Published in final edited form as:

Mamm Genome. 2012 June ; 23(5-6): 356–366. doi:10.1007/s00335-012-9393-3.

QTLs for murine red blood cell parameters in LG/J and SM/J F₂ and advanced intercross lines

Thomas B. Bartnikas¹, Clarissa C. Parker², Riyan Cheng², Dean R. Campagna¹, Jackie E. Lim³, Abraham A. Palmer^{2,4}, and Mark D. Fleming¹

¹Department of Pathology, Children's Hospital, Boston, MA, 02115, USA

²Department of Human Genetics, The University of Chicago, Chicago, IL, 60637 USA

³Department of Pharmacology & Cancer Biology, Duke University Medical Center, Durham, NC, 27708 USA

⁴Department of Psychiatry and Behavioral Neuroscience, The University of Chicago, Chicago, IL, 60637 USA

Abstract

Red blood cells are essential for oxygen transport and other physiologic processes. Red cell characteristics are typically determined by complete blood counts which measure parameters such as hemoglobin levels and mean corpuscular volumes; these parameters reflect the quality and quantity of red cells in the circulation at any particular moment. To identify the genetic determinants of red cell parameters, we performed genome-wide association analysis on LG/J × SM/J F₂ and F₃₄ advanced intercross lines using single nucleotide polymorphism genotyping and a novel algorithm for mapping in the combined populations. We identified significant quantitative trait loci for red cell parameters on chromosomes 6, 7, 8, 10, 12 and 17; our use of advanced intercross lines reduced the quantitative trait loci interval width from 1.6- to 9.4-fold. Using genomic sequences of LG/J and SM/J mice, we identified non-synonymous coding single nucleotide polymorphisms in candidate genes residing within quantitative trait loci and performed sequence alignments and molecular modeling to gauge the potential impact of amino acid substitutions. These results should aid in the identification of genes critical for red cell physiology and metabolism and demonstrate the utility of advanced intercross lines in uncovering genetic determinants of inherited traits.

Introduction

Blood cells perform an integral role in critical physiological processes from oxygen transport to blood clotting and assorted aspects of infection and immunity. Epidemiologic studies indicate that baseline hematopoietic traits are significant, independent risk factors for diseases with considerable morbidity and mortality. For example, total hemoglobin (Hgb) is a risk factor of sickle cell disease severity, with high Hgb associated with pain crises and acute chest syndrome and low Hgb levels associated with higher risk of stroke (Platt et al. 1994; Castro et al. 1994).

Substantial genetic contributions underlie variations in baseline hematopoietic parameters (Garner et al. 2000; Chen and Harrison 2002; Mahaney et al. 2005; Peters et al. 2005, 2006).

In humans, twin studies estimate that the heritable variation contributing to blood cell traits ranges from 40–90% (Whitfield and Martin 1985; Evans et al. 1999; Garner et al. 2000; Lin et al. 2007). Loci underlying this heritable variation can be identified by approaches such as genome-wide association studies (GWAS) in human populations and quantitative trait locus (QTL) studies in model organisms. Complementing human genetic studies, experiments in model organisms offer unique advantages including the ability to control for environmental factors, perform invasive procedures, conduct defined crosses, functionally evaluate candidate genes *in vitro* and *in vivo* and undertake rigorous mechanistic studies.

While QTL studies have identified chromosomal regions associated with blood cell traits in mice, identifying the causal genes is often a difficult task, mainly because studies in mice traditionally use recombinant inbred lines, backcrosses or intercrosses to identify QTLs. Due to a restricted number of recombination events, these panels offer relatively low resolution, identify large genomic regions and are often unsuitable for identifying genes that underlie QTLs (Flint et al. 2005; Peters et al. 2007). This limitation can be addressed by using populations with greater numbers of accumulated recombinations such as advanced intercross lines (AILs), which more closely approximate the amount of recombination found in human populations (Darvasi and Soller 1995). An AIL offers vastly improved mapping resolution while maintaining the desirable properties of polymorphic alleles, a known pedigree, and completely informative markers.

To date, identification of QTLs and candidate genes for red blood cell parameters has not been performed using AILs. In this paper, we report the results of such a study using genotype and complete blood count (CBC) data from F₂ and F₃₄ AILs from a cross between LG/J and SM/J mice. We identified QTLs in discrete regions on multiple chromosomes and identified candidate genes within these regions. Using LG/J and SM/J genomic sequences, we identified non-synonymous coding SNPs within these genes. To gauge the potential physiologic relevance of these SNPs, we performed protein sequence alignments and molecular modeling. Our results identify genes with known and putative roles in red cell physiology and metabolism and demonstrate the utility of AILs in QTL studies.

Materials and methods

Subjects

All procedures were approved by the University of Chicago Institutional Animal Care and Use Committee (IACUC) in accordance with NIH guidelines. Mice were housed and bred as previously described (Parker et al. 2011); rodent chow contained 225 ppm iron. Briefly, 462 F₂ mice (230 females, 232 males) were generated from F₁ mice bred from inbred male SM/J and female LG/J mice (Jackson Laboratories, Bar Harbor, ME). 472 F₃₄ mice (231 females, 241 males) were generated from 119 F₃₃ mice kindly provided by Dr. James Cheverud and of known pedigree back to the original inbred founders; each breeding pair produced only one F₃₄ litter and was rotated after each litter to minimize relatedness. All mice, as previously described (Cheng et al. 2010; Parker et al. 2011), were involved in a behavioral study after which anticoagulated whole blood was harvested by retro-orbital bleed at 13–14.5 weeks and analyzed by complete blood counts within 48 hours at Children's Hospital Boston using an Advia 120 Multispecies whole blood analyzer (Bayer Corporation, Tarrytown, NY).

Genotyping and data analysis

Genotyping was performed as previously described (Cheng et al. 2010; Parker et al. 2011); F₂ and F₃₄ mice were genotyped for 162 and ~4000 evenly-spaced SNPs respectively. Genome-wide association analysis was also performed as previously described (Cheng et al.

2010, 2011; Parker et al. 2011) using the R package QTLRel that accounted for complex relationships among F₃₄ mice. Genome-wide significance thresholds were estimated as previously described (Cheng et al. 2010). Sequences for LG/J and SM/J mice were kindly provided by Dr. Jim Cheverud (Norgard et al. 2011); sequences contained 4,299,800 autosomal polymorphisms between the two strains with 20-fold and 14-fold sequencing coverage for LG/J and SM/J respectively. SNPs were deemed 'previously documented' based on their inclusion in the mouse genome database at Ensembl (www.ensembl.org). Candidate genes in QTLs were identified using mouse genome data (www.ensembl.org), gene expression data (<https://biogps.gnf.org>), and published data (www.ncbi.nlm.nih.gov/pubmed). Genes of interest included known protein coding genes, novel processed transcripts, microRNAs and regulatory regions such as the hemoglobin locus control region. Gene expression patterns of interest included organs/tissues/cells with known involvement in absorption, distribution and recycling of nutrients and factors essential for red cell metabolism or the regulation of these processes and organs/tissues/cells required for hematopoiesis and red cell turnover. Published data was used to exclude candidates if animal models deficient in a gene of interest did not exhibit hematological phenotypes. Genomic sequence analysis was performed using Geneious (Biomatters); candidate genes were not necessarily excluded based on an absence of non-synonymous coding SNPs. Primary protein sequences were obtained from NCBI (www.ncbi.nlm.nih.gov). NCBI BLAST searches using mouse primary sequences identified 8–10 most highly similar sequences. Alignments were performed using ClustalW2 (Chenna et al. 2003). Molecular modeling was performed using NCBI Conserved Domain Search, HHPred (Biegert et al. 2006) and SwissPDB-Viewer (Guex and Peitsch 1997).

Results

We detected QTLs for mean corpuscular hemoglobin levels (MCH), mean corpuscular volumes (MCV), hemoglobin levels (HGB), red blood cell counts (RBC) and mean corpuscular hemoglobin concentrations (MCHC) on chromosomes 6, 7, 8, 10, 12 and 17 (Figs. 1, 2, 4–6; Table 1); several of these QTLs resided in regions identified in previous QTL studies or in regions syntenic to those identified by GWAS in human populations (Peters et al. 2004, 2006; Ganesh et al. 2009; Kamatani et al. 2010). We noted key differences between F₂ and F₂+F₃₄ QTLs: first, a chromosome 6 MCH QTL was noted in the F₂ but not in the F₂+F₃₄ analysis (Fig. 1A); second, a chromosome 6 HGB QTL was noted in the F₂+F₃₄ but not the F₂ analysis (Fig. 1B); third, most of the chromosome 12 MCV QTLs in the F₂+F₃₄ analysis did not reside within the QTL identified in the F₂ analysis (Fig. 5C); finally, the width of the intervals for QTLs decreased by 1.6- to 9.4-fold from the F₂ to the F₂+F₃₄ analysis (Figs. 1, 2, 4–6).

To identify candidate genes within QTLs, we used mouse genome, expression and published data and LG/J and SM/J genomic sequences and first focused on genes with non-synonymous coding SNPs. This produced three genes with known *in vivo* functions relevant to parameters of erythropoiesis, all of which resided on chromosome 7: *Hamp1*, *Hamp2* and *Klf13*. Previously documented SNPs Asn73Lys and Ser77Phe were noted in *Hamp1* and *Hamp2* respectively, two genes encoding isoforms of hepcidin, an essential regulator of iron absorption and recycling; these SNPs affect residues immediately adjacent to highly conserved cysteine residues involved in disulfide bonding (Figure 3A, B) (Fleming 2008). An Ala216Ser SNP, affecting a highly conserved residue in a zinc finger domain, was noted in *Klf13*, a gene encoding a transcription factor required for the regulation of erythropoiesis (Figure 3C, D) (Gordon et al. 2008). While a previously documented non-synonymous coding SNP was found in *Hbb-y*, a gene encoding an embryonic subunit of hemoglobin, this SNP affects a residue which, according to molecular modeling, is not directly involved with

heme binding or subunit-subunit interactions (data not shown); furthermore, deletion of the *Hbb-y* promoter in mice does not alter expression of fetal or adult globins (Hu et al. 2003).

Our analysis also identified several genes harboring non-synonymous coding SNPs but of unknown *in vivo* relevance to processes influencing red cell parameters: *Nfu1*, *Abcb10*, *Fancm*, *Neu1* and *Trim10*. Residing within the chromosome 6 HGB QTL, *Nfu1* encodes a gene implicated in iron sulfur cluster biosynthesis (Liu and Cowan 2007); a Val39Ala SNP affects a residue conserved in all species examined except for *B. taurus* cattle, although the functional significance of the N-terminus of this protein is unclear (Fig. 1C). Residing within the chromosome 8 MCH and RBC QTLs, *Abcb10* encodes an ATP-binding cassette transporter that interacts with mitoferrin-1, a mitochondrial membrane protein and iron transporter required for heme and iron-sulfur cluster biosynthesis (Chen et al. 2009); a Cys79Ser SNP affects a residue conserved within most sequences analyzed and lies within the mitochondrial targeting sequence of the preprotein (Fig. 4C) (Graf et al. 2004). Residing within the chromosome 12 MCV QTL, *Fancm* is the murine homolog of human *FANCM*, a gene mutated in Fanconi anemia, a disease of bone marrow failure and cancer predisposition due to defective DNA repair (Fig. 5C); although hematologic analysis of *Fancm*-deficient mice has not been reported (Bakker et al. 2009), we observed multiple novel and previously reported non-synonymous coding SNPs in *Fancm*, including Tyr50Phe, Lys655Asn, Phe797Cys*, His816Arg*, Pro879Arg, Pro1040Leu*, Ile1987Leu*, Gln2005Leu* and Lys2018Thr*, where '*' denotes a previously reported SNP.

Novel and previously documented non-synonymous coding region SNPs were also detected in *Neu1* and *Trim10*, two genes residing within the chromosome 17 MCH and HGB QTL (Fig. 6A, B). A Leu290Ile SNP affects a highly conserved residue in neuraminidase 1 or *Neu1* and has been suggested, along with a -519G>A promoter SNP also noted in our sequence analysis, to underlie the neuraminidase deficiency documented in SM/J mice (Fig. 5C; data not shown) (Rottier et al. 1998; Champigny et al. 2009); notably, the neuraminidase deficiency in SM/J mice has been linked to lipid accumulation in these mice in a variety of cell types including Kupffer cells, liver-resident macrophages that degrade senescent red blood cells (Champigny et al. 2009). We also detected novel and previously documented SNPs in *Trim10*, encoding Tripartite motif containing 10, a protein required for terminal differentiation of erythroid cells *in vitro* (Harada et al. 1999). An Arg133His SNP affects a residue within a zinc finger motif while Arg231His, His365Tyr and Met386Val SNPs, all previously documented, affect residues in regions of unknown function; all four of these SNPs affect strongly conserved residues (data not shown).

Finally, our analysis identified genes of potential relevance to parameters of erythropoiesis based largely on their expression patterns; these genes did not harbor non-synonymous coding SNPs nor has their relevance to processes such as erythropoiesis or iron metabolism been tested *in vitro* or *in vivo*. These genes include *Snca* and *March8*; their potential relevance will be discussed below.

Discussion

LG/J and SM/J mice have been used effectively in studies on a variety of physiologic and pathophysiologic processes including body weight, diabetes, lipid metabolism, wound healing and obesity (Ehrich et al. 2005; Blankenhorn et al. 2009; Lawson et al. 2010, 2011; Parker et al. 2011). To our knowledge, our study represents the first to use these strains and the first to use AILs to investigate genetic determinants of red blood cell parameters. We took advantage of an F₂ population of mice used for behavioral studies and mice bred from AILs kindly provided by Dr. Cheverud and phenotyped these mice with CBCs to measure several quantitative traits representative of red cell physiology and metabolism. We

analyzed data from F₂, F₃₄ and F₂+F₃₄ populations, the latter of which permitted us to exploit the detection power of an F₂ study and the mapping power of an AIL study. These strains and mice were also chosen given the known parental genotypes and genomic sequences and pedigrees of F₃₃ mice—this information permitted us to account for relatedness in our statistical analyses and identify non-synonymous coding SNPs in genes of interest.

LG/J and SM/J mice harbor known defects in specific physiologic processes. LG/J mice carry SNPs in *Ahr* (chromosome 12, 36 Mb, ~13 cM) encoding an aryl-hydrocarbon receptor and *Cdh23* (chromosome 10, 59–60 Mb, ~24–27 cM) encoding cadherin 23. While these two genes do lie within the range of QTLs identified in our study, they have no known role in determining red cell parameters measured by CBCs. Although we did not detect any QTLs on chromosome 14, SM/J mice carry a 5 base pair deletion in *Il3ra* (chromosome 14, 15 Mb) encoding an interleukin 3 receptor alpha chain which has been associated with impaired response of hematopoietic colony formation to interleukin 3 treatment (Ichihara et al. 1995). SM/J mice also carry coding and promoter SNPs in *Neu1* (chromosome 17, 35 Mb, 17 cM) encoding neuraminidase 1, which, as described above, are believed to underlie the neuraminidase deficiency observed in the strain. The lipid accumulation observed in SM/J mice, attributed to the neuraminidase deficiency, affects multiple cell types including liver-resident macrophages known as Kupffer cells (Champigny et al. 2009). Whether this lipid accumulation alters the efficacy of Kupffer cell-dependent red cell turnover or iron storage is not known. Anemia has been reported in a patient with a *Neu1* deficiency, also known as sialidosis type II, who presented at birth with hydrops, hepatomegaly and respiratory distress syndrome (Sergi et al. 2001).

As our analysis did not focus solely on genes with known roles in red cell physiology, we identified several candidate genes with expression patterns of interest but without any previously documented roles in red cell physiology or metabolism or any detected non-synonymous coding SNPs. As the sequence determinants of gene expression are much less well-defined than are the coding regions for most genes, we were unable to state that a non-coding SNP did not impact expression levels; an expression QTL (eQTL)-based study could address this shortcoming. Without expression data, the relevance of non-synonymous coding SNPs must be interpreted cautiously; for complex traits like MCH, HGB, RBC, MCV and MCHC, differences in gene expression levels or splicing patterns may play a more significant role in determining phenotype than do non-conservative amino acid substitutions in primary protein sequences. Nevertheless, two candidate genes with noteworthy expression patterns were *Snca* and *March2*. *SNCA* encodes α -synuclein, a component of intracellular inclusions present in neurons of Parkinson's disease patients (Crosiers et al. 2011), and is mutated in monogenic forms of this neurologic disease. While Parkinson's disease is not typically associated with hematologic manifestations, recent epidemiologic studies suggest that anemia is a feature of the preclinical period preceding the onset of the traditional motor, cognitive and behavioral characteristics (Savica et al. 2010). Notably, the α -synuclein is most highly expressed in early erythroid precursors in humans and in bone and bone marrow in mice (BioGPS); the protein is abundant in red blood cells (Barbour et al. 2008). Similar to α -synuclein, *March2* is highly expressed in bone marrow in mice and *MARCH2* in early erythroid precursors in humans while the protein functions in endosomal trafficking (Nakamura et al. 2005). *MARCH2* exhibits 20% identity and 39% similarity to *MARCH8* (data not shown), an E3 ubiquitin ligase that functions in clathrin-independent endocytosis (Eyster et al. 2011); SNPs in and near *MARCH8* associate with variation in MCV in human populations (Ganesh et al. 2009; Kamatani et al. 2010) while *March8* was identified as a candidate gene for an MCHC QTL in mice (Peters et al. 2010).

Our results demonstrate several advantages of using AILs over F_2 populations in mapping studies. As we observed in this study, widths of QTL intervals, and therefore the numbers of candidate genes per QTL, are greatly reduced from the F_2 to F_2+F_{34} analyses. Furthermore, broad individual peaks in our F_2 analyses resolved in several instances to multiple distinct peaks in the F_2+F_{34} analyses, thereby addressing the potential issue with F_2 studies that a single F_2 QTL represents multiple neighboring QTLs (Parker and Palmer 2011). (Notably, several QTL in this study were detected in one but not both crosses, as previously observed (Cheng et al. 2010)—this may reflect a lack of power to detect QTLs in both populations, genotyping or phenotyping errors, a cluster of many small QTLs segregating together in the F_2 population but disaggregating in the F_{34} population below detection threshold or neighboring QTLs with opposing effects in the F_{34} population merging together in the F_2 analysis.) The improved mapping resolution in our F_2+F_{34} analysis stems in part from the large number of recombinations in the F_{34} generation—which significantly degrades linkage disequilibrium present in the F_2 population—but also from our thorough genotyping of polymorphic markers at an average distance of ~5 Mb between markers. AILs also possess the advantage of balanced allele frequencies, as any marker polymorphic in the original inbred parental strains will remain polymorphic in advanced generations assuming a large population size (Parker and Palmer 2011), and equivalence between identity-by-state and identity-by-descent—as AIL mice are generated from inbred strains or some variant thereof, alleles must be shared in AIL mice by descent.

In this study, we identified several QTLs for red blood cell parameters and highlighted candidate genes within these QTLs using expression data and published reports. Using sequence data, alignments and molecular modeling, we refined the list of candidate genes and identified genes with known and potential roles in red cell metabolism. The use of AILs greatly enhanced our ability to narrow the size of QTLs from broad chromosomal regions to regions spanning 5-10 cMs. These results should guide future identification of genetic determinants of red cell homeostasis and demonstrate the utility of AILs in such studies.

Acknowledgments

The authors would like to thank Kaitlin Samocha, Sherika Blevins, Wen Chen, Barry Paw, Rakina Yaneva, Peter Cresswell, Lihua Huang, Chester Brown, Corey Hoehn and Carlo Brugnara for assistance and advice and James Cheverud for providing AIL lines and LG/J and SM/J genomic sequence data. This work was supported by NIH grants R01DA021336, R21DA024845, R01MH079103 and a Grant from the Schweppe Foundation to AAP, NIH grant T32DA007255 to CCP, NIH grant K99DK084122 to TBB, NIH grant R01DK080011 to MDF and NIH T32MH020065 to JEL.

References

- Bakker ST, van de Vrugt HJ, Rooimans MA, Oostra AB, Steltenpool J, Delzenne-Goette E, van der Wal A, van der Valk M, Joenje H, te Riele H, de Winter JP. Fancm-deficient mice reveal unique features of Fanconi anemia complementation group M. *Hum Mol Genet.* 2009; 18:3484–3495. [PubMed: 19561169]
- Barbour R, Kling K, Anderson JP, Banducci K, Cole T, Diep L, Fox M, Goldstein JM, Soriano F, Seubert P, Chilcote TJ. Red blood cells are the major source of alpha-synuclein in blood. *Neurodegener Dis.* 2008; 5:55–59. [PubMed: 18182779]
- Biegert A, Mayer C, Remmert M, Söding J, Lupas AN. The MPI Bioinformatics Toolkit for protein sequence analysis. *Nucleic Acids Res.* 2006; 34:W335–339. [PubMed: 16845021]
- Blankenhorn EP, Bryan G, Kossenkova AV, Clark LD, Zhang X-M, Chang C, Horng W, Pletscher LS, Cheverud JM, Showe LC, Heber-Katz E. Genetic loci that regulate healing and regeneration in LG/J and SM/J mice. *Mamm Genome.* 2009; 20:720–733. [PubMed: 19760323]
- Castro O, Brambilla DJ, Thorington B, Reindorf CA, Scott RB, Gillette P, Vera JC, Levy PS. The acute chest syndrome in sickle cell disease: incidence and risk factors. The Cooperative Study of Sickle Cell Disease. *Blood.* 1994; 84:643–649. [PubMed: 7517723]

- Champigny MJ, Mitchell M, Fox-Robichaud A, Trigatti BL, Igdoura SA. A point mutation in the neu1 promoter recruits an ectopic repressor, Nkx3.2 and results in a mouse model of sialidase deficiency. *Mol Genet Metab*. 2009; 97:43–52. [PubMed: 19217813]
- Chen J, Harrison DE. Quantitative trait loci regulating relative lymphocyte proportions in mouse peripheral blood. *Blood*. 2002; 99:561–566. [PubMed: 11781239]
- Chen W, Paradkar PN, Li L, Pierce EL, Langer NB, Takahashi-Makise N, Hyde BB, Shirihai OS, Ward DM, Kaplan J, Paw BH. Abcb10 physically interacts with mitoferrin-1 (Slc25a37) to enhance its stability and function in the erythroid mitochondria. *Proc Natl Acad Sci USA*. 2009; 106:16263–16268. [PubMed: 19805291]
- Cheng R, Abney M, Palmer AA, Skol AD. QTLRel: an R Package for Genome-wide Association Studies in which Relatedness is a Concern. *BMC Genet*. 2011; 12:66. [PubMed: 21794153]
- Cheng R, Lim JE, Samocha KE, Sokoloff G, Abney M, Skol AD, Palmer AA. Genome-wide association studies and the problem of relatedness among advanced intercross lines and other highly recombinant populations. *Genetics*. 2010; 185:1033–1044. [PubMed: 20439773]
- Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD. Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res*. 2003; 31:3497–3500. [PubMed: 12824352]
- Crosiers D, Theuns J, Cras P, Van Broeckhoven C. Parkinson disease: Insights in clinical, genetic and pathological features of monogenic disease subtypes. *J Chem Neuroanat*. 2011; 42:131–141. [PubMed: 21810464]
- Darvasi A, Soller M. Advanced intercross lines, an experimental population for fine genetic mapping. *Genetics*. 1995; 141:1199–1207. [PubMed: 8582624]
- Ehrich TH, Hrbek T, Kenney-Hunt JP, Pletscher LS, Wang B, Semenkovich CF, Cheverud JM. Fine-mapping gene-by-diet interactions on chromosome 13 in a LG/J × SM/J murine model of obesity. *Diabetes*. 2005; 54:1863–1872. [PubMed: 15919810]
- 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]
- Eyster CA, Cole NB, Petersen S, Viswanathan K, Früh K, Donaldson JG. MARCH Ubiquitin Ligases Alter the Itinerary of Clathrin-independent Cargo from Recycling to Degradation. *Mol Biol Cell*. 2011; 22:3218–3230. [PubMed: 21757542]
- Fleming MD. The regulation of hepcidin and its effects on systemic and cellular iron metabolism. *Hematology Am Soc Hematol Educ Program*. 2008:151–158. [PubMed: 19074074]
- 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]
- Ganesh SK, Zakai NA, van Rooij FJA, Soranzo N, Smith AV, et al. Multiple loci influence erythrocyte phenotypes in the CHARGE Consortium. *Nat Genet*. 2009; 41:1191–1198. [PubMed: 19862010]
- Garner C, Tatu T, Reittie JE, Littlewood T, Darley J, Cervino S, Farrall M, Kelly P, Spector TD, Thein SL. Genetic influences on F cells and other hematologic variables: a twin heritability study. *Blood*. 2000; 95:342–346. [PubMed: 10607722]
- Gordon AR, Outram SV, Keramatipour M, Goddard CA, Colledge WH, Metcalfe JC, Hager-Theodorides AL, Crompton T, Kemp PR. Splenomegaly and modified erythropoiesis in KLF13^{-/-} mice. *J Biol Chem*. 2008; 283:11897–11904. [PubMed: 18285334]
- Graf SA, Haigh SE, Corson ED, Shirihai OS. Targeting, import, and dimerization of a mammalian mitochondrial ATP binding cassette (ABC) transporter, ABCB10 (ABC-me). *J Biol Chem*. 2004; 279:42954–42963. [PubMed: 15215243]
- Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis*. 1997; 18:2714–2723. [PubMed: 9504803]
- Harada H, Harada Y, O'Brien DP, Rice DS, Naeve CW, Downing JR. HERF1, a novel hematopoiesis-specific RING finger protein, is required for terminal differentiation of erythroid cells. *Mol Cell Biol*. 1999; 19:3808–3815. [PubMed: 10207104]
- Hu X, Bulger M, Roach JN, Eszterhas SK, Olivier E, Bouhassira EE, Groudine MT, Fiering S. Promoters of the murine embryonic beta-like globin genes Ey and betah1 do not compete for interaction with the beta-globin locus control region. *Proc Natl Acad Sci USA*. 2003; 100:1111–1115. [PubMed: 12525692]

- Ichihara M, Hara T, Takagi M, Cho LC, Gorman DM, Miyajima A. Impaired interleukin-3 (IL-3) response of the A/J mouse is caused by a branch point deletion in the IL-3 receptor alpha subunit gene. *EMBO J*. 1995; 14:939–950. [PubMed: 7889941]
- Kamatani Y, Matsuda K, Okada Y, Kubo M, Hosono N, Daigo Y, Nakamura Y, Kamatani N. Genome-wide association study of hematological and biochemical traits in a Japanese population. *Nat Genet*. 2010; 42:210–215. [PubMed: 20139978]
- Lawson HA, Lee A, Fawcett GL, Wang B, Pletscher LS, Maxwell TJ, Ehrich TH, Kenney-Hunt JP, Wolf JB, Semenkovich CF, Cheverud JM. The importance of context to the genetic architecture of diabetes-related traits is revealed in a genome-wide scan of a LG/J $\times$ SM/J murine model. *Mamm Genome*. 2011; 22:197–208. [PubMed: 21210123]
- Lawson HA, Zelle KM, Fawcett GL, Wang B, Pletscher LS, Maxwell TJ, Ehrich TH, Kenney-Hunt JP, Wolf JB, Semenkovich CF, Cheverud JM. Genetic, epigenetic, and gene-by-diet interaction effects underlie variation in serum lipids in a LG/J $\times$ SM/J murine model. *J Lipid Res*. 2010; 51:2976–2984. [PubMed: 20601649]
- Lin J-P, O'Donnell CJ, Jin L, Fox C, Yang Q, Cupples LA. Evidence for linkage of red blood cell size and count: genome-wide scans in the Framingham Heart Study. *Am J Hematol*. 2007; 82:605–610. [PubMed: 17211848]
- Liu Y, Cowan JA. Iron sulfur cluster biosynthesis. Human NFU mediates sulfide delivery to ISU in the final step of [2Fe-2S] cluster assembly. *Chem Commun (Camb)*. 2007:3192–3194. [PubMed: 17653385]
- Mahaney MC, Brugnara C, Lease LR, Platt OS. Genetic influences on peripheral blood cell counts: a study in baboons. *Blood*. 2005; 106:1210–1214. [PubMed: 15870178]
- Nakamura N, Fukuda H, Kato A, Hirose S. MARCH-II is a syntaxin-6-binding protein involved in endosomal trafficking. *Mol Biol Cell*. 2005; 16:1696–1710. [PubMed: 15689499]
- Norgard EA, Lawson HA, Pletscher LS, Wang B, Brooks VR, Wolf JB, Cheverud JM. Genetic factors and diet affect long-bone length in the F34 LG,SM advanced intercross. *Mamm Genome*. 2011; 22:178–196. [PubMed: 21170743]
- Parker CC, Palmer AA. Dark matter: are mice the solution to missing heritability? *Front Gene*. 2011; 2:32.
- Parker CC, Cheng R, Sokoloff G, Lim JE, Skol AD, Abney M, Palmer AA. Fine-mapping alleles for body weight in LG/J $\times$ SM/J F(2) and F(34) advanced intercross lines. *Mamm Genome*. 2011; 22:563–571. [PubMed: 21761260]
- Peters LL, Lambert AJ, Zhang W, Churchill GA, Brugnara C, Platt OS. Quantitative trait loci for baseline erythroid traits. *Mamm Genome*. 2006; 17:298–309. [PubMed: 16596451]
- Peters LL, Robledo RF, Bult CJ, Churchill GA, Paigen BJ, Svenson KL. The mouse as a model for human biology: a resource guide for complex trait analysis. *Nat Rev Genet*. 2007; 8:58–69. [PubMed: 17173058]
- Peters LL, Shavit JA, Lambert AJ, Tsaih S-W, Li Q, Su Z, Leduc MS, Paigen B, Churchill GA, Ginsburg D, Brugnara C. Sequence variation at multiple loci influences red cell hemoglobin concentration. *Blood*. 2010; 116:e139–149. [PubMed: 20833975]
- Peters LL, Swearingen RA, Andersen SG, Gwynn B, Lambert AJ, Li R, Lux SE, Churchill GA. Identification of quantitative trait loci that modify the severity of hereditary spherocytosis in wan, a new mouse model of band-3 deficiency. *Blood*. 2004; 103:3233–3240. [PubMed: 15070709]
- Peters LL, Zhang W, Lambert AJ, Brugnara C, Churchill GA, Platt OS. Quantitative trait loci for baseline white blood cell count, platelet count, and mean platelet volume. *Mamm Genome*. 2005; 16:749–763. [PubMed: 16261417]
- Platt OS, Brambilla DJ, Rosse WF, Milner PF, Castro O, Steinberg MH, Klug PP. Mortality in sickle cell disease. Life expectancy and risk factors for early death. *N Engl J Med*. 1994; 330:1639–1644. [PubMed: 7993409]
- Rottier RJ, Bonten E, d' Azzo A. A point mutation in the neu-1 locus causes the neuraminidase defect in the SM/J mouse. *Hum Mol Genet*. 1998; 7:313–321. [PubMed: 9425240]
- Savica R, Rocca WA, Ahlskog JE. When does Parkinson disease start? *Arch Neurol*. 2010; 67:798–801. [PubMed: 20625084]

- Sergi C, Penzel R, Uhl J, Zoubaa S, Dietrich H, Decker N, Rieger P, Kopitz J, Otto HF, Kiessling M, Cantz M. Prenatal diagnosis and fetal pathology in a Turkish family harboring a novel nonsense mutation in the lysosomal alpha-N-acetyl-neuraminidase (sialidase) gene. *Hum Genet.* 2001; 109:421–428. [PubMed: 11702224]
- Whitfield JB, Martin NG. Genetic and environmental influences on the size and number of cells in the blood. *Genet Epidemiol.* 1985; 2:133–144. [PubMed: 4054596]

Figure 1. LOD scores for MCH and HGB for chromosome 6 and analysis of candidate protein Nfu1

LOD scores for mean corpuscular hemoglobin (MCH) levels (A) and red blood cell (HGB) counts (B) were plotted versus chromosome 6 genetic position in cM for F₂ (dark gray line), F₃₄ (light gray line) and F₂+F₃₄ (black line) analyses. Dashed horizontal lines indicates genome-wide significance thresholds (P<0.05) for F₂+F₃₄ (black line) and F₂ (gray line) analyses. Locations of candidate genes and sites syntenic to region identified by GWAS (Kamatani et al. 2010) are indicated by arrows. Bars above graphs indicate chromosomal regions where F₂+F₃₄ (black bar) and F₂ (dark gray bar) LOD scores exceed respective genome-wide significance thresholds; text within black bar indicates fold decrease in cumulative width of F₂ to F₂+F₃₄ bars. (C) Alignment of Nfu1 primary sequences was constructed using ClustalW using murine sequence and 8–10 highly similar sequences identified by BLAST searches. Identical and similar residues are highlighted in dark and light grey respectively; residue affected by SNP in LG/J vs. SM/J genome sequences is indicated by arrow. Species and accession numbers include *Mus musculus* (M.mus) NP_001164062.1, *Canis familiaris* (C.fam) XP_855433.1, *Ailuropoda melanoleuca* (A.mel) XP_002914967.1, *Equus caballus* (E.cab) XP_001491099.1, *Oryctolagus cuniculus* (O.cun) XP_002709869.1, *Bos taurus* (B.tau) NP_001040031.1, *Pan troglodytes* (P.tro) XP_525775.1, *Homo sapiens* (H.sap) AAQ73784.1, *Nomascus leucogenys* (N.leu) XP_003262545.1 and *Rattus norvegicus* (R.nor) NP_001100076.2. Full length of mouse primary sequence is indicated in parentheses next to protein name.

Figure 2. LOD scores for HGB, MCHC, MCV and RBC for chromosome 7
 LOD scores for hemoglobin levels (HGB) (A), mean corpuscular hemoglobin concentrations (MCHC) (B), mean corpuscular volumes (MCV) (C) and red blood cell counts (RBC) (D) were plotted versus chromosome 7 genetic position in cM, as in Fig. 1. Locations of candidate genes and published QTL (Peters et al. 2006, 2010) are indicated by arrows.

Figure 3. Analysis of candidate proteins for chromosome 7 QTL

Primary sequence alignments for Hamp1 and 2 (A) and Klf13 (C) were constructed and depicted as in Fig. 1; asterisks indicate previously documented SNPs in Ensembl. Molecular models of Hamp1 and 2 (B) and Klf13 (D) were constructed using HHPred and SwissPDBViewer. 'N' and 'C' denote respective amino and carboxyl termini. Residues affected by SNPs are labeled and shown by light grey space-filling. Dark grey ball-and-stick denotes residues of interest, specifically those involved in disulfide bonding in hepcidin (B) and zinc ion coordination in a Klf13 zinc finger (D). Species and accession numbers include *Mus musculus* (M.mus) NP_067341.2, NP_115930.1 and NP_899080.1, *Rattus norvegicus* (R.nor) EDM08379.1, *Bos taurus* (B.tau) NP_001077002.1, *Sus scrofa* (S.scr) ACT56535.1, *Pan troglodytes* (P.tro) XP_510269.2 and NP_001103163.1, *Callithrix jacchus* (C.jac) XP_002749103.1, *Homo sapiens* (H.sap) AF131292_1 and NP_057079.2, *Monodelphis domestica* (M.dom) XP_001365529.1, *Meleagris gallopavo* (M.gal) XP_003209439.1, *Gallus gallus* (G.gal) XP_425065.1, *Macaca mulatta* (M.mul) XP_001094273.1, *Macaca fuscata* (M.fus) ABU75217.1, *Trachypithecus obscurus* (T.obs) ABU75214.1, *Pongo pygmaeus* (P.pyg) ABU75212.1, *Pongo abelii* (P.abi) NP_001127676.1, *Hylobates lar* (H.lar) ABU75224.1 and *Ovis aries* (O.ari) NP_001182241.1.

Figure 4. LOD scores for MCH and RBC for chromosome 8 and analysis of candidate protein Abcb10

LOD scores for mean corpuscular hemoglobin (MCH) levels (A) and red blood cell (RBC) counts (B) were plotted versus chromosome 8 genetic position in cM, as in Fig. 1. Locations of candidate genes and site syntenic to region identified by GWAS (Ganesh et al. 2009) are indicated by arrows. (C) Alignment of Abcb10 primary sequences was constructed and depicted as in Fig. 1. Species and accession numbers include *Mus musculus* (M.mus) NP_062425.1, *Rattus norvegicus* (R.nor) NP_001012166.1, *Homo sapiens* (H.sap) NP_036221.2, *Pongo abei* (P.abe) XP_002809369.1, *Macaca mulatta* (M.mul) XP_001082734.1, *Oryctolagus cuniculus* (O.cun) XP_002717342.1, *Bos taurus* (B.tau) XP_001256449.1, *Sus scrofa* (S.scr) XP_001925414.1 and *Monodelphis domestica* (M.dom) XP_001379107.1.

Figure 5. LOD scores for MCH for chromosome 10 and for MCH and MCV for chromosome 12
 LOD scores for mean corpuscular hemoglobin (MCH) levels for chromosome 10 (A) and for MCH and mean corpuscular volumes (MCV) for chromosome 12 (B, C) were plotted versus chromosome 10 and 12 genetic position in cM, as in Fig. 1. Locations of candidate genes, published QTLs and sites syntenic to region identified by GWAS (Peters et al. 2004, 2006; Ganesh et al. 2009; Kamatani et al. 2010) are indicated by arrows.

Figure 6. LOD scores for MCH and HGB for chromosome 17 and analysis of candidate protein Neu1

(A, B) LOD scores for mean corpuscular hemoglobin (MCH) (A) and hemoglobin (HGB) (B) levels were plotted versus chromosome 17 genetic position in cM, as in Fig. 1. Location of candidate genes and site syntenic to regions identified by GWAS (Ganesh et al. 2009; Kamatani et al. 2010) are indicated by arrows and brackets. (C) Alignment of Neu1 primary sequences was constructed and depicted as in Fig. 1. Species and accession numbers include *Mus musculus* (M.mus) NP_035023.3, *Homo sapiens* (H.sap) NP_000425.1, *Nomascus leucogenys* (N.leu) XP_003272176.1, *Pongo abelii* (P.abe) CAH89635.1, *Macaca mulatta* (M.mul) XP_001113496.1, *Callithrix jacchus* (C.jac) XP_002746410.1, *Ailuropoda melanoleuca* (A.mel) XP_002931026.1, *Oryctolagus cuniculus* (O.cun) XP_002714358.1, *Rattus norvegicus* (R.nor) NP_113710.2 and *Cricetulus griseus* (C.gri) ACL52160.1.

Table 1

QTL interval lengths from F2 and advanced intercross lines.

Chr, trait	F2			F2+F34		
	Interval			Interval		
	cM	Mb	LOD	cM	Mb	LOD
6 MCH	27.7–80.5*	49.1–147.4*	5.0	30.3–35.9	54.2–73.4	6.1
6 HGB				41.1–42.5	84.0–89.1	4.9
7 HGB	2.0–62.2	13.5–123.7	7.1	11.4–14.6	30.1–30.2	5.3
				20.7–22.2	37.4–40.2	5.5
				23.8–30.4	42.3–56.4	6.9
				40.5–40.7	78.5–78.8	4.6
				48.1–50.5	99.2–107.4	7.2
7 MCH	19.2–62.2	35.8–123.7	7.3	45.8–51.7	92.7–110.7	12.4
7 MCV	23.5–54.1	42.2–115.2	5.1	27.9–33.8	44.9–68.3	7.6
				45.6–47.1	92.1–98.4	6.3
				48.1–51.7	99.2–110.7	5.8
7 RBC	9.1–48.1	27.4–101.6	5.5	25.4–35.4	44.9–69.9	69.9
8 MCH	59.5–77.8*	111.1–132.0*	7.4	67.4–77.8*	120.1–132.0*	11.3
8 RBC	31.4–77.8*	48.7–132.0*	6.4	69.0–77.8*	120.1–132.0*	8.5
10 MCH	12.4–71.2*	28.4–129.0*	5.6	41.3–42.0	88.6–90.4	5.0
				43.9–44.4	92.4–93.5	4.6
				46.3–49.9	95.0–104.1	8.0
				52.4–53.1	106.9–110.6	5.4
				55.6–56.3	111.7–114.2	4.6
				62.5–68.3	118.9–123.6	6.4
12 MCH	0*–15.8	0*–40.9	4.2	0*–1.6	4.2–8.9	6.4
				2.0–4.5	9.8–12.9	5.5
				5.1–5.6	15.8–17.0	5.3
				6.7–8.8	25.3–29.9	6.1
12 MCV	12.7–25.6	35.0–68.8	4.9	23.2–24.1	59.8–61.1	5.5
				25.2–26.5	67.2–71.3	5.4
				28.1–32.3	72.9–76.8	5.9
				35.7–36.5	81.4–83.3	4.9
17 MCH	0*–29.4	0–63.2	6.7	9.7–11.1	23.2–25.4	5.1
				12.8–19.7	27.2–42.5	8.1
17 HGB	9.0–30.0	17.5–63.2	6.7	10.4–10.8	23.3–25.4	4.5
	43.1–54.4	75.1–85.5	3.3	15.9–22.5	31.4–48.1	6.2

Traits are listed by chromosome (Chr). QTL intervals and LOD scores are indicated in cM and Mb for F2 and F2+F34. Asterisks indicate that QTL is bordered by chromosome end; intervals refer to chromosome positions where LOD score exceeds genome-wide significance thresholds.