16), and four of these correspond to loci for red blood cell traits reported in a recent human GWA study. For white blood cells, including granulocytes, monocytes, and lymphocytes, a total of six significant loci were identified on chromosomes 1, 6, 8, 11, 12 and 15. An average of ten candidate genes were found at each locus and those were prioritized by examining functional variants in the HMDP such as missense and expression variants. These new results provide intermediate phenotypes and candidate loci for genetic studies of atherosclerosis and cancer as well as inflammatory and immune disorders in mice.

Keywords

blood cell traits; genetics; association; linkage; red cell; white cell; mice

Correspondence: Aldons J. Lusis, University of California, Los Angeles, Med-Cardio/Microbio, Box 951679, 3730 MRL, Los Angeles, California 90095-1679, jlusis@mednet.ucla.edu.

*AvN and RCD contributed equally to this work

Authorship Contributions

AvN analyzed data and wrote the paper; RCD performed research and wrote the paper; BB performed research; LO performed research; CP analyzed data; CDR analyzed data; EE designed analysis software; AJL designed the research and wrote the paper.

Conflict of Interest Disclosures

The authors have nothing to disclose.

INTRODUCTION

Blood cell traits have been linked to risk of cancer, coronary artery disease, total mortality as well as other clinically relevant disorders²⁻⁴. Moreover, hematological parameters are under strong genetic control being both tightly controlled and highly heritable^{5,6}. Thus, an understanding of the genetic factors that regulate the production and properties of blood cells could contribute substantially to our ability to predict, treat or prevent a variety of diseases. Such considerations have led to a number of large genome-wide association studies (GWAS) of blood cell parameters, such as leukocyte count and hemoglobin concentration⁷⁻¹⁰. And, these studies have resulted in the identification of an impressive number of loci and associated candidate genes. However, while highly precise hematological data is relatively easy to collect in large human population studies, it is difficult to further characterize the role of any identified candidate genes and associated metabolic pathways in the same subjects, particularly when the likely tissue for such studies is not blood itself but tissues such as bone marrow or spleen.

Animal models are the logical alternative for investigating these candidate genes and pathways, and our laboratory has been using systems-genetics approach in mice to understand cardiometabolic traits such as atherosclerosis, obesity, diabetes, osteoporosis and heart failure¹¹. Inflammation is an important component of these traits, therefore we resolved to investigate the potential impact of blood cell variations by using association mapping to identify underlying genetic loci. Two previous studies have examined blood cell traits in mice using quantitative trait locus (QTL) analysis of genetic crosses^{12,13}. It is difficult to identify the causal genes in such linkage analyses because such analyses typically have excellent power but poor mapping resolution¹⁴. Both of these prior studies used additional strategies to improve resolution and to suggest candidate genes. Peters et al.,¹² carried out a dozen classical crosses involving a total of about 4000 mice. By combining data from multiple crosses and by using haplotype association mapping, they identified candidate genes for about a dozen loci, including the identification of *Hbb* as the causal gene for several red cell traits on chromosome 7. Kelada et al.,¹³ provided a demonstration of the potential power of the collaborative cross by mapping hematological parameters in 131 incipient strains to identify five loci. The QTL confidence intervals were from 3.8 to 14.0 Mb and in several cases they were able to narrow substantially by using a shared ancestry approach. Our approach for improving resolution in mouse genetic studies is a recently-developed association-based mapping strategy¹⁵. An essential aspect of this approach is the correction for population structure using an Efficient Mixed Model Algorithm (EMMA)¹⁶. Based on power calculations and strain diversity, we chose 100 inbred strains of mice for a panel which we term the Hybrid Mouse Diversity Panel (HMDP). Using transcript levels as a model phenotype, we have shown that the panel has sufficient power to map several thousand expression QTL (eQTL) with excellent resolution. In the current study, we identify six loci for red blood cell phenotypes and eight loci for white blood cell phenotypes. For each locus, we identified a linkage disequilibrium (LD) block that contains the peak-association SNP as a region most likely to contain the underlying genetic variation. These LD blocks were typically 1.5 to 2 Mb but ranged from 0.8 Mb to 4 Mb. We identified a set of candidate genes within the LD blocks based on the presence of a local (*cis*-acting) eQTL or genes carrying non-synonymous SNPs within the coding region or splice site sequence variations. In several cases it was possible to validate these loci based on published mapping results from other groups. Most strikingly, four of five loci that we identified for red blood cell traits corresponded to loci that were identified in a recent human GWA study¹⁰ including several cases where the mouse and human studies identified the same candidate gene.

METHODS

C57BL/6J x DBA/2J F2 intercross

We previously carried out an F2 intercross between the inbred strains *DBA/2* and *C57BL/6*¹⁷. These parental mice were obtained from The Jackson Laboratory (Bar Harbor, ME). The male *C57BL/6* parents carried heterozygous deficiency in the leptin receptor (*db +/-*) and F1 progeny were selected for the presence of the mutant allele. Among F2 progeny, only those with homozygous deficiency in leptin receptor (*db/db*) were selected for further phenotype and genotype analysis.

Genomic DNA was isolated by standard procedures. Genotyping for the *db* mutation was carried out essentially as described by Horvat and Bünger¹⁸ except that the *AccI* digestion was carried out at 55°C for 1.5 h to increase efficiency of cutting. Genome-wide analysis of single nucleotide polymorphism (SNP) genotypes for quantitative trait locus (QTL) mapping was carried out using the AffymetrixGeneChip Mouse Mapping 5K SNP panel¹⁹. This panel carries >2,800 informative makers between *C57BL/6* and *DBA/2* (average spacing: 0.5–0.6 cM)

Phenotypic traits were log transformed to normalize the trait distribution. Logarithm of the odds (LOD) score distributions were calculated using the scanone function in R/qtl (<http://www.rqtl.org>) (<http://www.cran.r-project.org>)²⁰. Sex and age were treated as interactive covariates as needed²¹. In most cases, we carried out separate analysis of the 5 wk and 12 wk cohorts using sex as an interactive covariate. Permutations of $n = 1,000$ were carried out to determine thresholds for significance at $P < 0.05$ as a means to account for the multiple testing of many markers for association with each trait.

Animals were housed in vivaria accredited by the Association for Assessment and Accreditation of Laboratory Animal Care, and all procedures were approved by the UCLA Institutional Animal Care and Use Committee. The breeding colony was maintained on a 12 h light-dark cycle with lights on from 6 AM to 6 PM, and animals were fed a chow diet with 6% fat by weight. Male mice from the hybrid HMDP panel were purchased from the Jackson Labs. Mice were between 6 and 10 wk of age and to ensure adequate acclimatization to a common environment the mice were aged until 16 wk of age. All mice were maintained on a chow diet (Ralston-Purina Co.) until sacrifice at 16 wk of age. Following a 16-h fast, mice were bled retro-orbitally under isoflurane anesthesia and plasma lipids determined as previously described^{22, 23}.

HMDP strains and genotypes

The HMDP strains have been described in detail¹⁵. Genotypes for these inbred strains were obtained from the Broad Institute (<http://www.broadinstitute.org/mouse/hapmap>), and combined with the genotypes from Wellcome Trust Center for Human Genetics (WTCHG). Genotypes of RI strains at the Broad SNPs were inferred from WTCHG genotypes by imputing alleles at polymorphic SNPs among parental strains, calling ambiguous genotypes missing. Of the 140,000 SNPs available, 107,145 were informative with an allele frequency greater than 5% and were used for GWAS in the HMDP. For phenotyping, we used 2 to 18 animals of each strain ($N_{ave} = 7.06$, $N_{median} = 6$, $N_{modal} = 6$). The exact numbers for each strain are as previously published¹⁵.

Blood Cell Traits

Blood for hematology analysis and for plasma was collected from mice that were fasted for 4–5 h and bled 2–3 h after the beginning of the light cycle from the retro-orbital plexus under isoflurane anesthesia. Complete blood cell profiling was carried out using the Heska

(Loveland, CO) HemaTrue(TM) Veterinary Hematology Analyzer. Blood was collected in 20ul EDTA- coated glass capillaries and processed using standard procedures as per instructions from Heska. Three male adult mice were studied per strain.

Genome-wide association mapping of the HMDP

We applied the following linear mixed model to account for the population structure and genetic relatedness among strains in the genome-wide association mapping $y = \mu + x\beta + u + e$: where μ represents mean, x represents SNP effect, u represents random effects due to genetic relatedness with $\text{Var}(u) = \sigma_g^2 K$ and $\text{Var}(e) = \sigma_e^2$, where K represents IBS (identity-by-state) matrix across all genotypes. A restricted maximum likelihood (REML) estimate of σ_g^2 and σ_e^2 are computed using EMMA, and the association mapping is performed based on the estimated variance component with a standard F-test to test $\beta = 0$.

Genome-wide significance threshold

Genome-wide significance threshold in genome-wide association mapping is determined by the family-wise error rate (FWER) as the probability of observing one or more false positives across all SNPs per phenotype. We ran 100 different sets of permutation tests and parametric bootstrapping of size 1000, and observed that the mean and standard error of the genome-wide significance threshold at FWER of 0.05 were $3.9 \times 10^{-6} \pm 0.3 \times 10^{-6}$ and $4.0 \times 10^{-6} \pm 0.3 \times 10^{-6}$, respectively. This is approximately an order of magnitude larger than the significance threshold obtained by Bonferroni correction (4.6×10^{-7}). We also performed parametric bootstrapping under simulated genetic background effect from population structure using EMMA. With 50% and 100% of variance explained by genetic background, the thresholds were determined to be $1.6 \times 10^{-6} \pm 0.2 \times 10^{-6}$ and $1.7 \times 10^{-6} \pm 0.2 \times 10^{-6}$. The reduction in the significance threshold compared to no genetic background effect is due to the fact that inter-SNP correlation due to long-range LD reduces when conditioning on the population structure.

Linkage Disequilibrium and candidate mutations

Linkage disequilibrium was determined based on same SNP database used in EMMA analyses. Correlations between each SNP on a chromosome was determined and plotted in a heatmap using MATLAB. Significant correlation of $R^2 > 0.8$ between blocks of SNPs were determined to be indicative of LD. LD blocks for each significant locus were determined by visual approximation of the decay in LD. The Wellcome Trust Mouse Genomes Project Query Server was used to identify potentially causal mutations in genes of interest. The query was limited to mice contained within the HMDP, and for mutations affecting splice sites and mutations resulting in either frame shifts or non-synonymous mutations.

RESULTS

Analysis of blood cell parameters across a panel of inbred mouse strains

Analysis of hematological parameters in the HMDP was carried on a panel of 100 classical and recombinant inbred strains. Blood cell profiling was performed using the Heska Hematruer Veterinary Hematology Analyzer on 3 male mice per strain. Figure 1 shows the results for mean corpuscular volume [fl] (MCV) and for total white blood cell counts [$10^3/\mu\text{l}$] (WBC). The traits show wide strain-dependent variation suggesting complex underlying genetics. Strain averages for each hematological trait are shown in Supplemental Table 1. Trait-trait correlations are shown in Supplemental Table 2.

Genetic control of red blood cell traits

Previous linkage mapping studies^{12,13} identified a strong QTL peak on distal chromosome 7 at about 110 Mb for mean corpuscular volume (MCV)¹³ and for CHCM¹², an alternate measure of mean corpuscular hemoglobin concentration (MCHC). In the HMDP, we observed the same QTL on chromosome 7 QTL for MCV (Figure 2 and Table 1) and for the related traits of mean corpuscular hemoglobin (MCH) and red cell distribution width (RDW_a) (Table 1) along with a suggestive QTL for hematocrit (HCT). For comparison, in Figure 2, we also show the MCV QTL determined by linkage in a previous F2 cross between C57BL/6 and DBA¹⁷. Typical of linkage studies, the QTL peak was broad and the hemoglobin Hbb-b1 gene variant that is postulated to underlie this QTL¹² is encoded at 110.96 Mb, more than 15 Mb and hundreds of genes from the linkage QTL peak at 94Mb. By contrast, the association peak in the HMDP is quite sharp and the peak SNP is less than 1 Mb from Hbb-b1 (Figure 2). Additional novel association loci for red blood cell phenotypes of hematocrit (HCT), hemoglobin concentration (HGB) and red blood cell counts (RBC) were observed on chromosomes 1, 11, 12 and 16 (Figure 3 and Table 1).

We examined the significant loci for red blood cell traits for long range LD, previously described²⁵ as a potential confounding factor in GWAS analysis. We see no evidence of significant ($R^2 > 0.8$) LD between any of the red blood cell trait loci, and only one suggestive ($R^2 > 0.5$) LD between the chromosome 1 and 11 peaks (Table S3). This peak may be due to population substructure, which was not taken into account when computing LD values.

Genetic control of white blood cell traits

Association analysis in the HMDP identified 6 loci for total white blood cells, granulocytes, monocytes, and lymphocytes shown in Figures 4A-4D and in Table 2. Two of these loci, on chromosomes 6 and 12, exhibited significant associations for all classes of white cells, suggesting that they likely influence early stem cell progenitors or some inflammatory process common to all. The chromosome 8 locus which appears to be specific for monocytes (monocyte levels) exhibited the strongest association for all loci. Five out of the six loci were shared by monocytes and granulocytes, likely reflecting their similar developmental origins (Table 2). None of the loci overlapped with those for red cell parameters.

The white cell loci were also examined for long range LD. As in the red cell loci, we observe no significant ($R^2 > 0.8$) LD between any loci, but do observe two suggestive ($R^2 > 0.5$) LDs between the chromosome 1 and 6 peaks, and the 15 and 18 peaks (Table S4).

Candidate Genes in Associated Loci

In order to identify candidate genes for these associated loci, we first calculated the boundaries of the linkage disequilibrium (LD) blocks that encompassed the peak SNPs for each locus based on the correlation of SNPs among the HMDP strains (see Methods). For example, Figure 5 shows LD blocks along chromosome 8. The peak SNP for monocyte and overall white blood cell counts lies at 8.1 Mb in a 700 kb LD block that spans from 7.9 to 8.6 Mb. Boundaries for LD blocks associated with the other blood-trait QTLs are listed in Table 3. The median LD block-size was about 1.5 Mb although the LD block on chromosome 7 determined in this manner was unusually large, spanning nearly 12 Mb. However, the p-values for SNPs associated with MCV on Chr.7 fell sharply (by > 2 logs) outside the region from 109.38 to 111.75Mb, providing a core LD block for this trait spanning only 2.4 Mb (Supplemental Figure 1).

To identify potential candidate genes for these QTLs, we determined which genes within the LD block exhibited either structural variation (non-synonymous variations in the coding sequence) or splice-site variations.

The genes exhibiting structural variation are shown in Table 3. The number of such candidate genes ranged from 1 to 16 per LD block except for the unusually gene-rich LD block on chromosome 7 that harbors nearly 110 such variants. However, 87 of these are from the olfactory-receptor gene family that populates this region of chromosome 7. Of the remaining 23 non-synonymous variants, 5 are from the hemoglobin gene family including *Hbb-b1*, thought to underlie the MCV QTL¹².

For expression variation we examined local (likely *cis*-acting) eQTL in peritoneal macrophages²⁶. The macrophages were studied under basal conditions (medium) or in the presence of an acute inflammatory agent (bacterial lipopolysaccharide, LPS) or a chronic inflammatory agent (oxidized phospholipids). We performed association analysis for global transcript levels to identify local (within 5 Mb of the gene) expression quantitative trait loci (eQTL), most likely acting in *cis*, and those distal to the gene likely acting in *trans*. Altogether, there are more than 9000 genes that show heritable expression at a false-discovery-rate (FDR) of <5%. We used these data to identify genes likely to have local (*cis*) eQTL in cells of the white blood cell lineage. Table 4 lists the genes for white blood cell loci that exhibited local eQTL within the associated LD block under basal conditions or following stimulation with LPS or oxidized lipids. Because these genes show genetically determined variation in expression that co-localizes with the white blood cell traits, they constitute additional candidates for genes that are causal for these traits.

DISCUSSION

We have used association analysis with correction for population structure to map multiple loci for blood cell traits. The results confirm some previous findings and identify a number of novel loci. They also demonstrate the greatly increased mapping resolution for association as compared to linkage analysis in mice.

There have been a number of recent efforts using linkage and GWAS to map loci underlying blood cell traits both in mice^{12,13} and in humans^{7,8,27}. The major QTL we observed on mouse chromosome 7 was also reported by previous investigators^{12,13} and appears to arise from ancestral blocks carrying coding variations in duplicated adult hemoglobin genes¹². Thus, for strains that carry different alleles of this locus, there is a strong impact on hemoglobin hydration resulting in the reported difference in mean corpuscular hemoglobin concentration. The related traits of HCT, RDW and MCV are logically modulated by the same structural modification of the hemoglobin molecule and show correspondingly strong QTLs at the same locus. None of the human GWAS studies of blood cell traits^{7,8,27} have reported a corresponding QTL. Thus, similar coding variations in hemoglobin among the studied human populations are not sufficiently common to have produced a significant corresponding QTL.

Several other red cell QTLs that we observed in mice are replicated by QTLs seen in a recent GWAS study of red cell phenotypes of 135 thousand individuals of European and South Asian ancestry¹⁰. For instance, the LD block encompassing the hematocrit/hemoglobin QTL on chromosome 1 (Table 1) contains a non-synonymous SNP in the *Atp2b4* gene (Table 3), a gene that was also identified as potentially causal for a MCHC QTL in the human study. *Atp2b4* is a peripheral membrane calcium ATPase with no obvious connection to hemoglobin synthesis or structure but, the coincidence of GWAS results in human and mouse suggest that this gene may play a novel role in red blood cell metabolism.

Similarly, potential candidate genes for the RBC QTL on Chr.11 (Table 3) include *Trim58*, a gene also identified in the human study¹⁰. *Trim58* (tripartite motif-containing 58) is strongly expressed in bone marrow but has no established function. For the Chr 12 and Chr 16 RBC QTLs (Tables 1 and 3), none of the candidate genes we identified within the central LD block correspond to candidate genes in human GWAS studies. However, the recent human GWAS¹⁰ does appear to have identified syntenic loci. For instance, the human study identified *FNTB* (farnesyltransferase, CAAX box, beta), *MAX* (max protein) and *SMOCL1* (SPARC related modular calcium binding 1) as candidates for genes impacting MCV or MCH. In mice, these genes are located on Chr 12 at 76.8 Mb, 76.9 Mb and 81 Mb, respectively, immediately flanking the central LD block (78.7 to 80.5 Mb) (Table 3). Moreover, both *Fntb* and *Smoc1* carry non-synonymous SNPs among inbred strains that might underlie functional variation affecting red blood cell traits. Similarly, human GWAS identified *YDJC* (YdjC bacterial homolog) and *UBE2L3* (ubiquitin-conjugating enzyme E2L3) as candidates for genes impacting MCV. In mice, these genes map to Chr 16 at about 17.1 Mb, immediately distal to the central QTL (Tables 1 and 3).

Peters et al.¹² observed a QTL for red cell hemoglobin content in 3 or 4 of the 12 mouse crosses studied. The fact that no such QTL is observed in the present study suggests that the responsible alleles are not widespread in the HMDP panel or that the effect of these alleles is relatively small, or both.

For white blood cell traits, the chromosome 18 QTL was seen in linkage studies of incipient lines from the collaborative cross¹³. In addition, for the same locus, we observed strong QTL for all the white cell subtypes with the most significant being that for monocyte counts (Table 2). Finally, we observed significant WBC QTLs on chromosomes 1, 6, 8, 11, 12, 15 and 16 that were not seen in the collaborative cross. In contrast to the very high frequency of RBC locus-replication that we see between our mouse study and published human data, none of the white blood cell loci found in the HMDP was replicated in recent human GWAS analyses^{7,9}. In part, this could be due to the size of the available human studies. The van der Harst mapping of RBC-related phenotypes¹⁰ involved over 135,000 subjects while two recent GWAS studies of WBC-related phenotypes analyzed only 19,509 subjects⁴ and 16,388 subjects⁶. Moreover, because of the very large numbers of biological processes and environmental factors that likely impact inflammatory pathways, it seems probable that WBC genetics is more complex than that of RBC and that differences in environment may play a larger role in obscuring WBC QTLs. Finally, there may be some major differences in the key determinants of WBC numbers between mice and men so that genetic variation at a locus may have much higher impact in one species than the other.

As discussed in a recent review of genome wide association studies in mice²⁸, there is a tradeoff in power, resolution, and repeatability of the various approaches. The classic linkage approach has high power to detect the impact of genetic variants but the resolution is limited because of the small number of generations in which recombination can occur. Also, the number of genetic variants is limited to those found in the two parental strains. Moreover, the necessity to breed all the animals in a cross confers a high cost in time and resources before phenotyping can begin and, because each of the progeny is unique, there is a requirement to genotype every animal and no opportunity to do follow-up studies on additional animals of any given genotype. Peters et al.¹² partially overcame the issue of resolution by combining the results a dozen different crosses involving thousands of animals, but such a herculean effort is daunting to contemplate for most genetic studies. The collaborative cross avoids the necessity of breeding or genotyping because it involves the use of stock recombinant inbred strains and, because the collaborative cross strains incorporate the genomes of several wild-derived strains, there is the opportunity to monitor the impact of a much larger set of genetic variants than is available in a classic cross or the

HMDP. Ultimately, while the collaborative cross has constraints in resolution because of the limited number of generations for recombination in each strain, when they become available the large number of strains should help to overcome any resolution issues. It is exciting to see the debut of this effort in the mapping of hematologic traits by Kelada et al.¹³ The HMDP combines results from recombinant inbred strains, which give power approaching that of an F2 cross, and results from classic inbred strains confer high resolution due to the high number of generations that separate each strain from ancestral stocks. In fact, the HMDP is predicted to have resolution of about 2 Mb²⁸ which is consistent with the size of the LD blocks that we report here. The need to correct for population structure in the HMDP reduces power²⁸ and residual population structure after correction has the potential to produce false positives. Although we observed several instances of suggestive LD between pairs of loci, we conclude that this is not a major problem in the present study due to the fact that all of the red blood cell loci were replicated in previous mouse studies or in human GWAS. Genetic diversity is limited in the HMDP relative to the collaborative cross due to the decision to exclude wild derived strains in favor of better population structure corrections. In principle, this would lead to decreased ability to detect relevant QTL. However, at least in comparison to the initial collaborative cross studies by Kelada et al., this does not seem to be a large problem.

In summary, GWAS in human populations have been very successful in identifying loci for diseases and disease-related traits, but such studies have limited ability to examine gene-by-gene or gene-by-environment interactions, and the loci generally account for a very small fraction of the variance of the traits. Studies of natural variation in animal models can complement human studies by allowing analyses under controlled conditions and the examination of intermediate phenotypes such as transcript levels in tissues. An understanding of the genetic variation contributing to hematopoietic traits among common inbred strains may help explain differences in susceptibility to atherosclerosis, infection, and cancer.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

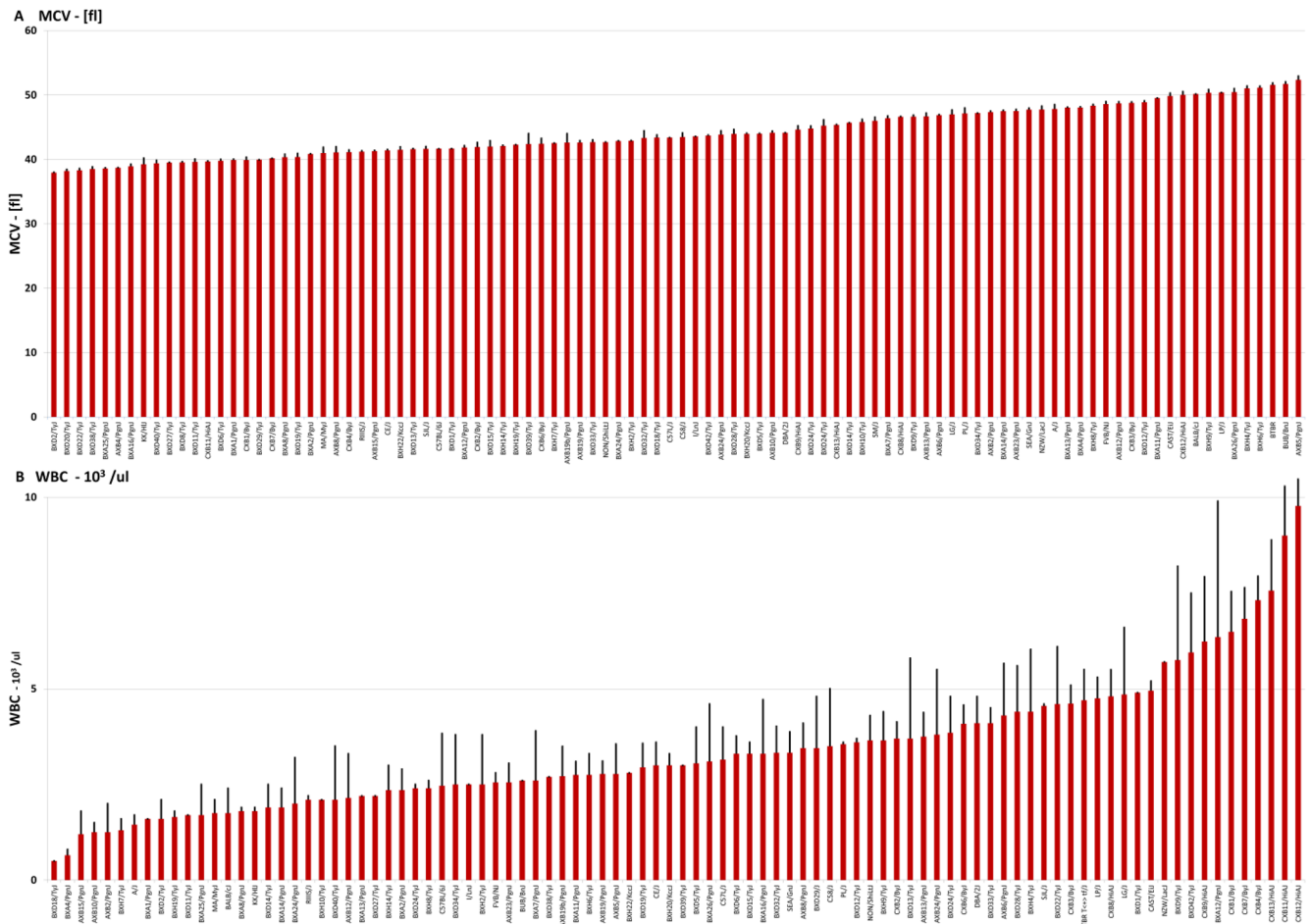
Acknowledgments

This work was supported by NIH grants HL30568 and HL28481.

References

1. Reshef DN, Reshef YA, Finucane HK, et al. Detecting novel associations in large data sets. *Science*. 2011; 334:1518–1524. [PubMed: 22174245]
2. Gillum RF, Mussolino ME, Madans JH. Counts of neutrophils, lymphocytes, and monocytes, cause-specific mortality and coronary heart disease: the NHANES-I epidemiologic follow-up study. *Ann Epidemiol*. 2005; 15:266–271. [PubMed: 15780773]
3. Lloyd-Jones DM, Camargo CA, Allen LA, Giugliano RP, O'Donnell CJ. Predictors of long-term mortality after hospitalization for primary unstable angina pectoris and non-ST-elevation myocardial infarction. *Am J Cardiol*. 2003; 92:1155–1159. [PubMed: 14609588]
4. Shankar A, Wang JJ, Rochtchina E, Yu MC, Kefford R, Mitchell P. Association between circulating white blood cell count and cancer mortality: a population-based cohort study. *Arch Intern Med*. 2006; 166:188–194. [PubMed: 16432087]
5. Evans DM, Frazer IH, Martin NG. Genetic and environmental causes of variation in basal levels of blood cells. *Twin Res*. 1999; 2:250–257. [PubMed: 10723803]
6. Garner C, Tatu T, Reittie JE, et al. Genetic influences on F cells and other hematologic variables: a twin heritability study. *Blood*. 2000; 95:342–346. [PubMed: 10607722]

7. Nalls MA, Couper DJ, Tanaka T, et al. Multiple loci are associated with white blood cell phenotypes. *PLoS Genet.* 2011; 7:e1002113. [PubMed: 21738480]
8. 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; 41:1182–1190. [PubMed: 19820697]
9. Reiner AP, Lettre G, Nalls MA, et al. Genome-wide association study of white blood cell count in 16,388 African Americans: the continental origins and genetic epidemiology network (COGENT). *PLoS Genet.* 2011; 7:e1002108. [PubMed: 21738479]
10. van der Harst P. Seventy-five genetic loci influencing the human red blood cell. *Nature.* 2012 advance online publication.
11. Farber CR, Bennett BJ, Orozco L, et al. Mouse genome-wide association and systems genetics identify *Asx12* as a regulator of bone mineral density and osteoclastogenesis. *PLoS Genet.* 2011; 7:e1002038. [PubMed: 21490954]
12. Peters LL, Shavit JA, Lambert AJ, et al. Sequence variation at multiple loci influences red cell hemoglobin concentration. *Blood.* 2010; 116:e139–149. [PubMed: 20833975]
13. Kelada SN, Aylor DL, Peck BC, et al. Genetic analysis of hematological parameters in incipient lines of the collaborative cross. *G3 (Bethesda).* 2012; 2:157–165. [PubMed: 22384394]
14. Flint J, Valdar W, Shifman S, Mott R. Strategies for mapping and cloning quantitative trait genes in rodents. *Nat Rev Genet.* 2005; 6:271–286. [PubMed: 15803197]
15. Bennett BJ, Farber CR, Orozco L, et al. A high-resolution association mapping panel for the dissection of complex traits in mice. *Genome Res.* 2010; 20:281–290. [PubMed: 20054062]
16. Kang HM, Zaitlen NA, Wade CM, et al. Efficient control of population structure in model organism association mapping. *Genetics.* 2008; 178:1709–1723. [PubMed: 18385116]
17. Davis RC, van Nas A, Castellani LW, et al. Systems genetics of susceptibility to obesity-induced diabetes in mice. *Physiol Genomics.* 2012; 44:1–13. [PubMed: 22010005]
18. Horvat S, Bunger L. Polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP) assay for the mouse leptin receptor (*Lepr(db)*) mutation. *Lab Anim.* 1999; 33:380–384. [PubMed: 10778787]
19. Laurie CC, Nickerson DA, Anderson AD, et al. Linkage disequilibrium in wild mice. *PLoS Genet.* 2007; 3:e144. [PubMed: 17722986]
20. Broman KW, Wu H, Sen S, Churchill GA. R/qtl: QTL mapping in experimental crosses. *Bioinformatics.* 2003; 19:889–890. [PubMed: 12724300]
21. van Nas A, Ingram-Drake L, Sinsheimer JS, et al. Expression quantitative trait loci: replication, tissue- and sex-specificity in mice. *Genetics.* 2010; 185:1059–1068. [PubMed: 20439777]
22. Hedrick CC, Castellani LW, Warden CH, Puppione DL, Lusis AJ. Influence of mouse apolipoprotein A-II on plasma lipoproteins in transgenic mice. *J Biol Chem.* 1993; 268:20676–20682. [PubMed: 8376417]
23. Puppione DL, Charugundla S. A microprecipitation technique suitable for measuring alpha-lipoprotein cholesterol. *Lipids.* 1994; 29:595–597. [PubMed: 7990668]
24. Krzywinski M, Schein J, Birol I, et al. Circos: an information aesthetic for comparative genomics. *Genome Res.* 2009; 19:1639–1645. [PubMed: 19541911]
25. Payseur BA, Place M, Weber JL. Linkage disequilibrium between STRPs and SNPs across the human genome. *Am J Hum Genet.* 2008; 82:1039–1050. [PubMed: 18423524]
26. Orozco LD, Bennett BJ, Farber CR, et al. Unraveling Inflammatory Responses using Systems Genetics and Gene-Environment Interactions in Macrophages. *Cell.* 2012; 151:658–670. [PubMed: 23101632]
27. Ganesh SK, Zakai NA, van Rooij FJ, et al. Multiple loci influence erythrocyte phenotypes in the CHARGE Consortium. *Nat Genet.* 2009; 41:1191–1198. [PubMed: 19862010]
28. Flint J, Eskin E. Genome-wide association studies in mice. *Nat Rev Genet.* 2012; 13:807–817. [PubMed: 23044826]



Linkage & Association of Mean Corpuscular Volume [fl] on Chromosome 7

Figure 1. Blood cell traits among the HMDP strains rank-ordered by strain average
 A. MCV mean corpuscular volume [fl], B. WBC - white blood cell counts [10³/μl]. Red bars reflect strain average for the trait and black bars indicate standard error.

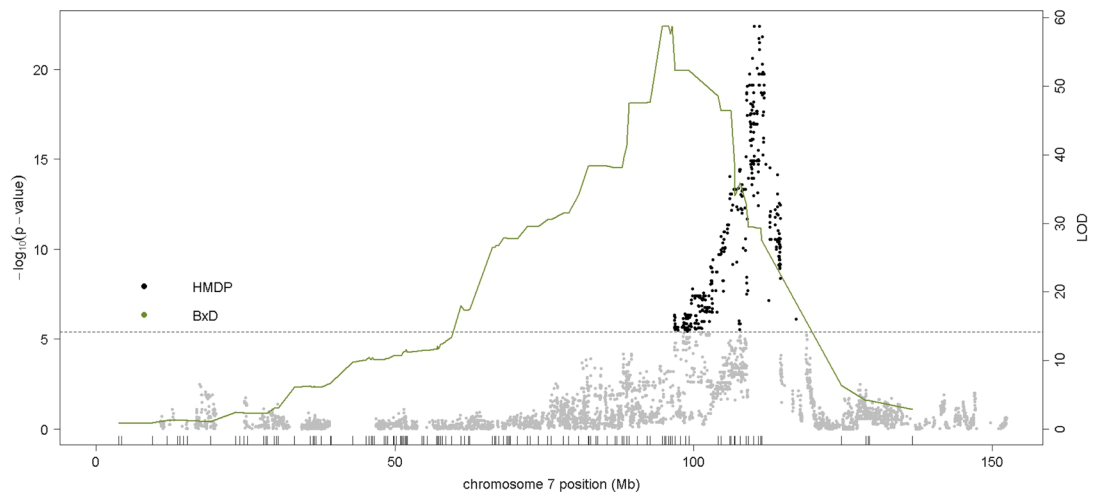
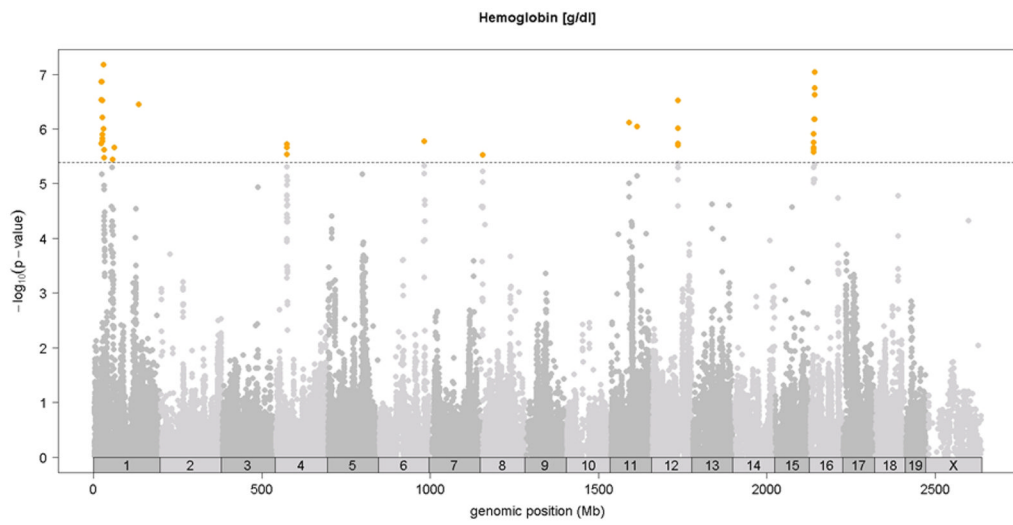
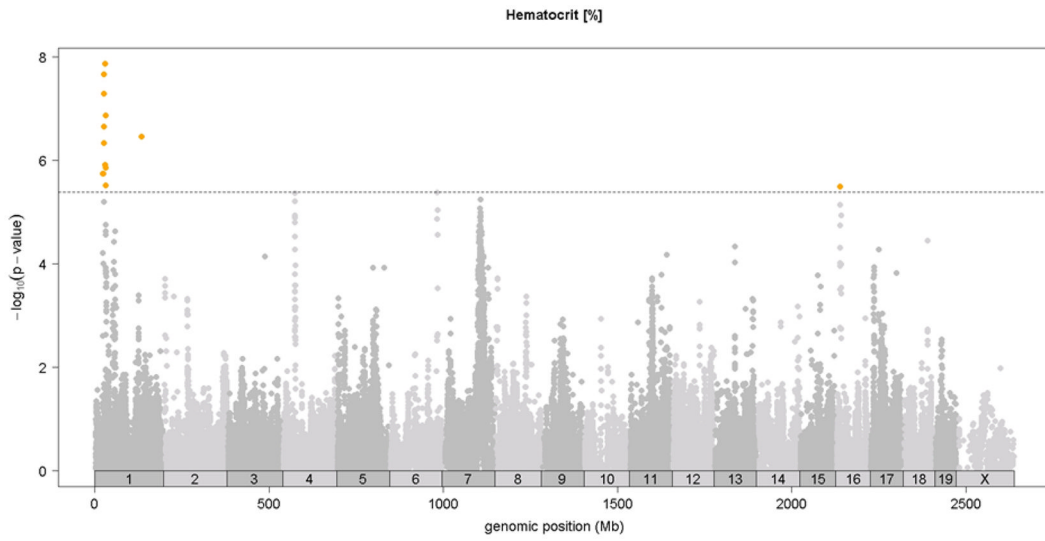
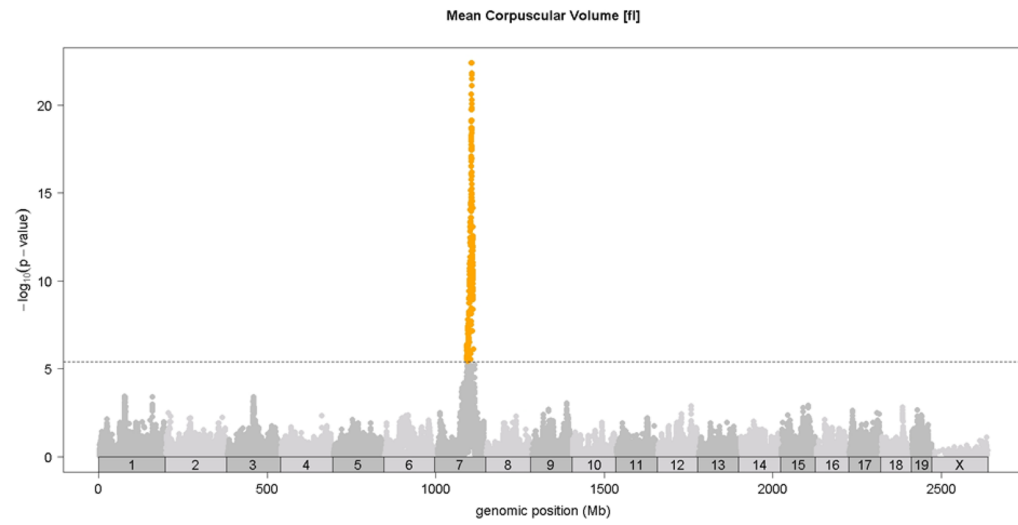


Figure 2. Comparison of linkage and association analysis of red cell MCV locus on chromosome 7
Linkage analysis for MCV in C57BL/6J x DBA/2J F2 mouse cross, shown in the solid green line, is plotted on chromosome 7 with peak LOD = 58.72. Association analysis of MCV is shown as a series of dots indicating the position and significance of association ($-\log_{10}$ p-value) for each SNP. The threshold for genome-wide significance ($p=4.1E-6$) is shown as a dashed line.



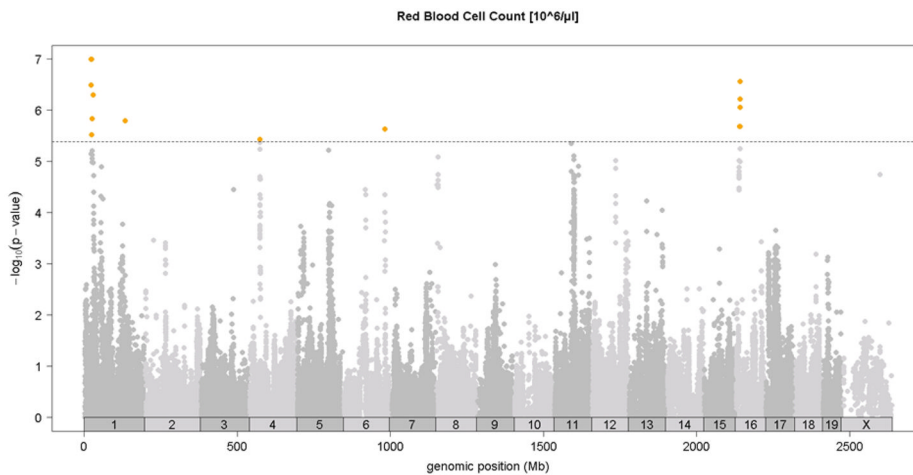
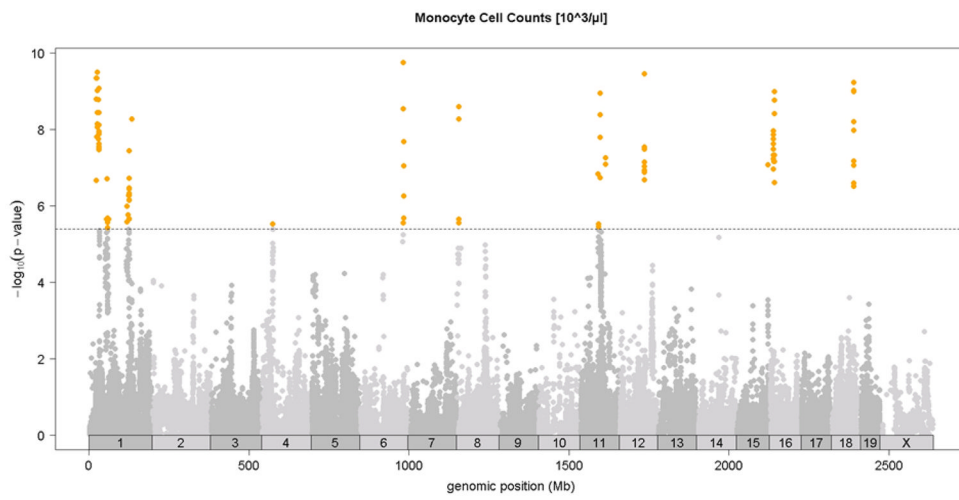
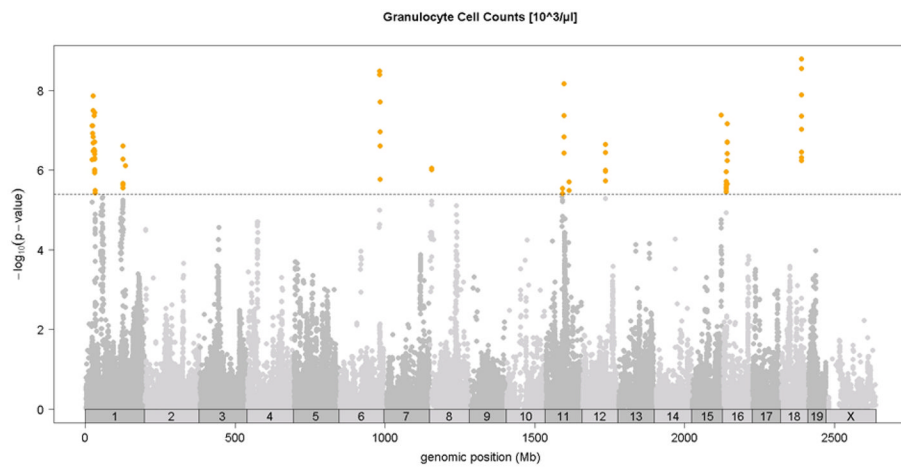
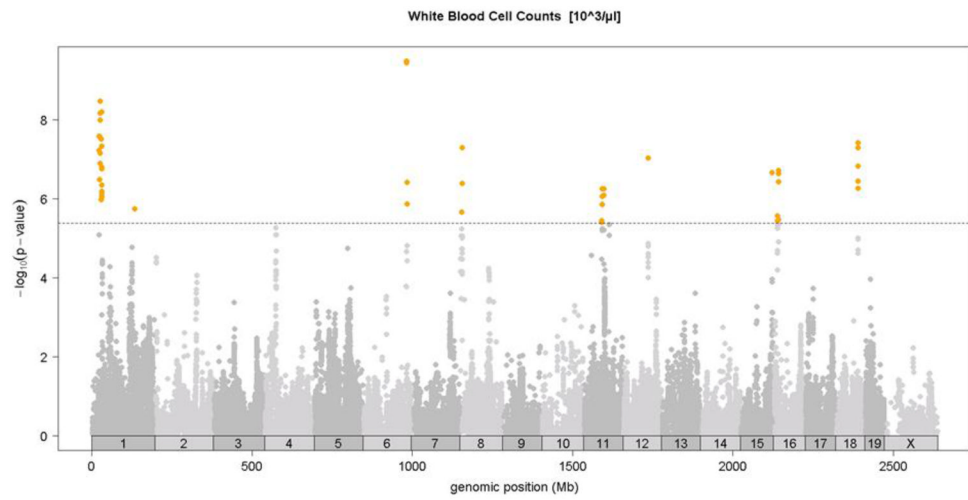


Figure 3. SNP association analysis of red blood cell traits in HMDP population

Shown are the associations for about 107,000 SNPs with genomic position on the x-axis and significance of association ($-\log_{10}$ p-value) on the y-axis. Dashed horizontal line at $p=4.1 \times 10^{-6}$ is threshold for genome-wide significance. A. MCV [fl], B. HCT [%], C. hemoglobin content (HGB) [g/dl], D. red blood cell count (RBC)[$10^6/\mu\text{l}$]. Absolute genomic position (Mb) for each SNP is indicated below the X-axis while chromosome number is shown immediately above.



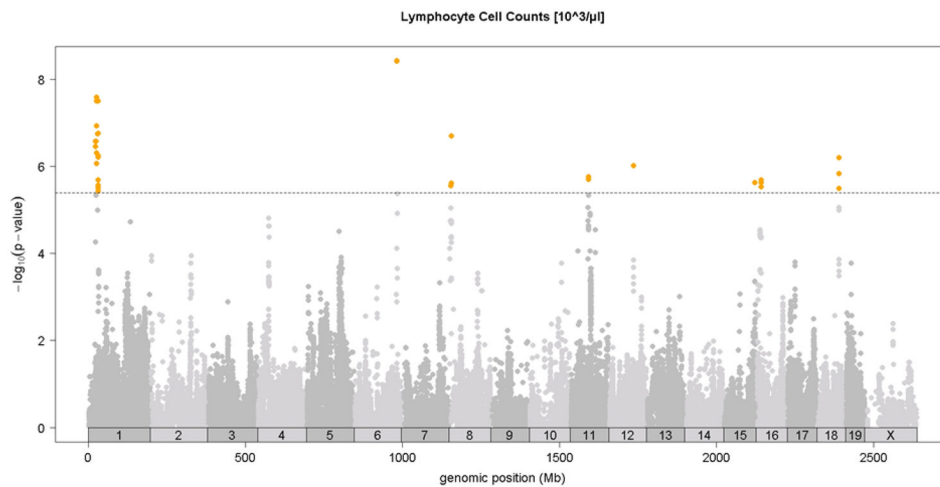


Figure 4. SNP association analysis of white blood cell counts in HMDP population

Shown are the associations for about 107,000 SNPs with genomic position on the x-axis and significance of association ($-\log_{10}$ p-value) on the y-axis. Dashed horizontal line at $p=4.1 \times 10^{-6}$ is threshold for genome-wide significance. Cell counts ($10^3/\mu\text{l}$) for A. total white blood cells (WBC), B. granulocytes (GRAN), C. monocytes (MONO) and, D. lymphocytes (LYMP). Absolute genomic position (Mb) for each SNP is indicated below the X-axis while chromosome number is shown immediately above.

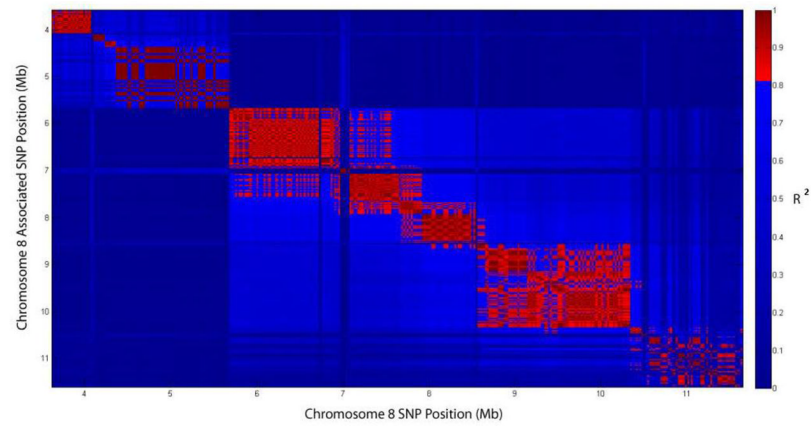


Figure 5. Linkage disequilibrium (LD) structure (visualized as $R^2 > .8$) on chromosome 8 encompassing the locus for monocyte and white blood cell counts
The peak SNP lies at 8.1 Mb in a ~700kb LD block from 7,915,722 to 8,647,388 bp. Horizontal axis is physical position of the SNP (Mb) while vertical axis is position of the associated SNP. Vertical bar on right shows color scale for correlation coefficient.

Table 1
SNPs Associated with Hemoglobin and Red Blood Cell Phenotypes in the HMDP*

chromosome	MCV		MCHC		RDW%		HCT		HGB		RBC	
	position	p-value	position	p-value	position	p-value	position	p-value	position	p-value	position	p-value
1							30,609,106	1.37E-08	29,056,427	6.80E-08	23,872,35	1.02E-07
1							134,905,722	3.51E-07	134,905,722	3.60E-07	134,905,722	1.62E-06
7	110,098,213	4.23E-23	110,098,213	7.83E-24	111,015,676	5.75E-14	110,618,166	5.80E-06				
11									56,841,408	7.72E-07	56,841,408	4.49E-06
12									80,028,884	3.06E-07	80,028,884	9.80E-06
16									15,916,062	9.21E-08	15,916,052	2.78E-07

Abbreviations:

MCV Mean Corpuscular Volume [fl]

MCHC Mean Corpuscular Hemoglobin Concentration [g/dl]

RDW% Red Cell Distribution Width [%]

HCT Hematocrit [%]

HGB Hemoglobin [g/dl]

RBC Red Blood Cell Count [$10^6/\mu\text{l}$]

Table 2

SNPs Associated with White Blood Cell Phenotypes in the HMDP

chromosome	WBC		GRAN		MONO		LYMPH	
	position	p-value	position	p-value	position	p-value	position	p-value
1	26,971,726	3.46E-09	26,294,215	1.39E-08	25,754,577	3.27E-10		
6	135,927,582	3.33E-10	136,211,567	3.27E-09	136,211,567	1.84E-10	136,211,567	3.79E-09
8	8,119,195	5.19E-08			8,119,195	2.57E-09		
11	63,825,134	5.63E-07	63,825,134	6.83E-09	63,825,134	1.15E-09		
12	79,259,640	9.35E-08	79,259,640	2.27E-07	79,259,640	3.55E-10	79,259,640	9.70E-07
15	99,555,171	2.20E-07	99,555,171	4.27E-08	99,555,171	8.48E-08	99,555,171	2.40E-06
16	15,916,062	1.96E-07			15,916,062	1.03E-09		
18	70,410,404	3.83E-08	70,197,956	1.66E-09	70,197,956	5.93E-10	70,410,404	6.38E-07

Abbreviations:

WBC White Blood Cell Counts [$10^3/\mu\text{l}$]

GRAN Granulocyte Cell Counts [$10^3/\mu\text{l}$]

MONO Monocyte Cell Counts [$10^3/\mu\text{l}$]

LYMPH Lymphocyte Cell Counts [$10^3/\mu\text{l}$]

Table 3

Candidate genes with non-synonymous mutations in LD blocks for blood phenotype QTLs

Hemoglobin + RBC			
chromosome	start position	end position	Genes with Nonsynonymous coding mutation or splice-site variations
1	134,905,722	136,420,211	Mdm4, Pik3c2b, Plekha6, Ren1, Etrik2, Sox13, Zc3h11a, Lax1, Atp2b4, Chi311, Mybch, Ppfia4, Cyb5r1, Rabif
7	109,376,676	111,745,069	Hbb-b1, Hbb-b2, Hbb-bh1, Hbb-bh2, Hbb-y and 19 other non-olofactory receptor genes
11	55,901,493	59,445,442	Fam 114a2, Mfap3, Galnt10, Larpl1, Gemin5, Mrpl22, Igtp, Irgm2, Trim58, zfp39, Butr1, Trim17, Trim11, Obscn, Gjc2, Mrpl55, Prss38, Jmjd4, Nlrp3
12	78,704,936	80,544,480	Mpp5, Atp6v1d, Plekhh1, Pigh, Arg2, Vti1b, Zfyve26
16	13,716,171	16,326,370	Pla2g10, pdxdc1, myh11, abcc1, prkdc

White Cell counts			
chromosome	start position	end position	Genes with Nonsynonymous coding mutations or splice-site variations
1	25,822,356	26,780,378	Gm9884,4931408C20Rik
6	135,036,055	139,317,009	gorc5d, hebp1, gsg1, pbp2 atf7ip, plbd1, Gucy2c, Wbp11, Art4, Mgp, Erp27, Ptpro, Eps8, Slc15a5, Mgst1, lgbp1b
8	7,915,722	8,647,388	Efnb2
11	63,581,579	64,044,219	Hs3st3b1
12	78,704,936	80,544,480	Mpp5, Atp6v1d, Plekhh1, Pigh, Arg2, Vti1b, Zfyve26
15	99,439,319	99,832,043	Accn2, Smarcd1, Gpd1, Lass5, Lima1
16	13,716,171	16,326,370	Pla2q10, pdxdc1, myh11, abcc1, prkdc
18	69,977, 676	70,802,044	ccdc68,2310002L13Rik, 4930503L19Rik, Stard6, Poli, Mbd2

Table 4

Macrophage eQTL for genes at loci associated with blood cell traits in the HMDP

gene	chromosome	position	p-value	macrophage	LPS	OxPAPC
Pip4a1	1	26,101,763	3.9663E-06	x		
Emp1	6	135,272,118	6.6586E-12	X	X	
Pipr1	6	135,272,118	2.8697E-11	X	X	X
Mgp	6	135,274,624	2.922E-10	x	X	X
H2afj	6	135,607,348	5.3959E-16	x	X	X
8430419L09Rik	6	135,607,348	8.14E-11			X
Eps8	6	135,701,623	1.3198E-09	X		X
P1bd1	6	136,040,066	6.3669E-14	X	X	X
Gpr19	6	136,040,694	1.5169E-17	X	X	X
Arhgd1b	6	136,040,694	9.92E-09	X		X
Klra2	6	136,073,887	8.63E-07	X		X
Cdkon1b	6	136,401,458	9.99 E-07	X		X
Ddx47	6	136,515,939	5.1766E-07	X	X	
Dusp16	6	136,515,939	1.57E-06	X		X
Reeq1	6	137,667,376	1.67E-06	X		X
Mgst1	6	138,540,068	1.8599E-08	X	X	X
Arg2	12	79,687,143	1.2743E-11	X	X	X
Alp6v1d	12	79,687,143	1.6421E-06	X	X	X
Eif2s1	12	79,687,143	2.0495E-21	X	X	X
Vt11b	12	79,687,143	5.2777E-07	X		X
Comt1	16	13,922,036	1.5192E-12	X	X	X
Litaf	16	14,087,146	5.2865E-08	X	X	X
Pdxdc1	16	14,311,427	1.154E-39	X	X	X
Mld2	16	14,438,272	1.0535E-09	X	X	
Clec16a	16	15,144,252	5.3751E-08	X	X	X
Me2	18	69,977,676	2.8453E-07	X		
Smad4	18	70,571,511	4.3765E-12	X		

gene	chromosome	position	p-value	macrophage	LPS	OxPAPC
Stard6	18	70,627,066	5.6035E-12	X	X	X
Chmp1b	18	70,802,044	3.4346E-13	X	X	X

¹ under basal [medium] condition

² Following LPS treatment

³ Following oxidized phospholipid treatment