Genome-wide Trans-ethnic Meta-analysis Identifies Seven Genetic Loci Influencing Erythrocyte Traits and a Role for RBPMS in Erythropoiesis


Genome-wide association studies (GWASs) have identified loci for erythrocyte traits in primarily European ancestry populations. We conducted GWAS meta-analyses of six erythrocyte traits in 71,638 individuals from European, East Asian, and African ancestries using a Bayesian approach to account for heterogeneity in allelic effects and variation in the structure of linkage disequilibrium between ethnicities. We identified seven loci for erythrocyte traits including a locus (RBPMS/GTF2E2) associated with mean corpuscular hemoglobin and mean corpuscular volume. Statistical fine-mapping at this locus pointed to RBPMS at this locus and excluded nearby GTF2E2. Using zebrafish morpholino to evaluate loss of function, we observed a strong in vivo erythropoietic effect for RBPMS but not for GTF2E2, supporting the statistical fine-mapping at this locus and demonstrating that RBPMS is a regulator of erythropoiesis. Our findings show the utility of trans-ethnic GWASs for discovery and characterization of genetic loci influencing hematologic traits.

Introduction

Erythrocyte disorders are common worldwide, contributing to substantial morbidity and mortality.1 Erythrocyte counts and indices are heritable (estimated $h^2 = 0.40–0.90^2$), exhibit different patterns across ethnic groups, and have been influenced by selection in various ethnic groups, most notably for protection against infection by parasites such as those that cause malaria.5–7 Erythrocyte traits have been studied most extensively in European

1Department of Epidemiology, Erasmus MC, 3000 CA Rotterdam, the Netherlands; 2GeneSTAR Research Program, Johns Hopkins University School of Medicine, Baltimore, MD 21287, USA; 3Faculty of Medicine, University of Iceland, 101 Reykjavik, Iceland; 4Icelandic Heart Association, 210 Kopavogur, Iceland; 5Harvard Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA 02138, USA; 6Stem Cell Program and Division of Hematology/Oncology, Children’s Hospital Boston, Pediatric Hematology/Oncology at DFCI, Harvard Medical School and Howard Hughes Medical Institute, Boston, MA 02115, USA; 7Department of Cardiology, Leiden University Medical Center, 2300 AC Leiden, the Netherlands; 8Department of Gerontology and Geriatrics, Leiden University Medical Center, 2300 AC Leiden, the Netherlands; 9National Institute on Aging, NIH, Baltimore, MD 21224, USA; 10Laboratory of Neurogenetics, National Institute on Aging, NIH, Bethesda, MD 20892, USA; 11Institute of Biomedical Sciences, Academia Sinica, Taipei 115, Taiwan; 12Institute of Molecular Biology and Biochemistry, Centre for Molecular Medicine, Medical University of Graz, 8010 Graz, Austria; 13Division of Cardiovascular Medicine, Department of Internal Medicine, Department of Human Genetics, University of Michigan, 1500 E. Medical Center Drive, Ann Arbor, MI 48109, USA; 14Department of Neurology, Boston University School of Medicine, Boston, MA 02218, USA; 15Framingham Heart Study, Population Sciences Branch, Division of Intramural Research, National Heart, Lung, and Blood Institute, NIH, Framingham, MA 01702, USA; 16Inich Institute for Multiscule Biology, Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; 17Department of General and Interventional Cardiology, University Heart Centre Hamburg-Eppendorf, 20246 Hamburg, Germany; 18Department of Genetics, University of North Carolina, Chapel Hill, NC 27599, USA; 19Department of Medicine, University of Washington, Seattle, WA 98195-6420, USA; 20Department of Genetics, School of Medicine, Mashhad University of Medical Sciences, 91375-345 Mashhad, Iran; 21National Institute on Aging, NIH, Bethesda, MD 20892-9020, USA; 22Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA; 23University of Michigan Medical School, Ann Arbor, MI 48109, USA; 24Department of Biostatistical Sciences, Wake Forest School of Medicine, Winston-Salem, NC 27101, USA; 25School of Chinese Medicine, China Medical University, Taichung 40402, Taiwan; 26Clinical Division of Neurogeriatrics, Department of Neurology, Medical University of Graz, 8010 Graz, Austria; 27Human Genetics Center, School of Public Health, University of Texas Health Science Center at Houston, Houston, TX 77030, USA; 28Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo 108-8639, Japan; 29Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences, Yokohama 230-0045, Japan; 30Center for Thrombosis and Hemostasis (CTH), University Medical Center Mainz, 55131 Mainz, Germany; 31German Center for Cardiovascular Research (DZHK), Partner Site RhineMain, Mainz, Germany; 32Preventive Cardiology and Preventive Medicine, Center for Cardiology, University Medical Center of the Johannes Gutenberg-University Mainz, 55131 Mainz, Germany; 33Institute of Cardiovascular and Medical Sciences, Faculty of Medicine, University of Glasgow, Glasgow G12 8QQ, UK; 34Robertson Center for Biostatistics, University of Glasgow, Glasgow G12 8QQ, UK; 35Department of Medical...
ancestry populations, with smaller studies in non-European populations, and have shown both shared and distinct genetic loci influencing erythrocyte traits. Trans-ethnic meta-analysis of genome-wide association studies (GWASs) offers improved signal detection in a combined meta-analysis when heterogeneity of allele effects, allele frequencies, and differences in linkage disequilibrium (LD) between ethnicities are accounted for. Trans-ethnic meta-analysis can also enable fine-mapping of association intervals by evaluating differences in LD structure between diverse populations, thereby enhancing the detection of causal variants.

We conducted trans-ethnic GWAS meta-analyses with the goal of elucidating the genetic architecture of erythrocyte traits and to evaluate (1) whether combining data across populations of diverse ancestry may improve power to detect associations for erythrocyte traits and (2) whether differences in LD structure can be exploited to identify causal variants driving the observed associations with common SNPs. In this study, we analyzed GWAS summary statistics from 71,638 individuals from three diverse populations of European (EUR), East Asian (EAS), and African (AFR) ancestry. We conducted replication analyses in independent samples and performed functional testing to support our approach to fine-mapping.

Subjects and Methods

Study Samples
We aggregated HapMap-imputed GWAS results from 71,638 individuals represented in 23 cohorts embedded in the CHARGE Consortium (40,258 individuals of EUR ancestry), the RIKEN/BioBank Japan Project and AGEN cohorts (15,252 individuals of EAS ancestry), and the COGENT Consortium (16,128 individuals of AFR ancestry). Phenotypic information on all participating
cohort size \( N = 16,389 \) has been reported previously. We conducted replication analyses of the identified trait-loci associations in six independent studies: the Gutenberg Health Study (GHS cohorts 1 and 2, both EUR ancestry), the Genes and Blood-Clotting Study (GBC, EUR ancestry), the NEO study (EUR ancestry), the JUPITER trial (EUR ancestry), and the HANDLS study (AFR ancestry). The COGENT consortium cohorts comprise unrelated individuals of East Asian ancestry (EAS), the JUPITER trial (EUR ancestry), and the HANDLS study (AFR ancestry). SNPs with a minor allele frequency \(< 1\%\), or \(\text{HWE } p < 1 \times 10^{-7}\) were excluded. Genotypes were imputed to approximately 2.5 million SNPs using HapMap Phase II CEU. The RIKEN and the BioBank Japan Project and AGEN cohorts comprise unrelated individuals of East Asian ancestry (EAS). SNPs with a minor allele frequency < 0.01, missingness > 1\%, or \(\text{HWE } p < 1 \times 10^{-7}\) were excluded. Individuals with a call rate < 98\% were excluded as well. Genotypes were imputed to approximately 2.5 million SNPs using HapMap Phase II JPT and CHB. The COGENT consortium cohorts comprise individuals of African American ancestry (AFR). SNPs with a minor allele frequency < 1\% or missingness > 10\% were excluded. Genotypes were imputed to approximately 2.5 million SNPs using HapMap Phase II CEU and YRI.

Cohort-Specific GWASs

For the initial GWA analyses, each cohort used linear regression to assess the association of all SNPs meeting the quality control criteria with each of the six traits separately. An additive genetic model was used and the regressions were adjusted for age, sex, and study site (if applicable). The Framingham Heart Study and the GeneSTAR study used linear mixed effects models to account for relatedness, and these models included adjustment for principal components.

Ethnic-Specific GWAS Meta-analyses

GWAS results of SNPs with a minor allele frequency (MAF) \(\geq 1\%\) and an imputation quality \(> 30\%\) were analyzed in a fixed-effect meta-analysis (METAL software) within each ancestry group, with genomic control (GC) correction of the individual GWAS results of each contributing cohort and the final meta-analysis results.

Trans-ethnic Meta-analyses

For the trans-ethnic meta-analyses, the three sets of the ethnic-specific meta-analysis summary statistics were then combined with three approaches. First, we performed for each trait a trans-ethnic fixed-effect inverse variance-weighted meta-analysis of the EUR, EAS, and AFR GWAS summary statistics using METAL. Second, the ethnic-specific GWAS summary statistics were also combined using the MANTRA (Meta-Analysis of Trans-ethnic Association Studies) package, a meta-analysis software tool allowing for heterogeneity in allelic effects due to differences in LD structure in different ancestry clusters. MANTRA results are reported as \(\log_{10}\) Bayes’s factors \((\log_{10}\text{BF})\). Finally, the three sets of ethnic-specific results were analyzed by means of the Han and Eskin RE2 model, a meta-analysis method developed for higher statistical power under heterogeneity. We used the METASOFT 3.0c tool as developed by the Buhm Han laboratories.

Replication in Human Cohorts

The six independent replication studies—the Gutenberg Health Study (GHS cohorts 1 and 2, both EUR ancestry), the Genes and Blood-Clotting Study (GBC, EUR ancestry), the NEO study (EUR ancestry), the JUPITER trial (EUR ancestry), and the HANDLS study (AFR ancestry) (total replication size \(N = 16,389\))—provided linear regression results for the nine trait-locus combinations. Their results were meta-analyzed with a fixed effects inverse variance weighted method (METAL) and the RE2 methodology. Additionally, we meta-analyzed replication results with the discovery data using fixed-effects, MANTRA, and RE2 methods. For the replication analyses of the nine individual trait-locus combinations, we applied a threshold of \(p < 0.05/9\). Additional human replication findings are provided in Supplemental Data.

Fine-Mapping

We used the MANTRA results to fine-map the regions of trait-associated index SNPs. We defined regions by identifying variants within a 1 Mb window around each index SNP (500 kb upstream and 500 kb downstream). For each SNP in a region, the posterior probability that this SNP is driving the region’s association signal was calculated by dividing the SNP’s BF by the summation of the BFs of all SNPs in the region. Credible sets (CSs) were subsequently created by sorting the SNPs in each region in descending order based on their BF (starting with the index SNP since this SNP has the region’s largest BF by definition). Going down the sorted list, the SNPs’ posterior probabilities were summed until the cumulative value exceeded 99% of the total cumulative posterior probability for all SNPs in the region. The length of a CS was expressed in base pairs. We compared 99% CSs for the trans-ethnic results and the results of a EUR-only MANTRA analysis. For the MANTRA fine-mapping analyses, a less stringent threshold value of \(\log_{10}\text{BF} > 5\) was applied, because we wanted to include...
confirmed by peptide sequence homology study and gene synteny study. 30,31 and hours post-fertilization (hpf) to minimize differences in embryonic development staging caused by the MO injection. 30,31

Heterogeneity Analysis
Heterogeneity of the associations across the different ethnicities was assessed by the I² and Cochran’s Q statistics as reported by METAL.25 and the posterior probability of heterogeneity as reported by MANTRA.24

ENCOD EnCODE Annotation
We evaluated the SNPs identified in the discovery analyses against the ENCODE Project Consortium’s database of functional elements in the K562 erythroleukemic line.29

Experiments in Zebrafish
To substantiate the fine mapping of the RBPMS/GTF2E2 region biologically, we tested the effect of morpholino knockdown in zebrafish for both RBPMS and GTF2E2 orthologous genes, followed by assays of erythrocyte development. Zebrafish rbpms, rbpms2, and gtf2e2 were identified and confirmed by peptide sequence homology study and gene synteny analysis. For rbpms, we relied solely on peptide homology comparison and domain structure since no syntenic region was previously annotated and found by this study.

For each morpholino (MO), its design incorporated information about gene structure and translational initiation sites (Gene-Tool Inc.). MOs targeting each transcript were injected into single-cell embryos at 1, 3, and 5 ng/embryo to find an optimal dose at which there was minimal non-specific toxicity. The stepwise doses also give a range of phenotypes from a hypomorph to a near complete knockdown. After injection, embryos were assayed for hematopoietic development at 1, 3, and 5 ng/embryo to find an optimal dose at which there was minimal non-specific toxicity. The stepwise doses also give a range of phenotypes from a hypomorph to a near complete knockdown. After injection, embryos were assayed for hematopoietic development at 1, 3, and 5 ng/embryo to find an optimal dose at which there was minimal non-specific toxicity. The stepwise doses also give a range of phenotypes from a hypomorph to a near complete knockdown. After injection, embryos were assayed for hematopoietic development at 1, 3, and 5 ng/embryo to find an optimal dose at which there was minimal non-specific toxicity. The stepwise doses also give a range of phenotypes from a hypomorph to a near complete knockdown. After injection, embryos were assayed for hematopoietic development at 1, 3, and 5 ng/embryo to find an optimal dose at which there was minimal non-specific toxicity. The stepwise doses also give a range of phenotypes from a hypomorph to a near complete knockdown. After injection, embryos were assayed for hematopoietic development at 1, 3, and 5 ng/embryo to find an optimal dose at which there was minimal non-specific toxicity. The stepwise doses also give a range of phenotypes from a hypomorph to a near complete knockdown. After injection, embryos were assayed for hematopoietic development at 1, 3, and 5 ng/embryo to find an optimal dose at which there was minimal non-specific toxicity. The stepwise doses also give a range of phenotypes from a hypomorph to a near complete knockdown. After injection, embryos were assayed for hematopoietic development at 1, 3, and 5 ng/embryo to find an optimal dose at which there was minimal non-specific toxicity. The stepwise doses also give a range of phenotypes from a hypomorph to a near complete knockdown. After injection, embryos were assayed for hematopoietic development at 1, 3, and 5 ng/embryo to find an optimal dose at which there was minimal non-specific toxicity. The stepwise doses also give a range of phenotypes from a hypomorph to a near complete knockdown. After injection, embryos were assayed for hematopoietic development at 1, 3, and 5 ng/embryo to find an optimal dose at which there was minimal non-specific toxicity. The stepwise doses also give a range of phenotypes from a hypomorph to a near complete knockdown. After injection, embryos were assayed for hematopoietic development at 1, 3, and 5 ng/embryo to find an optimal dose at which there was minimal non-specific toxicity. The stepwise doses also give a range of phenotypes from a hypomorph to a near complete knockdown. After injection, embryos were assayed for hematopoietic development at 1, 3, and 5 ng/embryo to find an optimal dose at which there was minimal non-specific toxicity. The stepwise doses also give a range of phenotypes from a hypomorph to a near complete knockdown. After injection, embryos were assayed for hematopoietic development at 1, 3, and 5 ng/embryo to find an optimal dose at which there was minimal non-specific toxicity. The stepwise doses also give a range of phenotypes from a hypomorph to a near complete knockdown. After injection, embryos were assayed for hematopoietic development at 1, 3, and 5 ng/embryo to find an optimal dose at which there was minimal non-specific toxicity. The stepwise doses also give a range of phenotypes from a hypomorph to a near complete knockdown. After injection, embryos were assayed for hematopoietic development at 1, 3, and 5 ng/embryo to find an optimal dose at which there was minimal non-specific toxicity. The stepwise doses also give a range of phenotypes from a hypomorph to a near complete knockdown. After injection, embryos were assayed for hematopoietic development at 1, 3, and 5 ng/embryo to find an optimal dose at which there was minimal non-specific toxicity. The stepwise doses also give a range of phenotypes from a hypomorph to a near complete knockdown. After injection, embryos were assayed for hematopoietic development at 1, 3, and 5 ng/embryo to find an optimal dose at which there was minimal non-specific toxicity. The stepwise doses also give a range of phenotypes from a hypomorph to a near complete knockdown. After injection, embryos were assayed for hematopoietic development at 1, 3, and 5 ng/embryo to find an optimal dose at which there was minimal non-specific toxicity. The stepwise doses also give a range of phenotypes from a hypomorph to a near complete knockdown. After injection, embryos were assayed for hematopoietic development at 1, 3, and 5 ng/embryo to find an optimal dose at which there was minimal non-specific toxicity. The stepwise doses also give a range of phenotypes from a hypomorph to a near complete knockdown. After injection, embryos were assayed for hematopoietic development at 1, 3, and 5 ng/embryo to find an optimal dose at which there was minimal non-specific toxicity. The stepwise doses also give a range of phenotypes from a hypomorph to a near complete knockdown. After injection, embryos were assayed for hematopoietic development at 1, 3, and 5 ng/embryo to find an optimal dose at which there was minimal non-specific toxicity. The stepwise doses also give a range of phenotypes from a hypomorph to a near complete knockdown. After injection, embryos were assayed for hematopoietic development at 1, 3, and 5 ng/embryo to find an optimal dose at which there was minimal non-specific toxicity. The stepwise doses also give a range of phenotypes from a hypomorph to a near complete knockdown. After injection, embryos were assayed for hematopoietic development at 1, 3, and 5 ng/embryo to find an optimal dose at which there was minimal non-specific toxicity. The stepwise doses also give a range of phenotypes from a hypomorph to a near complete knockdown. After injection, embryos were assayed for hematopoietic development at 1, 3, and 5 ng/embryo to find an optimal dose at which there was minimal non-specific toxicity. The stepwise doses also give a range of phenotypes from a hypomorph to a near complete knockdown. After injection, embryos were assayed for hematopoietic development at 1, 3, and 5 ng/embryo to find an optimal dose at which there was minimal non-specific toxicity. The stepwise doses also give a range of phenotypes from a hypomorph to a near complete knockdown. After injection, embryos were assayed for hematopoietic development at 1, 3, and 5 ng/embryo to find an optimal do...
(Life Technologies, 12027). A tail was added to the end-repaired DNA using NEB Klenow Fragment Enzyme (3'-5'exo, M0212L), 1× NEB buffer 2, and 0.2 mM dATP (Invitrogen, 18252-015) and incubating the reaction mix at 37°C for 30 min. A-tailed DNA was cleaned up using AMPure beads (1.8× reaction volume). Subsequently, cleaned-up a-Tailed DNA was through Adapter ligation reaction using Quick Ligation Kit (NEB, M2200L) according to the manufacturer’s protocol. Adapter-ligated DNA was first cleaned up using AMPure beads (1.8× of reaction volume), eluted in 100 μL and then size-selected using AMPure beads (0.9× of the final supernatant volume, 90 μL). Adapter ligated DNA fragments of proper size were enriched with PCR reaction using Fusion High-Fidelity PCR Master Mix kit (NEB, M0531S) and specific index primers supplied in NEBNext Multiplex Oligo Kit for Illumina (Index Primer Set 1, NEB, E7335L). Conditions for PCR used are as follows: 98°C, 10 s; 65°C, 30 s; 72°C, 30 s × 15 to 18 cycles; 72°C, 5 min; hold at 4°C. PCR-enriched fragments were further size selected by running the PCR reaction mix in 2% low-molecular-weight agarose gel (Bio-Rad, 161-3107) and subsequently purifying them using QIAquick Gel Extraction Kit (28704). Libraries were eluted in 25 μL elution buffer. After measuring concentration in Qubit, all the libraries went through quality-control analysis using an Agilent Bioanalyzer. Samples with proper size (250–300 bp) were selected for next generation sequencing using Illumina HiSeq 2000 or 2500 platform.

Alignment and visualization ChIP-seq reads were aligned to the human reference genome (hg19) using bowtie with parameters -k 2 -m 2 -s. 34 WIG files for display were created using MACS with parameters -w -S-space = 50-nomodel-shiftsize = 200 and were displayed in IGV. 36, 37 High-confidence peaks of ChIP-seq signal were identified using MACS with parameters --keepdups = auto -p 1e-9 and corresponding input control. Bound genes are RefSeq genes that contact a MACS-defined peak between −10,000 bp from the TSS and +5,000 bp from the TSS.

For the assay for transposase accessible chromatin (ATAC-seq), CD34+ cells were expanded and differentiated using the protocol mentioned above. Before collection, cells were treated with 25 ng/mL hrBMP4 for 2 hr. 5 × 10⁶ cells per differentiation stage were harvested by spinning at 500 × g for 5 min, 4°C. Cells were washed once with 50 μL of cold 1× PBS and spun down at 500 × g for 5 min, 4°C. After discarding supernatant, cells were lysed using 50 μL cold lysis buffer (10 mM Tris-HCl [pH 7.4], 10 mM NaCl, 3 mM MgCl₂, 0.1% IGEPAL CA-360) and spun down immediately at 500 × g for 10 min, 4°C. The cells were then precipitated and kept on ice and subsequently resuspended in 25 μL 2X TD Buffer (Illumina Nextera kit), 2.5 μL transposase enzyme (Illumina Nextera kit, 15028252), and 22.5 μL nuclease-free water in a total of 50 μL reaction for 1 hr at 37°C. DNA was then purified using QIAGEN MinElute PCR purification kit (28004) in a final volume of 10 μL. Libraries were constructed according to Illumina protocol using the DNA treated with transposase, NEB PCR master mix, Sybr green, and universal and library-specific Nextera index primers. The first round of PCR was performed under the following conditions: 72°C, 5 min; 98°C, 30 s; (98°C, 10 s; 63°C, 30 s; 72°C, 1 min) × 5 cycles; hold at 4°C. Reactions were kept on ice and, using a 5 μL reaction aliquot, the appropriate number of additional cycles required for further amplification was determined in a side qPCR reaction: 98°C, 30 s; (98°C, 10 s; 63°C, 30 s; 72°C, 1 min) × 20 cycles; hold at 4°C. Upon determining the additional number of PCR cycles required further for each sample, library amplification was conducted using the following conditions: 98°C, 30 s; (98°C, 10 s; 63°C, 30 s; 72°C, 1 min) × appropriate number of cycles; hold at 4°C. Libraries prepared went through quality-control analysis using an Agilent Bioanalyzer. Samples with appropriate nucleosomal laddering profiles were selected for next generation sequencing using Illumina HiSeq 2500 platform.

All human ChIP-seq datasets were aligned to build version NCB137/HG19 of the human genome using Bowtie2 (v.2.2.1)34 with the following parameters:-end-to-end, -N0, -L20. We used the MACS2 v.2.1.0L35 peak finding algorithm to identify regions of ATAC-seq peaks, with the following parameter: --nomodel-shift = −100-extsize 200. A q-value threshold of enrichment of 0.05 was used for all datasets.

Evaluation in Mouse Crosses
To further affirm the trait loci we identified, and in an attempt to further fine-map the intervals identified in our discovery analyses through cross-species comparisons, we evaluated the new loci in syntenic regions in 12 inter-strain mouse QTL crosses. 38 In brief, mice from 12 different strains were inter-crossed and the same erythrocyte traits we have studied by GWAS were measured in peripheral blood. The Jackson Laboratory Animal Care and Use Committee approved all protocols. The number of markers genotyped per cross varied by the platform used, and the total number per cross is provided in Table S9. QTL analysis was performed for each erythrocyte trait using R/qtl v1.07-12 (Web Resources). Genetic map positions of all markers used were updated to the new mouse genetic map using online mouse map converter tool (Web Resources). All phenotypic data were ranked-Z transformed to approximate the normal distribution prior to analysis. The QTL analysis was performed as a genome-wide scan with sex as an additive covariate. Permutation testing (1,000 permutations) was used to determine significance, and LOD scores greater than the 95th percentile (p < 0.05) were considered significant. QTL confidence intervals were determined by the posterior probability. For each candidate region in the mouse, the coordinates were obtained from the Mouse Genome Database, which is part of Mouse Genome Informatics (MGI), using the “Genes and Markers” query (Web Resources). Protein coding genes, non-coding RNA genes, and unclassified genes were queried.

Results
In this study we analyzed the association of genetic variation in 71,638 individuals and 6 clinically relevant erythrocyte traits which are commonly measured, accounting for the diverse ethnic background of the participants. We identified 44 previously reported loci and 9 other significant trait-locus associations at 7 loci (p < 5 × 10^{-8} or log₁₀BF > 6.1, Table 1). SHROOM3 was simultaneously identified in an exome chip analysis by our group in overlapping samples. Ethnic-specific results are presented in Table S4. Regional association plots are shown for each region in Figure S1, showing ethnic-specific results, the trans-ethnic meta-analysis, and plots of pairwise LD across the regions for EUR, EAS, and AFR ancestry.

Five of the discovered trait loci showed a significant association in the fixed-effects trans-ethnic METAL analyses, in

The American Journal of Human Genetics 100, 51–63, January 5, 2017 55
Furthermore, we found nominal significance for MCV-effects model and nominally significant in the RE2 model. RBPMS achieved a Bonferroni-corrected significance for MCH in the trans-ethnic meta-analyses because non was accompanied by greater evidence of heterogeneity and thus yield a stronger result as compared to MCV in the AFR meta-analyses. This phenotype was observed stronger associations for Hct-TEMEM163/ACMSD, MCH-PLCL2, MCV-ID2, and MCV-RBPMS in all three models (fixed-effects, MANTRA, and RE2). For MCH-RBPMS, we found a stronger association in the fixed-effects analysis (Table S6).

Statistical Fine-Mapping
We found that 31 trait-specific trans-ethnic 99% CSs showed a decrease in length of at least 50% as compared to their EUR-only CS counterparts (26 unique loci across the 6 erythrocyte traits) (Table S7).

Among the loci identified in this study, the chromosome 8 RBPMS locus showed fine-mapping according to this criterion (Table 2, Figure 1). For MCH, the EUR credible set spanned 204,200 bp, encompassing RBPMS and GTF2E2. The multi-ethnic credible set comprised just one SNP, rs2979489, within the first intron of RBPMS (Figure 1). Remarkably, this associated SNP rs2979489 is located adjacent to a GATA-motif where a gradual switch of binding from GATA2 to GATA1 takes place during commitment of human CD34 progenitors toward erythroid lineage (Figure 2, bottom left). Moreover, an assay for chromatin accessibility sites (ATAC-seq) and H3K27a ChIP-seq clearly identify that the genomic region proximal to this SNP is actively regulated during human erythroid differentiation (Figure 2, bottom right).

Among the known loci, fine mapping narrowed signals to their EUR-only CS counterparts (26 unique loci across the 6 erythrocyte traits) (Table S7). The multi-ethnic credible set comprised just one SNP, rs2979489, within the first intron of RBPMS (Figure 1). Remarkably, this associated SNP rs2979489 is located adjacent to a GATA-motif where a gradual switch of binding from GATA2 to GATA1 takes place during commitment of human CD34 progenitors toward erythroid lineage (Figure 2, bottom left). Moreover, an assay for chromatin accessibility sites (ATAC-seq) and H3K27a ChIP-seq clearly identify that the genomic region proximal to this SNP is actively regulated during human erythroid differentiation (Figure 2, bottom right).
analysis of their GWAS results.\textsuperscript{10} By means of subsequent formaldehyde-assisted isolation of regulatory elements followed by next-generation sequencing (FAIRE-seq), they pinpointed rs6791816 as an NDR SNP in LD with their initial index SNP for MCH and MCV.

Furthermore, fine-mapping of both the \textit{MPND} locus (MCH) and \textit{SH3GL1} locus (MCV) pointed to the rs8887 SNP within the 3' UTR of \textit{PLIN4}. The rs8887 SNP minor allele has been shown experimentally to create a novel seed site for miR-522, resulting in decreased \textit{PLIN4} expression.\textsuperscript{49} miR-522 is expressed in circulating blood,\textsuperscript{50} and these data suggest that an allele-specific miR-522 regulation of \textit{PLIN4} by rs8887 could serve as a functional mechanism underlying the identified association.

We additionally showed fine mapping in several other intervals (Table S7) with fine-mapped genes about which less is known about their potential biologic role in erythropoiesis or red blood cell function. These regions are of interest for further hypothesis generation based upon the GWAS findings.

\textbf{ENCODE Analyses}

We further evaluated the SNPs from the chromosome 8 \textit{RBPMS} region against the ENCODE Project Consortium's database of numerous functional elements in the K562 erythroleukemic line.\textsuperscript{29} The lone SNP that was fine mapped at the locus, rs2979489, was found in a strong enhancer element as defined by Segway, supporting a functional role for this SNP and \textit{RBPMS}. The other SNPs in the \textit{RBPMS} region, excluded by the statistical fine-mapping exercise, were not annotated as regulatory in the ENCODE data (Table S8).

\begin{table}[h]
\centering
\begin{tabular}{|l|l|l|l|l|l|l|l|}
\hline
\textbf{Trait} & \textbf{Chr} & \textbf{Gene} & \textbf{EUR topSNP} & \textbf{log}_{10}\textbf{BF} & \textbf{n\_SNPs} & \textbf{width (bp)} & \textbf{Multi-ethnic topSNP} & \textbf{log}_{10}\textbf{BF} & \textbf{n\_SNPs} & \textbf{width (bp)} \\
\hline
MCH & 8 & \textit{RBPMS} & rs2979502 & 6.32982 & 21 & 241480 & rs2979489 & 9.72267 & 1 & 1 \\
MCV & 8 & \textit{RBPMS} & rs2979489 & 6.13733 & 11 & 241480 & rs2979489 & 7.96132 & 1 & 1 \\
\hline
\end{tabular}
\caption{Fine Mapping of a Chromosome 8 Locus Identified in European Ancestry Meta-analysis by MANTRA Trans-ethnic Analysis}
\textbf{Abbreviations are as follows: chr, chromosome number; log}_{10}\textbf{BF}, logarithm of Bayes Factor; n\_SNPs, number of SNPs in the region.}
\end{table}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure1.jpg}
\caption{Fine Mapping of the Chromosome 8 \textit{RBPMS/GTF2E2} Locus}
\textit{99\%} credible sets (red dots) around the top hit rs2979489 (red diamond). European Ancestry MANTRA analyses (top) for MCH (left) and MCV (right) are shown, compared to 99\% credible sets of the trans-ethnic MANTRA analyses (bottom, MCH on the left and MCV on the right).}
Experiments in Zebrafish

We identified an erythropoietic effect for the zebrafish rbpms. Both embryonic globin expression at 16 ss and o-dianisidine/benzidine staining at 48 hpf significantly decreased in morphants, indicating a decrease in both globin transcription and Hb levels (Figure 3). This loss-of-function finding is consistent with a decreased mean erythrocyte Hb content observed in our human association results. In zebrafish, the rbpms orthology mapping included rbpms2a, rbpms2b, and rbpms, and loss-of-function phenotypes of all orthologs were tested experimentally. The results suggested a clear erythropoietic effect with limited functional compensation of the genes in the rbpms family in zebrafish during embryonic erythropoiesis. On the other hand, morpholino knockdown experiments with the zebrafish ortholog of GTF2E2 did not show an apparent erythropoietic effect.

Review of the human association results showed no evidence of pleiotropy across the RBPMS family of genes and denote that the human association is specific to RBPMS (Supplemental Data). This review was conducted because the orthology in the fish led to inclusion of rbpms2 in the zebrafish analyses as well. These findings indicate that the statistical fine-mapping was useful to home in on RBPMS as a causal gene influencing erythropoiesis.

Evaluation in Mouse Crosses

In the eight regions from our discovery analysis, six had evidence of cross-species validation by evidence of syntenic gene within the linkage peak in the mouse QTL results (Table 3). However, the human GWAS intervals were not narrowed by the mouse QTL results for any of these loci (Table S9).

Discussion

We conducted GWASs and meta-analyses of six erythrocyte traits (Hb, Hct, MCH, MCHC, MCV, and RBC) in 71,638 individuals from European, Asian, and African American ancestry. While prior genome-wide association studies have identified loci associated with erythrocyte traits through the analysis of ancestrally homogeneous cohorts and consortia, largely biased toward European ancestry studies, trans-ethnic analysis has not previously been performed while accounting for differences in genetic architecture in ethnically diverse groups.

We identified seven loci for erythrocyte traits (nine locus-trait combinations) and replicated 44 previously identified loci. We fine-mapped several known and new loci. One fine-mapped locus led us to a region on chromosome 8 associated with MCH and MCV.

In the chromosome 8 RBPMS/GTF2E2 locus, the index variant rs2979489, which was associated with MCV and MCH and highlighted in the trans-ethnic fine-mapping analyses, is located within the first intron of RBPMS (RNA binding protein with multiple splicing), notably at an open chromatin site at which a switch of GATA1/2 binding occurs during erythroid differentiation. The RBPMS protein product regulates a variety of RNA processes, including pre-mRNA splicing, RNA transport, localization, translation, and stability. RBPMS is expressed at relatively low levels in mammalian erythroblasts and the protein product has not been detected in mature human erythrocytes.

The rs2979489 polymorphism showed remarkably high heterogeneity in effect on the MCH trait across the different ethnicities, with different directions of effect for the AFR meta-analysis results compared to the EUR and ASN findings. If the variant is causal, this pattern of association could reflect gene-environment interaction. In this case, different exposures in AFR compared to EUR/ASN...
populations may lead to a marginal effect of the SNP in opposing directions by different selection pressures. If, however, rs2979489 is not causal, but rather a marker in LD with the causal variant, then the opposing direction of effects could reflect very different LD structures in the different populations, also indicating selection, or theoretically it could even reflect different causal variants in AFR and EUR/EAS—and rs2979489 being just in strong LD with both causal variants.

The SNP rs2979489 is located adjacent to a GATA-motif where a gradual switch of binding from GATA2 to GATA1 takes place during commitment of human CD34 progenitors toward erythroid lineage. These observations suggest that rs2979489 localizes at a potential regulatory site where a modulation of erythroid cell differentiation occurs and the presence of rs2979489 may lead to observed red cell trait alterations in human populations, possibly through regulation of RBPMS expression timing, level, and/or splicing variation. Although RBPMS previously had no known role in hematopoiesis or more specifically in erythropoiesis, RBPMS has been previously shown to be upregulated in transcriptional profiles of murine and human hematopoietic stem cells. Its role may be at much earlier stages during the differentiation of erythrocytes from erythroblasts and/or hematopoietic stem cells. RBPMS is known to physically interact with Smad2, Smad3, and Smad4 and stimulate smad-mediated transactivation through enhanced Smad2 and Smad3 phosphorylation and associated promotion of nuclear accumulation of Smad proteins. RBPMS has four annotated transcript isoforms, and further delineation of the tissue specificity, timing of expression, and function of these transcripts in the context of the genetic variant we identified warrants further study.

Among the additional six loci, we identified two loci in which the index SNP was located within annotated genes, rs6430549 in ACMSD (aminocarboxymuconate semi aldehyde dehydrogenase, intronic) and rs2299433 in MET (mesenchymal epithelial transition factor, intronic). No previous hematologic role has been described for either region. Variants in the chromosome 2q21.3 ACMSD region

Figure 3. Loss-of-Function Analysis of the RBPMS, RBPMS2, and GTF2E2 Orthologs in Zebrafish

After injection of 0–3 ng ATG and splicing morpholinos (MOs) against the RBPMS zebrafish ortholog (row E), both the o-dianisidine/benzidine staining (arrows) in embryos at 48 hpf (right) and the embryonic β3 globin expression in embryos at 16–18 ss (left) are obviously decreased, indicating a dose-dependent disruption in erythropoiesis in the experimentally treated embryos as compared to uninjected and gtf2e2-, rbpms2a-, and rbpms2b-MO-injected controls (rows A–D). Representative results are shown for the embryos injected with MOs against the RBPMS ortholog in (E) as well as for the embryos injected with MOs against rbpms2a (C) and rbpms2b (D) at higher doses. Injections of MO against the zebrafish GTF2E2 ortholog (B) also at a higher dose show no obvious effect on β3 globin expression at 16–18 ss and o-dianisidine/benzidine staining at 48 hpf. Expression pattern of vascular marker gene kdrl (A–E, middle) is relatively normal in all MO-injected embryos at 24–26 hpf, suggesting grossly normal development of cells in other organs. The numbers on the lower right corner of each image indicate the number of embryos with phenotypes similar to the ones shown on each of the images over the total number of embryos examined in each of the experimental groups.
have previously been associated with blood metabolite levels, obesity, and Parkinson disease.\textsuperscript{60–62} A genetic variant in the first intron of \textit{MET} was significantly associated with both Hb and Hct; however, association was not observed in replication samples, possibly due to lower power in the replication experiment. Three additional loci were intergenic but close to a coding gene (rs10929547 near \textit{ID2} [inhibitor of DNA binding 2, dominant-negative helix-loop-helix protein], rs6121246 near \textit{FOXS1} [forkhead box S1], and rs2060597 approximately 40 kbp upstream of \textit{PLCL2} [phospholipase C-like 2]). The roles of variants in these regions in determining erythrocyte traits are unknown.\textsuperscript{53,63}

In the statistical fine-mapping analyses, the trans-ethnic meta-analysis approach resulted in smaller 99\% credible intervals in all of the loci identified in this study. Since these loci were identified in analyses that accounted for heterogeneity in allelic effects between ethnic groups, in which the heterogeneity may be due to variation in LD patterns, we examined the LD patterns in these loci. Not surprisingly, we noted that the consistent decrease in the size of 99\% credible interval across all loci is likely due to the inclusion of cohorts of African ancestry, an ethnic group with generally smaller LD blocks throughout the genome. The loss-of-function screens in zebrafish for the chromosome 8 signal suggested that these analyses successfully identified a single gene (\textit{RBPMS}) with erythropoietic effect within one of the fine-mapped intervals. We also fine-mapped previously known regions such as the chromosome 6p21.1 region associated with RBC count and highlighted \textit{CCND3}, which has been experimentally shown to regulate RBC count experimentally in a knock-out mouse model.\textsuperscript{64} These examples suggest that attempts to refine association signals using these types of approaches in existing samples may yield functional candidates for further mechanistic hypothesis testing, which is a major goal of GWASs.

Trans-ethnic genome-wide meta-analyses of common variants have aided in the characterization of genetic loci for various complex traits.\textsuperscript{13,65–67} Our data demonstrate the benefits of trans-ethnic genome-wide meta-analysis in identifying and fine-mapping genetic loci of erythrocyte traits. By exploiting the differences in genetic architecture of the associations within these loci in various ethnic groups, we may identify causal genes influencing clinically relevant hematologic traits. Use of a similar approach for other complex traits is likely to provide deeper insights into the biological mechanisms underlying human traits.

**Accession Numbers**

Summary data have been deposited in the database of Genotypes and Phenotypes (dbGaP) under CHARGE (Cohorts for Heart and Aging Research in Genomic Epidemiology) Consortium Summary Results from Genomic Studies. The dbGaP study accession number is phs000930.

**Supplemental Data**

Supplemental data include Supplemental Acknowledgments, individual study methods and cohort descriptions, pleiotropy analysis, 10 tables, and a figure with 123 panels.

**Acknowledgments**

B.M.P. serves on the DSMB of a clinical trial funded by the manufacturer (Zoll LifeCor) and on the Steering Committee of the Yale Open Data Access project funded by Johnson & Johnson.

Received: February 15, 2016
Accepted: November 16, 2016
Published: December 22, 2016

**Web Resources**

Center for Genome Dynamics, http://cgd.jax.org
METASOFT 3.0c, http://www.buhmhan.com/software
MGI Genes and Markers Query, http://www.informatics.jax.org/marker
R/qtl v1.07-12, http://www.rqtl.org

\begin{table}[h]
\centering
\caption{Mouse QTL Validation of the Findings from MANTRA Trans-ethnic Analyses}
\begin{tabular}{|c|c|c|c|c|c|}
\hline
\textbf{Trait} & \textbf{Chr} & \textbf{Gene} & \textbf{Human (hg18/Build 36) (Chromosome:Position)} & \textbf{Mouse (37 mm9) (Chromosome:Position)} & \textbf{Significant and Suggestive Mouse QTL\(^a\)}
\hline
Hct & 2 & \textit{TMEM163}/\textit{ACMSD} & chr2: 135,196,450–135,438,613 & chr1: 129,581,372–129,711,586\(^c\) & 141.0 (54.8–158.9)\(^a\) 3.72\(^a\)
\hline
Hct & 4 & \textit{SHROOM3} & chr4: 77,586,311–77,629,342 & chr5: 93,112,461–93,394,344 & 46.0 (19.6–106.5) 2.34
\hline
Hct & 7 & \textit{MET} & chr7: 116,118,114–116,131,947 & chr6: 17,432,318–17,477,418\(^b\) & 37.6 (6.6–127.9) 2.75
\hline
MCH & 8 & \textit{RBPMS} & chr8: 30,400,375–30,400,375 & chr8: 34,893,115–35,040,335 & 78.9 (28.0–96.1) 3.98\(^a\)
\hline
MCV & 3 & \textit{PLCL2} & chr3: 16,860,239–16,945,942 & chr17: 50,604,848–50,698,773\(^b\) & 46.0 (28.6–55.3) 5.46\(^a\)
\hline
MCV & 20 & \textit{FOXS1} & chr20: 29,684,484–29,897,013 & chr2: 152,576,419–152,758,874\(^c\) & 170.1 (147.6–179.3) 4.69\(^a\)
\hline
\hline
\textsuperscript{a}Gene found in a significant (indicated with asterisk) or suggestive 95\% CI mouse QTL, not corresponding to the human interval.
\textsuperscript{b}Within the corresponding human interval (\pm 250 kbp).
\end{tabular}
\end{table}
References


