

## Supplementary Note for

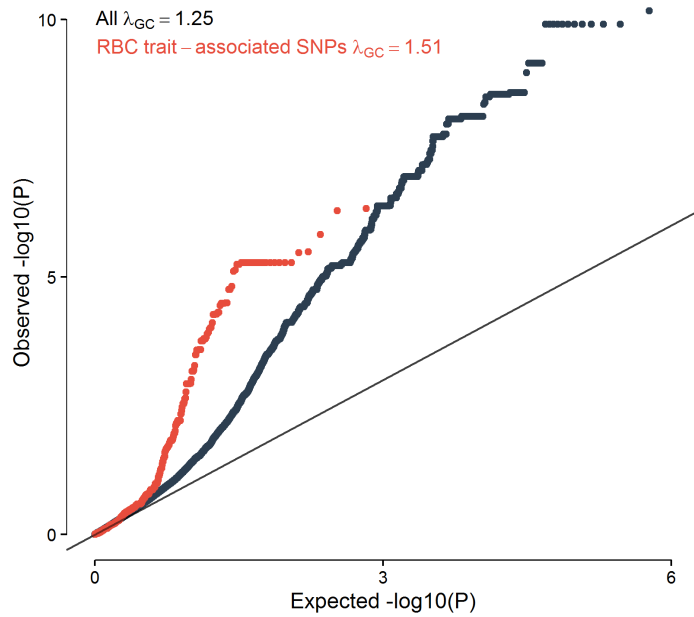
An erythroid-specific enhancer of *ATP2B4* mediates red blood cell hydration and malaria susceptibility

Samuel Lessard, Emily N. Stern, Mélissa Beaudoin, Patrick G. Schupp, Falak Sher, Syed Adnan Ali, Sukhpal Prehar, Ryo Kurita, Yukio Nakamura, Esther Baena, Jonathan Ledoux, Delvac Oceandy, Daniel E. Bauer, Guillaume Lettre

correspondence to: D.E.B. ([Daniel.Bauer@childrens.harvard.edu](mailto:Daniel.Bauer@childrens.harvard.edu)) or G.L. ([guillaume.lettre@umontreal.ca](mailto:guillaume.lettre@umontreal.ca)).

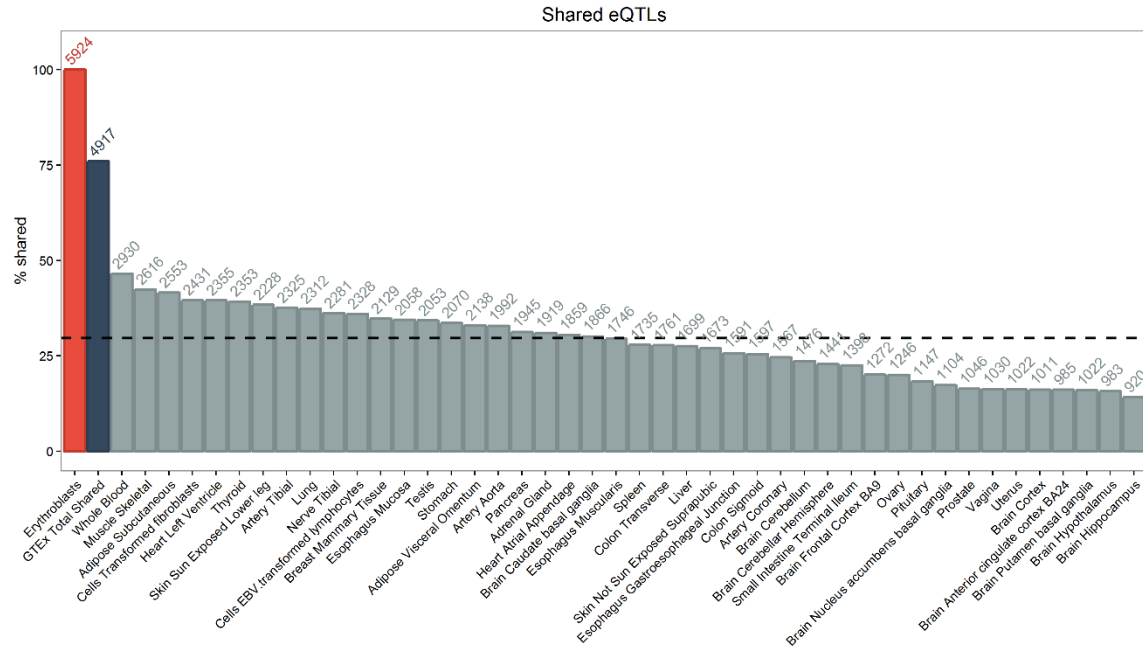
**This PDF file includes:**

Figures S1 to S8  
Tables S1 to S7



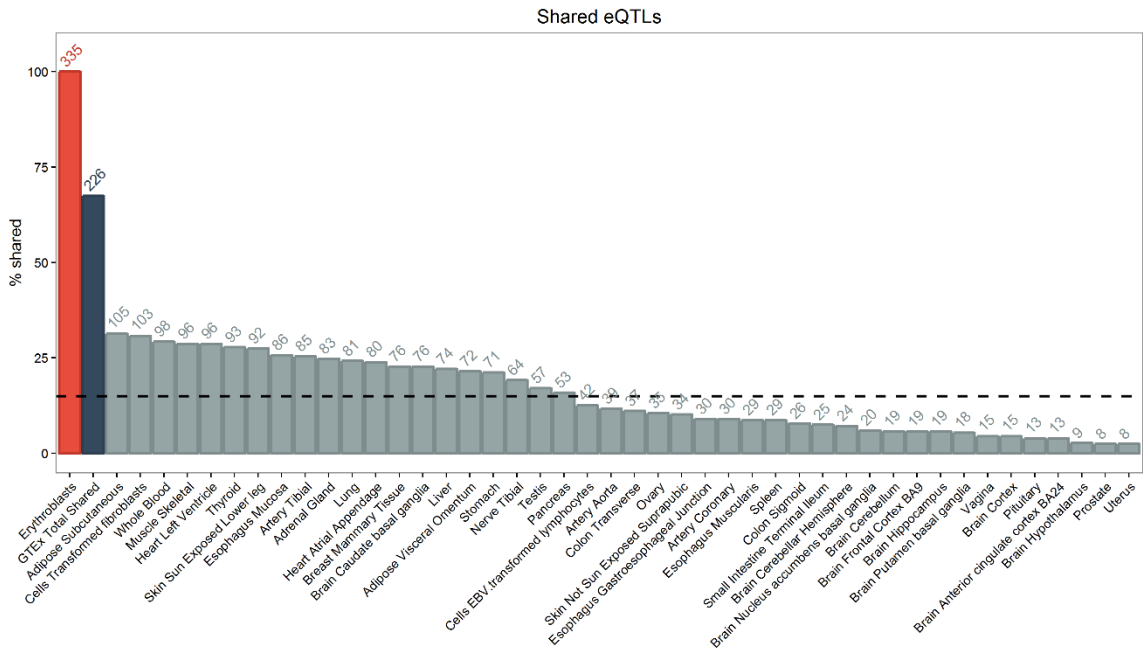
**Figure S1.**

Erythroblasts eQTL are enriched for RBC trait-associated SNPs. Quantile-quantile (QQ) plot of eQTL  $P$ -values for variants located within 100 kb of 479 AI genes in human erythroblasts (black, genomic control lambda ( $\lambda_{GC}$ )=1.25). We find an enrichment of eQTL  $P$ -values for SNPs associated with RBC traits (red,  $\lambda_{GC}$ =1.51)(7). The number of different SNPs included when considering all tested variants and when subsetting on RBC trait-associated SNPs is 258,663 and 178 respectively.

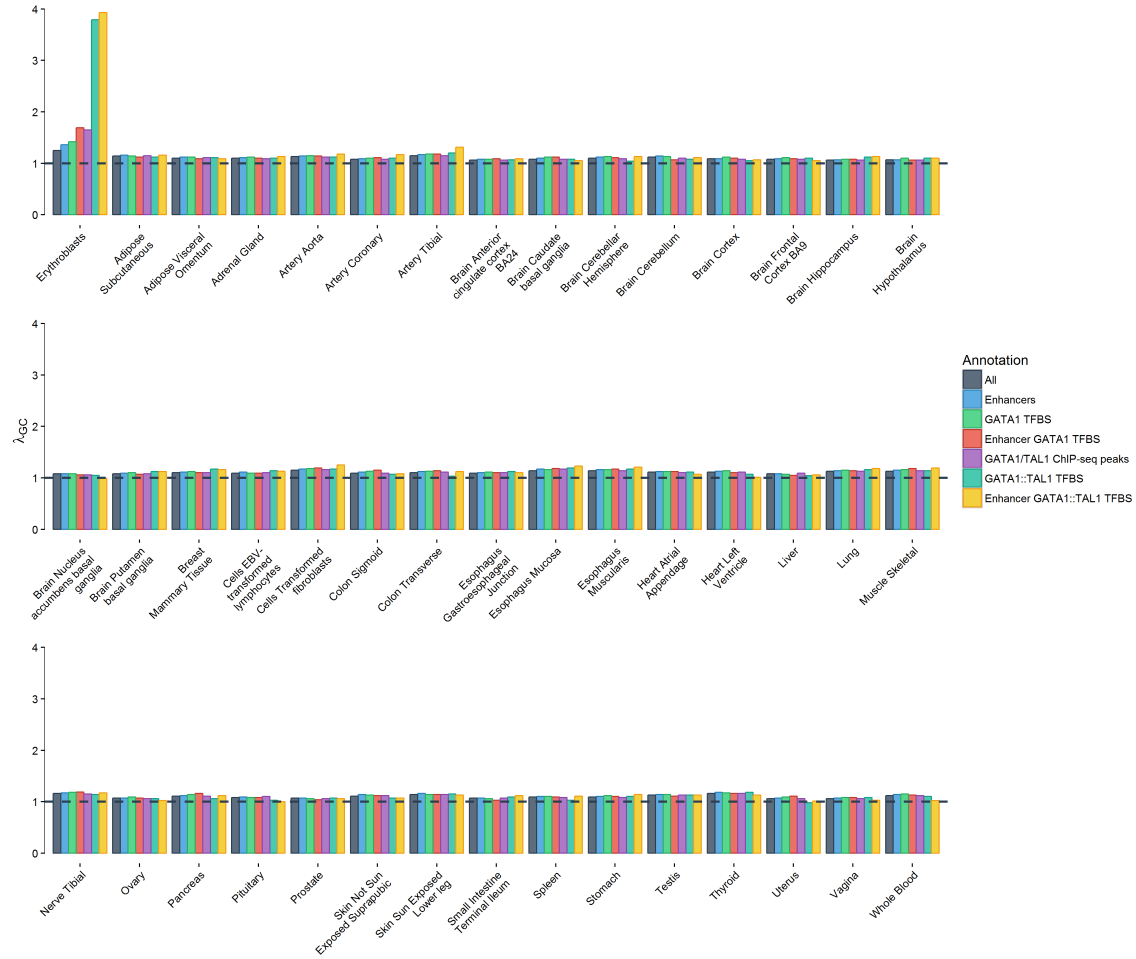


**Figure S2.**

Number of shared eQTLs between erythroblasts and GTEx samples. The dotted line represents the mean percentage of eQTLs shared between human erythroblasts and any of the GTEx tissues (mean=29.6%).

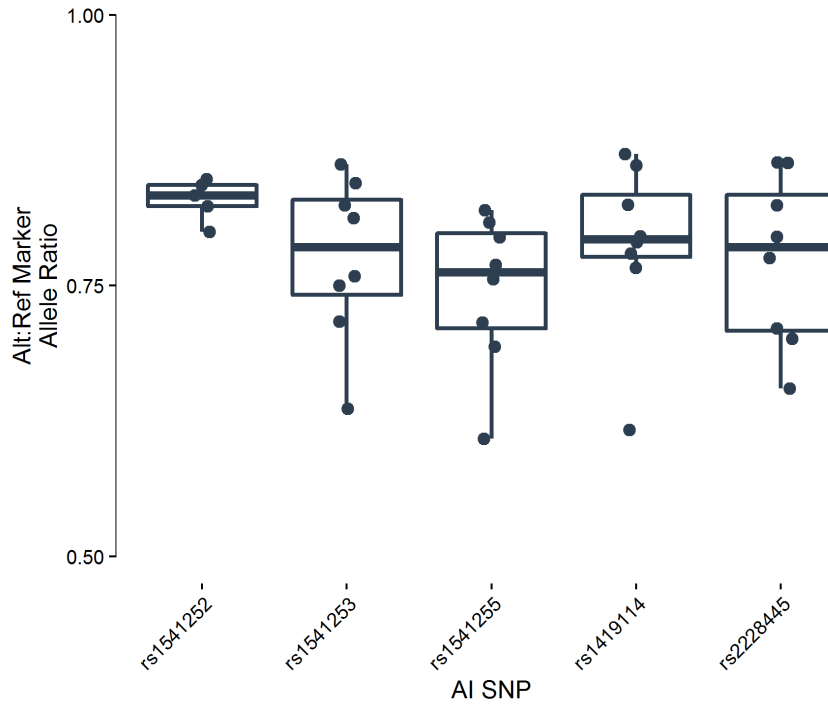


**Figure S3.** Number of eQTLs in GATA1/TAL1 ChIP-seq peaks shared between erythroblasts and GTEx samples. The dotted line represents the mean percentage of shared eQTLs that map to GATA1/TAL1 ChIP-seq peaks between human erythroblasts and any of the GTEx tissues (mean=14.9%).



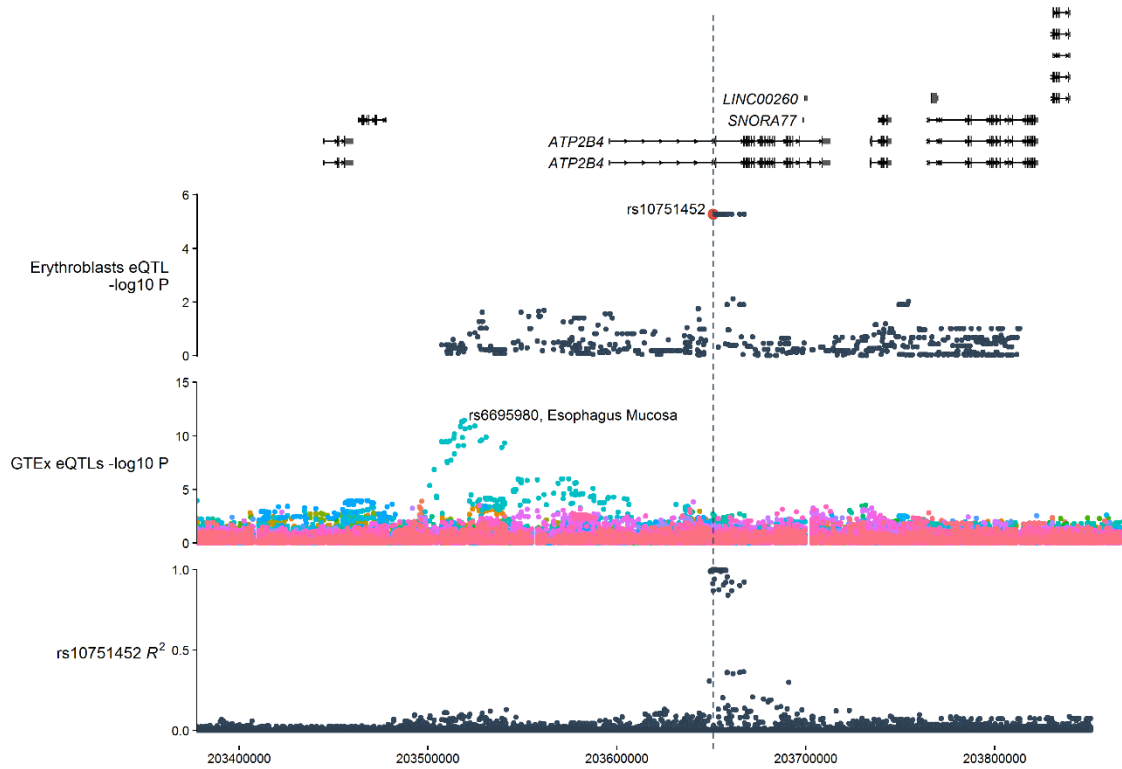
**Figure S4.**

Genomic control lambda ( $\lambda_{GC}$ ) of eQTLs subsetted by erythroid regulatory regions. In this analysis, we compared inflation of eQTL signals observed in human erythroblasts (this study) with the inflation of the same SNPs but in different tissues available in the GTEx resource. We note a marked inflation of eQTL signals for erythroid annotations only in human erythroblasts. Dotted line is  $\lambda_{GC}=1$ .



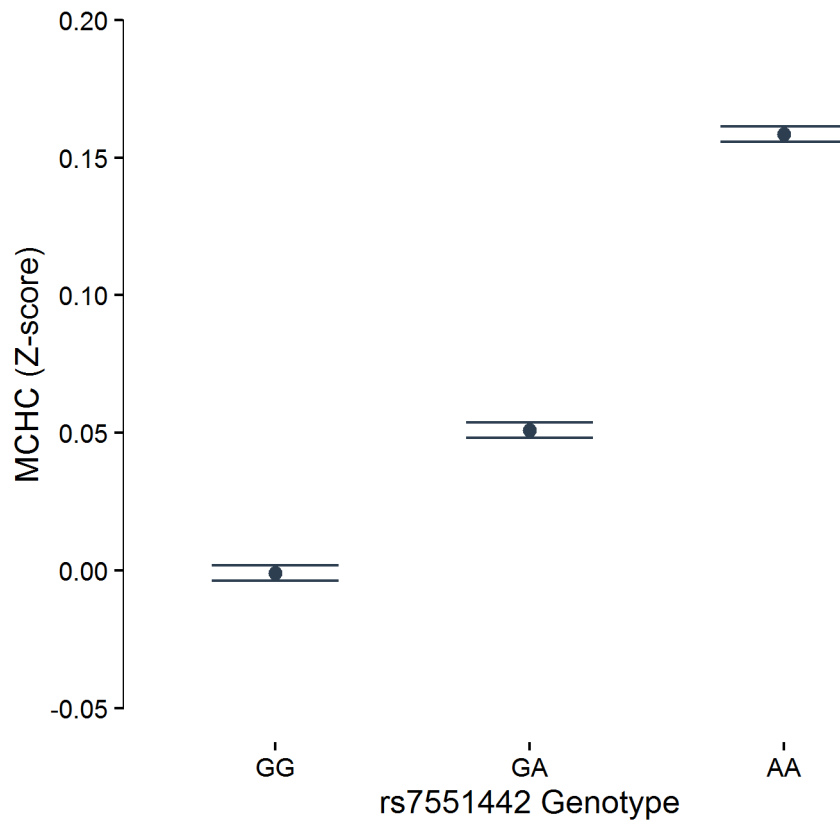
**Figure S5.**

*ATP2B4* allelic imbalance in human erythroblasts. Ratio of reads of the alternate allele over the total number of reads overlapping *ATP2B4* exonic SNPs. In the absence of allelic imbalance and read mapping biases, the expected ratio is 0.5. All heterozygous samples at these different exonic SNPs are also heterozygotes for the top *ATP2B4* candidate eQTLs.



**Figure S6.**

Erythroblasts show an eQTL signal that is independent from *ATP2B4* eQTLs found in GTEx. The bottom track shows pairwise linkage disequilibrium (LD,  $r^2$ ) between rs10751452 and surrounding SNPs. There is essentially no LD between the *ATP2B4* erythroblast eQTL (rs10751452) and eQTLs found in esophagus mucosa from the GTEx resource.



**Figure S7.**

Association of rs7551442 with mean corpuscular hemoglobin concentration (MCHC) in the UK Biobank. Dots represent the mean normalized MCHC by genotype, while the error-bars represent the standard errors.





**Figure S8.**

ENCODE and Roadmap Epigenomic Consortia DNase I hypersensitive sites (DHSs) signals at the *ATP2B4* locus.

**Table S1 (separate file).**

Genes with significant allelic imbalance in human erythroblasts. For each sample, we report exonic SNPs displaying significant allelic imbalance along with their overlapping genes. REF: Reference allele of the SNP. ALT: Alternative allele. REF\_count: Number of RNA-sequencing reads with the reference allele. ALT\_count: Number of reads containing the alternative allele. Total\_count: Total number of overlapping reads. Ratio: Ratio of reads with the reference allele over the total number of reads. P: Binomial test P-value (see Methods for details).

**Table S2 (separate file).**

Significant eQTLs around genes with allelic imbalance in human erythroblasts. eQTL SNPs correspond to candidate regulatory SNPs, whereas AI SNPs correspond to the exonic SNP which we used to test for allelic imbalance. N hets: Number of samples that are heterozygote for the eQTL SNP and for the AI SNP. These samples should display AI if the eQTL SNP is a real regulatory variant. N homs: Number of samples that are homozygote for the eQTL SNP, but heterozygote for the AI SNP. These samples should not display AI. Mean hets: Mean AI ratio of the previously defined heterozygotes. Mean homs: Mean AI ratio of the previously defined homozygotes. AI P: one-sided t-test P-value between homozygotes and heterozygotes. LM Beta: Beta statistic of the linear regression model between the eQTL SNP and gene expression. LM P: P-value of the linear regression model. Combined P: Combined P-values of the linear regression model and AI ratio t-test (see Methods for details). The combined P-value corresponds to the linear regression P-value when it was not possible to test for the difference in AI ratios due to a limited number of samples. q-value: False discovery rate-adjusted P-value. Associated Blood Index corresponds to SNP associated with red blood cell trait from Astle *et al.*, *Cell*, 2016. RDW: Red cell distribution width; MCH: Mean corpuscular hemoglobin; MCV: Mean corpuscular volume; IRF: Immature fraction of reticulocytes; MCHC: Mean corpuscular hemoglobin concentration; HGB: Hemoglobin; HLSR#: High light scatter reticulocyte count; HLSR%: High light scatter reticulocyte percentage of red cells; RBC#: Red blood cell count.

**Table S3 (separate file).**

Genes with erythroblast-specific eQTLs. This table includes genes without any shared eQTL in GTEx. See also Figure 1D.

**Table S4.**

Top gene ontology terms and mouse phenotypes significantly enriched among erythroblast eGenes.

ID	Name	P-value	FDR B&H	Genes from Input	Genes in Annotation
<b>GO: Molecular Function</b>					
GO:0046977	TAP binding	2.18E-05	1.92E-02	3	7
GO:0016404	15-hydroxyprostaglandin dehydrogenase (NAD+) activity	7.52E-05	2.80E-02	2	2
GO:0005344	oxygen transporter activity	2.17E-04	2.80E-02	3	14
GO:0002060	purine nucleobase binding	2.22E-04	2.80E-02	2	3
GO:0016822	hydrolase activity, acting on acid carbon-carbon bonds	2.22E-04	2.80E-02	2	3
GO:0004351	glutamate decarboxylase activity	2.22E-04	2.80E-02	2	3
GO:0016823	hydrolase activity, acting on acid carbon-carbon bonds, in ketonic substances	2.22E-04	2.80E-02	2	3
<b>GO: Biological Processes</b>					
GO:0002486	antigen processing and presentation of endogenous peptide antigen via MHC class I via ER pathway, TAP-independent	6.59E-07	2.42E-03	3	3
GO:0006779	porphyrin-containing compound biosynthetic process	4.78E-06	4.93E-03	5	29
GO:0019885	antigen processing and presentation of endogenous peptide antigen via MHC class I	5.24E-06	4.93E-03	4	14
GO:0033014	tetrapyrrole biosynthetic process	7.94E-06	4.93E-03	5	32
GO:0006335	DNA replication-dependent nucleosome assembly	9.29E-06	4.93E-03	5	33
GO:0034723	DNA replication-dependent nucleosome organization	9.29E-06	4.93E-03	5	33
GO:0002483	antigen processing and presentation of endogenous peptide antigen	9.39E-06	4.93E-03	4	16
GO:0019883	antigen processing and presentation of endogenous antigen	1.96E-05	9.00E-03	4	19
GO:0001916	positive regulation of T cell mediated cytotoxicity	3.62E-05	1.29E-02	4	22
GO:0048534	hematopoietic or lymphoid organ development	4.19E-05	1.29E-02	21	905
GO:0071103	DNA conformation change	4.27E-05	1.29E-02	11	285
GO:0006778	porphyrin-containing compound metabolic process	4.90E-05	1.29E-02	5	46
GO:1901566	organonitrogen compound biosynthetic process	5.27E-05	1.29E-02	28	1445
GO:0002484	antigen processing and presentation of endogenous peptide antigen via MHC class I via ER pathway	5.28E-05	1.29E-02	3	9
GO:0002480	antigen processing and presentation of exogenous peptide antigen via MHC class I, TAP-independent	5.28E-05	1.29E-02	3	9
GO:0030097	hematopoiesis	6.08E-05	1.40E-02	20	858
<b>Mouse Phenotypes</b>					
MP:0008956	decreased cellular hemoglobin content	8.88E-07	1.22E-03	4	8
MP:0011177	abnormal erythroblast number	1.48E-06	1.22E-03	6	34
MP:0011188	increased erythrocyte protoporphyrin level	1.58E-06	1.22E-03	4	9

MP:0008954	abnormal cellular hemoglobin content	2.62E-06	1.51E-03	4	10
MP:0011176	abnormal erythroblast morphology	4.02E-06	1.85E-03	6	40
MP:0000611	jaundice	4.93E-06	1.90E-03	5	24
MP:0010375	increased kidney iron level	6.07E-06	2.00E-03	4	12
MP:0002642	anisocytosis	8.16E-06	2.23E-03	6	45
MP:0008742	abnormal kidney iron level	8.70E-06	2.23E-03	4	13
MP:0002812	spherocytosis	1.63E-05	3.14E-03	4	15
MP:0004147	increased porphyrin level	1.63E-05	3.14E-03	4	15
MP:0011989	abnormal porphyrin level	1.63E-05	3.14E-03	4	15
MP:0003657	abnormal erythrocyte osmotic lysis	2.18E-05	3.87E-03	5	32
MP:0010163	hemolysis	2.96E-05	4.88E-03	5	34
MP:0010067	increased red blood cell distribution width	4.14E-05	5.73E-03	7	87

**Table S5.**

Comparison of the association of the rs7551442 A-allele with red blood cell (RBC) traits in the first release of the UK Biobank and effect of the *Atp2b4* targeted deletion on RBC phenotypes in mice. In the UK Biobank, effect sizes for the A-allele at rs7551442 (betas and standard errors (SE)) are in standard deviation units. In the mouse, effect sizes for the deletion allele are in the following units: hemoglobin (HGB, g/dL), hematocrit (HCT, ratio), mean corpuscular hemoglobin (MCH, pg), mean corpuscular hemoglobin concentration (MCHC, g/dL), RBC count (RBC,  $10^9/L$ ), reticulocyte count (RETIC, %), RBC distribution width (RDW, %), mean corpuscular volume (MCV, fL).

Trait	UK Biobank			<i>Atp2b4</i> <sup>-/-</sup> mice		
	N	Beta (SE)	P	N	Beta (SE)	P
<b>HGB</b>	128,311	0.034 (0.0066)	$2.1 \times 10^{-7}$	26	0.0557 (0.098)	0.58
<b>HCT</b>	128,311	0.011 (0.0066)	0.098	26	-0.009 (0.003)	$2.6 \times 10^{-3}$
<b>MCH</b>	128,253	0.012 (0.0065)	0.062	26	-0.095 (0.106)	0.38
<b>MCHC</b>	128,224	0.058 (0.0065)	$2.6 \times 10^{-19}$	26	0.893 (0.267)	$2.7 \times 10^{-3}$
<b>RBC</b>	128,312	0.021 (0.0065)	$1.4 \times 10^{-3}$	26	0.099 (0.056)	0.088
<b>RETIC</b>	128,348	0.022 (0.0066)	0.022	26	-0.295 (0.130)	0.033
<b>RDW</b>	128,231	-0.064 (0.0065)	$1.2 \times 10^{-22}$	26	0.409 (0.173)	0.026
<b>MCV</b>	128,311	-0.019 (0.0065)	$2.8 \times 10^{-3}$	26	-1.420 (0.214)	$7.1 \times 10^{-7}$

**Table S6.**  
Single guide RNA design.

<b>Single-guide RNA Name</b>	<b>Single-guide RNA sequence</b>	<b>Genomic cleavage position (hg19)</b>
<b>ATP2B4 sgL1</b>	GGACTACGTAACCGGGGCTG	chr1:203650624
<b>ATP2B4 sgR1</b>	TAAATGGTTACGACCTCTGA	chr1:203651551
<b>ATP2B4 sgL2</b>	TTAGCACAGCACCAGAACAT	chr1:203650637
<b>ATP2B4 sgR2</b>	CACAGGTCCCTCAAATAGAC	chr1:203651526
<b>ATP2B4 sg1</b>	CTCCCTTGCTAACCTTGCTG	chr1:203650938
<b>ATP2B4 sg2</b>	TCTCAATTCAGATAGAGGAT	chr1:203650951
<b>ATP2B4 sg3</b>	ATACCTCTCAATTCAGATAG	chr1:203650956
<b>ATP2B4 sg4</b>	TATCCTCTATCTGAATTGAG	chr1:203650964
<b>ATP2B4 sg5</b>	CATTAAAGTTTCTCTGGAGT	chr1:203650988
<b>ATP2B4 sg6</b>	AGCATTAAGTTTCTCTGGAG	chr1:203650989
<b>ATP2B4 sg7</b>	CCTGAGCATTAAAGTTTCTC	chr1:203650994
<b>ATP2B4 sg8</b>	ACCTCTGAGGAGGGAAGATA	chr1:203651026
<b>ATP2B4 sg9</b>	GGACAGAAGGACCTCTGAGG	chr1:203651036
<b>Non-targeting sg1</b>	CACGGAGGCTAAGCGTCGCAA	
<b>Non-targeting sg2</b>	GCGCTTCCGCGGCCCGTTCAA	
<b>Non-targeting sg3</b>	GATCGTTTCCGCTTAACGGCG	
<b>Non-targeting sg4</b>	GTAGGCGCGCCGCTCTCTAC	
<b>Non-targeting sg5</b>	GCCATATCGGGGCGAGACATG	
<b>Non-targeting sg6</b>	GTAATAACGCCGCTCCTACAG	
<b>Non-targeting sg7</b>	GTGAGGATCATGTTCGAGCGCC	
<b>Non-targeting sg8</b>	GGGCCCGCATAGGATATCGC	
<b>Non-targeting sg9</b>	GTAGACAACCGCGGAGAATGC	
<b>Non-targeting sg10</b>	GACGGGCGGCTATCGCTGACT	

**Table S7.**  
qPCR design.

<b>Forward Primer</b>	<b>Reverse Primer</b>	<b>Description</b>
<i>ATP2B4_enhancer_out_F</i> AGAGGATTGTCAGGAATCGG C	<i>ATP2B4_enhancer_out_R</i> TGA <sup>-</sup> CTCAAGAGAGGGCCCGTT	Deletion amplicon detection
<i>ATP2B4_enhancer_in_F</i> CCCTAGTTAGCATGCGTGAGA	<i>ATP2B4_enhancer_in_R</i> CTTTGGCAGTTGGTGACGCA	Non-deletion (internal) amplicon detection
<i>ATP2B4_enhancer_out_F</i> AGAGGATTGTCAGGAATCGG C	<i>ATP2B4_enhancer_in_R2</i> CAGCATTCTCCCCTCAGAA	Non- inversion amplicon detection
<i>ATP2B4_enhancer_out_F</i> AGAGGATTGTCAGGAATCGG C	<i>ATP2B4_enhancer_in_F2</i> CTAACGGGCTTTGGAGGATTT	Inversion amplicon detection
<i>ATP2B4_enh_core_F</i> TTCTGAGGGGAGAAATGCTG	<i>ATP2B4_enh_core_R</i> AAGAGGCTTTTGGGGAGAAG	Deletion/non- deletion amplicon detection
<i>ATP2B4_enh_core_F2</i> CCCTAGTTAGCATGCGTGAGA	<i>ATP2B4_enh_core_R2</i> AAATCCTCAAAGCCCGTTA G	Amplicon Sanger sequencing
<i>ATP2B4_RT_F1</i> AAAGACCCCATGTTGCTCTC	<i>ATP2B4_RT_R1</i> CCCCTTCGTCATCCTCATTG	RT-qPCR