Aus dem
Institut für Chemie und Biochemie
der Freien Universität Berlin
und der
Division of Hematology/Oncology, Boston Children's Hospital,
Harvard Medical School, Boston

DISSERTATION

Functional Studies of Genetic Variants in Human Erythropoiesis

zur Erlangung des akademischen Grades
Doctor medicinae (Dr. med.)

vorgelegt der Medizinischen Fakultät
Charité – Universitätsmedizin Berlin

von
Leif Si-Hun Ludwig
aus Berlin

Datum der Promotion: 25.06.2017

Thesis Summary

1. Abstract

Introduction

Genome-wide association and exome sequencing studies have identified numerous genetic variants associated with human traits and disease. In many instances, a detailed understanding of the underlying molecular mechanisms of these associations is lacking. In this study, we functionally followed up on common and rare genetic variants associated with human red blood cell traits, anemia and erythrocytosis. First, we followed up on genetic population studies associating common variants near the cyclin A2 (*CCNA2*) locus to red blood cell traits and investigated the thus far undefined role of *CCNA2* in erythropoiesis. Next, we identified a rare coding mutant (Y365C) in the X-chromosomal *ALAS2* gene, encoding for 5'-aminolevulinate synthase 2, in a single affected family presenting with macrocytic anemia and dyserythropoiesis and studied the functional effects of the mutant. Finally, we utilized insight from studies on common and rare genetic variation to determine that coding variants in *SH2B3,* a negative regulator of cytokine signaling, are associated with erythrocyte counts and explored the potential of this finding to improve the process of *in vitro* red blood cell production.

Methods

In vitro culture models were utilized to study associated genes in primary erythroid cells derived from murine fetal liver or healthy human donors. Patient derived cells, shRNA-mediated knockdown and genome editing approaches were utilized to study the consequences of disrupting each gene of interest. Cultured cells were characterized using a variety of methods, including flow cytometry assays, microscopic imaging, western blotting, quantitative PCR and bioinformatic analysis.

Results

Careful analysis of a murine *Ccna2* knockdown revealed an important role for this cell cycle gene in cytokinesis during terminal erythropoiesis, which led to the production of red blood cells with increased size. Characterization of the Y365C mutation showed disrupted pyridoxine cofactor binding, destabilization and resulting loss of function of the ALAS2 enzyme. Finally, genetic disruption of *SH2B3* expression in stem and progenitor cells increased erythroid proliferation, maturation and ultimately production by 3-7 fold.

Conclusions

By following up on genetic variants associated with erythroid phenotypes, we have revealed an unrecognized role of cyclin A2 in cytokinesis and characterize a new loss of function mutation in *ALAS2*. Furthermore, targeted disruption of *SH2B3* may provide a useful avenue to enhance red blood cell production for cell replacement therapies. Together these findings illustrate how functional studies of genetic variants can result in important applications and provide a deeper understanding of human variation in physiology and disease.

1.1 Abstract (German)

Einleitung

Genomweite Assoziationsstudien und die vermehrte Anwendung von Exom-Sequenzierung haben zur Entdeckung einer Vielzahl von neuen und mit humanen Merkmalen und Krankheiten assoziierten genetischen Varianten geführt. In vielen Fällen fehlt jedoch ein über die Assoziation hinausgehendes detailliertes Verständnis der zugrunde liegenden molekularen Mechanismen. Im Rahmen dieser Arbeit haben wir eine Reihe von mit Erythrozyten-Parametern und mit Anämie sowie Erythrozytose assoziierte genetische Varianten funktionell untersucht. Zunächst haben wir auf Grundlage von Populationsstudien, welche häufige Polymorphismen im cyclin A2 (*CCNA2*) Lokus mit der Größe von roten Blutkörperchen assoziiert haben, die bisher undefinierte Rolle von *CCNA2* in der Erythropoese untersucht. Weiterhin haben wir eine seltene Mutation (Y365C) im X-chromosomalen für die 5-Aminolävulinatsynthase kodierenden *ALAS2* in einer Familie mit makrozytärer Anämie und Dyserythropoese entdeckt und die funktionellen Effekte der Mutante charakterisiert. Schließlich haben wir aufgrund von mit der Erythrozytenzahl assoziierten häufigen und seltenen genetischen Varianten im Gen *SH2B3*, einem negativen Regulator der Zytokin Signaltransduktion, die Möglichkeit der Inaktivierung von *SH2B3* zur Optimierung der *in vitro* Produktion von roten Blutkörperchen exploriert.

Methodik

Für die Untersuchung der assoziierten Gene wurden *in vitro* Kulturmodelle primärer erythropoetischer Zellen aus der murinen fetalen Leber und von humanen Spendern eingesetzt. Um den Funktionsverlust und dessen Konsequenzen der relevanten Gene zu ermitteln wurden shRNA-vermittelter Knockdown bzw. Genome Editing angewendet und von Patienten isolierte Zellen genutzt. Die Charakterisierung der kultivierten Zellen erfolgte durch eine Reihe an Methoden, einschließlich durchflusszytometrischer Verfahren, Lichtmikroskopie, Westernblot, quantitativer PCR und bioinformatischer Analysen.

Ergebnisse

Die Charakterisierung des Knockdowns von murinem *Ccna2* offenbarte eine wichtige Funktion dieses Zellzyklus-Regulators in der Zytokinese im Rahmen der terminalen Erythropoese und führte zur Produktion von größeren roten Blutkörperchen. Die Untersuchung der Y365C Mutante zeigte eine gestörte Bindung des Kofaktors Pyridoxins,

sowie Destabilisierung und Funktionsverlust des ALAS2-Enzyms. Die Inaktivierung der *SH2B3* Expression in humanen Stamm- und Vorläuferzellen begünstigte die Proliferation sowie Differenzierung von erythropoetischen Zellen und erhöhte letztlich die Produktion von roten Blutkörperchen um das drei- bis siebenfache.

Schlussfolgerung

Mittels funktioneller Studien von genetischen Varianten haben wir eine bisher unbekannte Rolle von cyclin A2 im Rahmen der Zytokinese offenbart und charakterisieren eine neue Funktionsverlustmutante in *ALAS2*. Weiterhin stellt die Inaktivierung von *SH2B3* eine praktische Möglichkeit zur Steigerung der Produktion von roten Blutkörperchen für Zellersatzverfahren und ähnliche Therapien dar. Zusammen zeigen diese Studien, wie die funktionelle Untersuchung genetischer Varianten und menschlicher Variation zur Entwicklung von neuen Anwendungen und zu einem tieferen Verständnis von Physiologie und Krankheit beitragen kann.

2. Introduction

2.1 The human genome and studies of genetic variation

The initial publication of the sequence of the human genome marked a milestone for the understanding of human biology and evolution [1-3]. Numerous projects followed aiming to more fully comprehend the information embedded in the genome, describe patterns and establish a detailed database of human genetic variation to identify genetic variants affecting human physiology and disease [1]. The Encyclopedia of DNA Elements (ENCODE) and Roadmap Epigenomic projects aimed to identify functional elements in the genome and characterize modifications in chromatin and DNA and their effects on gene expression and other functions of the genome [4-6]. The International HapMap and the 1000 Genomes Project applied single nucleotide polymorphism (SNP) genotyping and whole-genome sequencing of individuals from multiple populations to characterize human genetic variation and create a resource to study genotype and phenotype relationships and their role in human disease [7-9]. Upon completion of sequencing the genomes of 2,504 individuals from 26 populations, a total number of over 88 million variants were revealed [10]. To study their effects on human physiology genome-wide association studies (GWAS) utilized these comprehensive catalogues of common genetic variants in populations to identify variants associated with a specific trait or disease [1, 11, 12]. As of December 14, 2015 the GWAS catalog lists over 23,000 variants that have been associated with various phenotypes (https://www.ebi.ac.uk/gwas/home). The Genotype-Tissue Expression (GTEx) project presents another approach aiming to understand the functional consequences on human gene expression and regulation due to genetic variation [13]. Additionally, the targeted sequencing of protein coding sequences (exome) further fueled the identification of rare genetic variants in Mendelian disease, caused by mutations in a single locus or gene [14, 15]. To date, the Online Mendelian Inheritance in Man (OMIM) database encompasses over 23,000 entries and over 5,600 phenotypes with a known molecular basis (www.omim.org, December 14, 2015).

For the majority of these genetic variants implicated in human traits and disease the molecular mechanisms fundamental to their associations have not been thoroughly established. This represents a significant challenge in order to fully understand the underlying biology and pathophysiology and eventually harness the embedded insight for the development of targeted therapies and other applications. Causal rare variants in the context of Mendelian disorders usually have large effect sizes, thus individually resulting in disease

[16], and the clinical implications may appear more straightforward with respect to the subsequent development of targeted drugs or gene-based therapies. In contrast, the utility of the study of common variants as identified by GWAS has been met with mixed responses [17, 18]. Most common variants have rather small effect sizes with respect to a specific trait or phenotype, explaining their broad prevalence compared to rare disease-causing variants that are more likely to be selected against during evolution [16]. Thus, the significance of common variants and their contribution to disease phenotypes has been debated, also in light of the large number of individuals that have to be genotyped to identify meaningful associations [17, 18]. The majority of GWAS-derived SNPs are located in non-coding regions, many of which appear to regulate transcription of neighboring genes [6, 19]. However, the phenotypic effects of such modulations may not need to correlate with the pathophysiological and potential therapeutic significance of the linked gene, as more extreme perturbation may reveal dramatic phenotypes and insight into previously unappreciated pathways [1, 18, 20]. Prominent examples include common variants in the *HMGCR* and *BCL11A* loci that have modest effect sizes, but their regulated genes are major targets for the cholesterol-lowering statins or for therapeutic efforts aimed at modulating the expression of fetal hemoglobin to treat human ß-chain related hemoglobinopathies, respectively [21-23]. Furthermore, additive effects or the epistatic interplay of multiple variants and their contribution to disease manifestation and modulation has only been delineated in selected cases and for the most part remains incompletely understood [1, 24].

2.2 Genetic studies in erythropoiesis

Erythropoiesis is the process of red blood cell or erythrocyte production, characterized by the proliferation and maturation of lineage-committed erythroid precursor cells through defined stages [25, 26]. Red blood cells are carriers of hemoglobin, the major transporter of oxygen in the peripheral blood, and a pathological decrease in the number of circulating erythrocytes is known as anemia [25, 26]. While most instances of anemia are caused by nutritional deficiencies, chronic kidney disease or infectious agents, a large number of genetic lesions are known that affect erythroid phenotypes and cause anemia [25]. These include the prevalent hemoglobinopathies sickle cell disease and the thalassemias, metabolic enzyme defects predisposing to hemolytic anemia, such as in glucose-6-phosphate dehydrogenase deficiency, red blood cell membrane defects as observed in hereditary spherocytosis, and disorders affecting red blood cell production like Diamond-Blackfan anemia or congenital

dyserythropoietic anemia [25, 27]. Yet, the wider availability of whole-exome sequencing approaches in the clinic has led to the identification of novel rare anemia-associated genes and subtypes, followed by additional studies revealing hitherto unappreciated pathophysiological mechanisms and phenomena [28-32]. Moving forward, similar approaches will almost certainly result in the identification of previously unrecognized disease genes and molecular pathways. Similarly, recent GWAS have identified over 75 loci associated with red blood cell traits, including red blood cell count (RBC) and red blood cell size or mean corpuscular volume (MCV) as well as variants associated with fetal hemoglobin levels [11, 12]. Functional studies have subsequently provided new molecular insight into these associations and provided a deeper understanding of homeostatic and pathophysiologic mechanisms in erythropoiesis and importantly have generated novel therapeutic approaches in a number of cases [22, 23, 33, 34]. However, most of the GWAS-derived loci remain uncharacterized and thus provide a valuable resource and starting point for future functional studies.

2.3 Aims of the projects and thesis

The aims of the thesis were to follow up on three independent genetic observations in hematopoiesis and erythropoiesis to ultimately gain a deeper molecular insight into the underlying mechanisms and pathophysiology and explore the potential of biological applications. The genetic findings and related studies will be introduced and discussed below along with the specific questions and aims set for the individual projects. The specific contributions of the author of the thesis are stated in section 8 (page 29-30).

2.3.1 GWAS associate the cyclin A2 locus with red blood cell size

One of the loci associated with MCV by GWAS contained the gene *CCNA2*, encoding for cyclin A2 [11]. Cyclin A2 is ubiquitously expressed in all proliferating cells and by binding to and activating its catalytic partners, the cyclin dependent kinases CDK1 and CDK2, it has been considered a key driver of S phase progression and mitosis [35, 36]. Targeted homozygous deletion in mice resulted in early embryonic lethality and underlined its key role as a core component of the cell cycle machinery [35]. Conditional deletions in mice confirmed its essential role in embryonic stem cells and in the hematopoietic stem cell compartment, where its ablation led to pancytopenia and soon death [36]. Cyclins have previously been implicated by GWAS to affect red blood cell count and size [11]. Subsequent studies identified a common variant that modulates cyclin D3 levels, genetic perturbation of

which affected the number of cell divisions in erythropoiesis and ultimately impacted the number and size of red blood cells produced [33]. However, a specific role for cyclin A2 in erythropoiesis and how common genetic variation in the locus can affect red blood cell traits like MCV has not previously been explored. The major aims of this project were to

1. Characterize the phenotype and possible role of cyclin A2 in erythropoiesis upon knockdown in an *in vitro* culture model of murine erythropoiesis.
2. Explore a connection to the initial genetic association with red blood cell size.

2.3.2 An *ALAS2* mutation in a family with macrocytic anemia and dyserythropoiesis

Extensive clinical and initial genetic evaluation of a family presenting with macrocytic anemia and dyserythropoiesis remained unrevealing and eventually family members were forwarded to a whole-exome sequencing study to identify a possible genetic cause of the anemia. The analysis revealed a previously unreported Y365C mutation in all affected females in the X chromosomal *ALAS2* gene. *ALAS2* encodes for 5'-aminolevulinate synthase 2, which is the first and rate-limiting enzyme of the biosynthesis of heme, an iron containing cofactor of hemoglobin essential for binding oxygen [37]. While the related *ALAS1* gene is widely expressed, *ALAS2* is the erythroid-specific isoform, loss of function mutations in which typically present as a hypochromic, microcytic sideroblastic anemia in male carriers or in females with highly skewed X inactivation [37, 38]. In contrast, the Y365C mutation identified in the proband was associated with macrocytic anemia and dyserythropoiesis, which has only rarely been observed in patients carrying lesions in the *ALAS2* locus, and was more reminiscent of the congenital dyserythropoietic anemias (CDA) [27, 39]. However, targeted sequencing of genes associated with CDA indicated no causal mutations. Additionally, initial bone marrow evaluation did not reveal any blasts with siderotic granules that are typically associated with this disorder [37], and thus the relevance and whether the identified mutation is indeed causal for the observed phenotype remained unclear. The major aims of the project were to

1. Characterize the Y365C mutant by structural modeling, protein expression and evaluation of enzymatic activity.
2. Assess to what extent the observed phenotype could be due to extensively skewed X inactivation.
3. Evaluate the expression of the wildtype and mutant transcripts over time in an *in vitro* culture system utilizing primary proband cells.

2.3.3 Coding variants in *SH2B3* are associated with red blood cell count and erythrocytosis

GWAS identified a common coding SNP in the *SH2B3* gene resulting in a R262W substitution and is associated with red blood cell count and hemoglobin levels in humans [11]. Similarly, rare loss of function variants have been identified in patients with erythrocytosis, presenting with even greater elevations in the red blood cell count and hemoglobin levels [40, 41]. SH2B3 is a member of a family of adaptor proteins, binding to phosphotyrosines and localizing to the plasma membrane via its src homology 2 (SH2) and pleckstrin homology (PH) domains, respectively [42, 43]. By binding to cytokine receptors, SH2B3 is thought to negatively regulate receptor signaling, as loss of SH2B3 has been shown to modulate sensitivity to cytokines and proliferative capacity of hematopoietic progenitor cells [42, 44, 45]. Interestingly, mice deficient in *SH2B3* do not present with an ostensible erythroid phenotype, suggesting differences in murine and human erythroid physiology [44]. In light of these observations, we explored the possibility that suppression of *SH2B3* could increase the efficiency of the *in vitro* production of human red blood cells. The use of stem cells for cellular replacement therapies has drawn increasing interest in the past years as certain predictions suggest a future shortage in blood availability due to the increasing demands of ageing populations in the world [46]. Various efforts have been put forward to optimize culture conditions or utilize genetic perturbations to improve production of red blood cells, but costs and scale remain significant challenges for these approaches to become feasible alternatives to traditional blood donation and transfusion [46-50]. The major aims of this project were to

1. Explore the effects of knockdown of *SH2B3* in a primary human *in vitro* culture system on proliferation and differentiation.
2. Characterize and compare the quality of *SH2B3* knockdown cells to those derived from control cultures.

3. Methods

Cell culture

293T cells were maintained in DMEM with 10 % fetal bovine serum (FBS), 2 mM L-glutamine, and 1 % penicillin/streptomycin. TF-1 human erythroid cells were cultured in RPMI media supplemented with 10 % FBS and 2 ng/ml granulocyte-macrophage colony-stimulating factor (GM-CSF) for maintenance and 1 % penicillin/streptomycin. For cytokine starvation of TF-1 cells, all cytokines were removed and cells were maintained overnight in media with FBS only. Erythropoietin (EPO) and stem cell factor (SCF) were added at time 0 min and cells were collected at indicated time points for analysis.

Mouse fetal liver erythroid progenitor purification and *in vitro* culture

E14.5-15.5 murine fetal liver cells were homogenized in PBS supplemented with 2 % FBS and 100 mM EDTA. Mature erythrocytes were lysed by the addition of ammonium chloride solution (StemCell Technologies, Inc.). After washing, the remaining cells were incubated with a cocktail of biotin-conjugated antibodies, including Lineage Cocktail (BD 559971), Ter119 (eBioscience 13-5921-85), CD16/32 (Abcam 25249), Sca-1 (BD553334), CD34 (MCA1825B), and CD41 (MCA2245B). After magnetic depletion with streptavidin beads (BD 557812), a pure fetal liver Ter119-negative erythroid progenitor population was obtained [51]. For differentiation, primary murine cells were resuspended in IMDM containing 15 % FBS, 1 % bovine serum albumin, 1 % penicillin/streptomycin, 2 mM L-Glutamine, 500 µg/ml holo-transferrin, 10 µg/ml insulin and 0.5 U/ml EPO (Amgen) for up to 66 h at 37 °C, 5 % CO_2 [33].

***In vitro* culture of human stem and progenitor cells**

$CD34^+$ cells from G-CSF mobilized peripheral blood, bone marrow, or cord blood were differentiated into erythroid cells as described [31], or utilizing a three-phase culture protocol: In phase 1 (day 0-7), cells were cultured at a density of 10^5-10^6 cells/ml in IMDM supplemented with 2 % human AB plasma, 3 % human AB serum, 1 % penicillin/streptomycin, 3 IU/ml heparin, 10 µg/ml insulin, 200 µg/mL holo-transferrin, 1 IU EPO, 10 ng/ml SCF and 1 ng/ml IL-3. In phase 2 (day 7-12), IL-3 was omitted from the medium. In phase 3 (day 12-18), cells were cultured at a density of 10^6 cells/ml, with both IL–3 and SCF omitted from the medium and the holo-transferrin concentration was increased to 1 mg/ml. Alternatively, an expansion phase lasting a total of 5 days prior to initiation of

erythroid differentiation and starting phase 1 of the culture was added, similar to what has been described [52]. The expansion medium was composed of StemSpan II serum free expansion medium (Stem Cell Technologies) and 1X CC100 cytokine mix composed of FLT3L, IL-3, IL-6, and SCF (Stem Cell Technologies), as described previously [22, 53]. Media was changed every 2-3 days and cells were maintained at a density between 10^5-10^6 cells/ml. During the expansion phase >85% of cells remain $CD34^+$. All cells were cultured at 37 °C and 5 % CO_2.

Retroviral and lentiviral constructs, production and infection of primary cells

The shRNA sequences targeting murine *Ccna2* and human *SH2B3* were obtained from the RNAi Consortium of the Broad Institute (http://www.broadinstitute.org/rnai/trc). *Ccna2* targeting shRNAs were cloned into the BbsI restriction sites of the linearized MSCV-pgkGFP-U3-U6P retroviral vector, which co-expresses GFP driven by the PGK promoter and had the following sequences:

sh4 - AAAAGTTAATGAAGTACCTGACTATGTCGACATAGTCAGGTACTTCATTAAC

sh5 - AAAAGCTTCGAAGTTTGAAGAAATAGTCGACTATTTCTTCAAACTTCGAAGC

SH2B3 targeting shRNAs were in the pLKO.1-puro lentiviral vector, co-expressing a puromycin resistance cassette and had the following sequences:

sh83 - CCGGCCTGACAACCTTTACACCTTTCTCGAGAAAGGTGTAAAGGTTGTCAG GTTTTTG

sh84 - CCGGGCCTGACAACCTTTACACCTTCTCGAGAAGGTGTAAAGGTTGTCAGG CTTTTTG

For retroviral and lentiviral production 293T cells were transfected with the appropriate viral packaging and genomic vectors using FuGene 6 reagent (Promega) according to the manufacturer's protocol. For retroviral production and infection of primary murine erythroid cells, the retroviral constructs described above along with the pCL-eco packaging vector were used. For lentiviral production and infection of primary human cells, the constructs described above along with the VSV-G envelope and pΔ8.9 packaging vector were used. The culture media was changed the day after transfection and viral supernatant was collected and filtered at 48 h post-transfection. Primary erythroid cells were mixed with viral supernatant and polybrene, before spinfection at 2000 rpm for 90 min. Depending on the utilized vectors, infected cells were identified by GFP expression or selected with 1 µg/ml puromycin for 30-48 h.

Flow cytometry analyses

For flow cytometry analysis, cells were washed in PBS and stained with 7AAD or propidium iodide (PI) for dead cell exclusion, and with one or multiple of the antibodies as indicated: APC-conjugated anti-mouse Ter119 (17-5921-83; eBioscience), PE-conjugated anti-mouse CD71 (12-0711-83; eBioscience), APC conjugated anti-human CD235a (17-9987, eBioscience), FITC conjugated anti-human CD71 (11-0719, eBioscience), PE conjugated anti-human CD41a (12-0419, eBioscience), and PE conjugated anti-human CD11b (12-0118, eBioscience). Where indicated, cells were additionally stained with 1 μg/mL Hoechst 33342 (Life Technologies). For Eosin-5-maleimide (EMA) binding, cells were washed with PBS and subsequently labeled with 0.5 mg/ml EMA eosin-5-maleimide (Life Technologies), as described previously (King et al., 2000). Cells were subsequently forwarded to analysis on a LSR II or Canto II flow cytometer (BD Bioscience) [54]. Data were analyzed using FlowJo v10 (Tree Star).

Cell cycle and division analyses, phospho-histone H3 staining

For cell cycle analysis, cells were pulsed with 5-ethynyl-20-deoxyuridine (EdU) and EdU incorporation was detected using an EdU flow kit (C10418; Invitrogen) as described by the manufacturer's protocol. PI was added to stain for DNA content after RNAse digestion. The PI signal data were acquired on a linear scale.

To allow tracking of the number of cell divisions cells were labeled with the PKH26 red fluorescent cell linker kit (PKH26GL-1KT; Sigma-Aldrich) as described previously [33]. An aliquot of the labeled cells was used to measure the mean fluorescence intensity of PKH26 immediately after labeling. The number of cell divisions was calculated as described previously [33].

For phospho-histone H3 staining, erythroid cells were fixed and permeabilized using reagents from an EdU flow kit (C10418; Invitrogen). For staining an antiphospho-histone H3 rabbit monoclonal antibody (Ser10; clone MC463; Millipore) and a secondary antibody donkey anti-rabbit AlexaFluor647 (711-605-152; Jackson Labs) were used. Anti-GFP-FITC antibody (ab6662; Abcam) was used to identify GFP-positive cells. PI was added to stain for DNA content after RNAse digestion. The PI signal data were acquired on a linear scale.

May-Grünwald–Giemsa staining

50,000–200,000 cells were centrifuged on to poly-L-lysine–coated slides and stained with May-Grünwald–Giemsa as described previously [33]. Then, slides were mounted with coverslips and examined. Stained cells were captured, processed, and analyzed using Axiovision Microscopy Software (Carl Zeiss).

Quantitative RT-PCR

Isolation of RNA was performed using the miRNeasy Mini Kit (Qiagen). An on-column DNase (Qiagen) digestion was performed according to manufacturer's instructions. RNA was quantified by a NanoDrop spectrophotometer (Thermo Scientific). Reverse transcription was carried out using the iScript cDNA Synthesis Kit (Bio-Rad). Real-time PCR was performed using the ABI 7900 Machine Real-Time PCR System and SYBR Green PCR Master Mix (Applied Biosystems). The following primers were used for quantitative RT-PCR:

Ccna2 forward: 5'-TGGATGGCAGTTTTGAATCACC-3'
Ccna2 reverse: 5'-CCCTAAGGTACGTGTGAATGTC-3'
Ubc forward: 5'-GAGTTCCGTCTGCTGTGTGA-3'
Ubc reverse: 5'-CCTCCAGGGTGATGGTCTTA-3'

Western blotting

Cells were harvested at indicated timepoints and processed as previously described [33]. After SDS gel electrophoresis and Western blotting, membranes were blocked with 3 % BSA-PBST and probed with primary antibodies as indicated. After washing membranes were incubated with appropriate HRP-conjugated secondary antibodies, followed by washing and incubation for 1 min with Western Lightning Plus-ECL substrate (Perkin Elmer). Proteins were visualized by exposure to scientific imaging film (Kodak).

***In silico* analyses of cyclin A2 gene regulation and expression**

Human and murine cyclin A2 mRNA expression patterns were obtained from publicly available microarray and RNA sequencing data [55-58]. Compiled GATA1 occupancy and nucleosome depleted region data were obtained and analyzed as described [59, 60]. Expression data were analyzed as previously described [55-59].

Statistical analysis

All pairwise comparisons were assessed by an unpaired two-tailed Student's t-test, unless indicated otherwise. Differences were considered significant if the P value was below 0.05.

Additional methods

A more detailed description and additional methods related to experiments and results primarily conducted at collaborating laboratories can be found in the supplementary information of the respective publications as printed in the thesis or in the online version of the manuscripts:

http://dx.doi.org/10.1016/j.stem.2015.09.015

http://dx.doi.org/10.1172/JCI78619

4. Results

References to figures in the following paragraphs correspond to the respective figures in the published manuscripts.

4.1 Genome-wide association study follow-up identifies cyclin A2 as a regulator of the transition through cytokinesis during terminal erythropoiesis

GWAS identified common genetic variants within the cyclin A2 locus to be associated with red blood cell size (MCV) [11], but the underlying biology of this association remained elusive. Gene expression and chromatin immunoprecipitation analysis revealed high expression of cyclin A2 in human and murine erythroid cell populations and potential regulation by the erythroid transcription factor GATA1 (Figure 1) [61]. Two independent short hairpin RNAs (shRNAs) expressed by retroviral vectors were utilized to reduce the levels of cyclin A2 in a primary murine *in vitro* erythroid culture system (Figure 2A,B) [51]. Erythroid differentiation as assessed by surface marker expression of CD71 and Ter119 appeared unaffected and measurement of cell size by flow cytometry and light microscopy revealed no significant differences by 24 h of culture post infection of shRNA and control constructs (Figure 2C, 3A,C). At 48 h the number of enucleated Hoechst negative cells was decreased by 1.4- to 2-fold (Figure 2C), while the size of nucleated and enucleated cells was increased in shRNA-infected cells compared to the shluc control (Figure 3B,D). This suggested that modulation of cyclin A2 expression levels led to perturbations during the late stages of erythroid differentiation resulting in differences in cell size. Moreover, this finding is consistent with the notion that common genetic variants in the cyclin A2 locus affect red blood cell size in humans.

Given the primary role of cyclin A2 during the cell cycle [35], the number of cell divisions and the progression through the cell cycle were analyzed to gain a better understanding of the underlying mechanisms of perturbation of cyclin A2. Knockdown of cyclin A2 showed no significant differences in the number of average cell divisions by 24 h, but resulted in a reduction of 0.3 to 0.8 of the number of average cell divisions by 48 h of culture as examined by PKH26 membrane labeling of erythroid progenitor cells (Figure 4A). Similarly, concomitant measurement of incorporation of the EdU nucleoside into newly synthesized DNA revealed no differences at 24 h (Figure 4B), but showed a significant

increase in the frequency of cells in the G2/M phase (40-60 % in knockdown vs. approximately 25 % in control cultures) and simultaneous reduction of cells in the G1 phase of the cell cycle for cyclin A2 knockdown cells after 48 h of culture (Figure 4C). Microscopic and phospho-Histone 3 staining, a marker of mitosis (before anaphase) was used to distinguish at which stage within the G2/M phases the accumulation may occur [62]. While phospho-Histone 3 staining showed no significant differences between cyclin A2 knockdown and control cells (Figure 4D), microscopic imaging revealed a significant increase in the number of bi-nucleated orthochromatic erythroblasts upon knockdown of cyclin A2 at 48 h of culture, many of which appeared to continue to enucleate (Figure 4E,F). These observations suggest that reduction of cyclin A2 does not interfere with initial progression through G2 phase and mitosis, but perturbs completion of cell division by interfering with cytokinesis. As a result, erythroid cells with reduced levels of cyclin A2 remain larger in size, but as they continue to enucleate this results in the production of larger than normal reticulocytes.

4.2 X-linked macrocytic dyserythropoietic anemia in females with an *ALAS2* mutation

Whole-exome sequencing has led to the identification of a previously unreported rare variant in the X-chromosomal *ALAS2* gene in a female proband presenting with macrocytic dyserythropoietic anemia and evidence of iron overload. Multiple female relatives of the proband also exhibited macrocytic anemia. A heterozygous A-to-G variant in *ALAS2* was identified in all affected family members, all of which were female (Figure 1). This variant results in a coding change of Y365C in the ALAS2 protein, and structural modeling revealed this region to be critical for binding of the essential cofactor pyridoxal 5'-phosphate (PLP) (Figure 2A). Accordingly, the recombinantly expressed and purified ALAS2 C365 mutant showed a 70-fold lower affinity for PLP than the wildtype enzyme (Table 1). Furthermore, the mutant enzyme was unstable (Figure 2B,C), consistent with the observation that PLP plays a fundamental role in stabilizing ALAS2 [63, 64], and supporting the notion that the C365 mutant enzyme is severely impaired in PLP binding.

To investigate the possibility of skewed X inactivation, a human androgen receptor gene polymorphism assay (HUMARA) of peripheral blood mononuclear cells was utilized and revealed 62-82 % of the active X chromosomes in the affected family members to contain the *ALAS2* mutant (Figure 2D). The degree of skewing appeared to correlate with the phenotype, as the proband most severely affected showed a greater extent of skewing towards

the mutant allele (82 %, Figure 2D). The severity in impairment of PLP binding to the mutant enzyme suggested that erythroid cells expressing the ALAS2 C365 mutant would not differentiate properly as functional ALAS2 is essential for heme biosynthesis, an essential component of hemoglobin and other hemeproteins [38]. In fact, sanger sequencing of cDNA derived from mature reticulocytes from the peripheral blood of the proband showed only the wildtype allele (Figure 2E), supporting the notion that *ALAS2* loss of function cells do not fully mature [65]. To confirm these observations, peripheral blood mononuclear cells of the proband were cultured *in vitro* under conditions favoring erythroid differentiation. As cells transitioned from the early progenitor towards more terminally differentiated erythroid cells, there was a loss of cells expressing the mutant over the course of the culture as assessed by sequencing of cDNA derived from the cultured cells at indicated timepoints (Figure 3A). These findings support the notion that erythroid cells expressing the ALAS2 C365 mutant are not viable and together with the clinical presentation, suggest that the remaining erythroid cells expressing the wildtype allele of *ALAS2* are unable to fully compensate for the presence of erythroid cells expressing the mutant.

4.3 Targeted application of human genetic variation can improve red blood cell production from stem cells

Various genetic studies identified common and rare coding variants within the *SH2B3* gene to be associated with an increased red blood cell count (RBC) and erythrocytosis, respectively [11, 40, 41]. Similarly, analysis of a population of 4,678 individuals subjected to whole-exome sequencing for rare putative damaging or loss of function variants in *SH2B3* led to the identification of numerous coding variants associated with higher hematocrit and hemoglobin values (Figure 1A,B). Given these associations, we evaluated whether targeted disruption of *SH2B3* expression may enhance erythropoiesis *in vitro*. Lentiviral vectors expressing two independent shRNAs were applied to reduce expression of SH2B3 protein in a culture system of primary human hematopoietic stem and progenitor cells (Figure 1C,D). Reduction of *SH2B3* resulted in improved erythroid maturation as suggested by surface marker phenotyping and cell morphology and led to a significant 1.6- to 2-fold increase in the frequency of enucleated (Hoechst negative) and $CD235a^+$ red blood cells (Figure 1E-H). Gene expression analysis by microarrays of *SH2B3* knockdown erythroblasts revealed an overall similar global expression profile with enhanced expression of genes associated with terminal

erythroid maturation compared to control cells (Figure 1I,J). Improved erythroid maturation upon reduction of *SH2B3* was accompanied by an augmentation in proliferative capacity and an increase in the number of cell divisions as measured by PKH26 membrane labeling during differentiation of erythroid progenitors cells (Figure S1A,D,E). Overall a 3- to 7-fold increase in the absolute number of red blood cells produced was observed, regardless of the source of progenitor cells (adult mobilized peripheral blood or cord blood) or culture conditions applied (Figure 2A-C). This observation is consistent with previous studies that associated genetic variants in *SH2B3* with an increase in red blood cell count and hemoglobin values.

SH2B3 negatively regulates signaling downstream of multiple receptors associated with hematopoiesis [45]. Accordingly, reduction of *SH2B3* expression enhanced phosphorylation of the KIT receptor and phosphorylation of STAT5, downstream of the erythropoietin receptor (Figure S1F). Thus, loss of *SH2B3* appears to improve erythroid maturation by enhancing signaling through the KIT and erythropoietin receptors, favoring or enhancing expression of key erythroid genes (Figure 1I,J, S1G). Importantly, erythroid cells produced in *SH2B3* loss of function cultures showed similar characteristics to control cells, including comparable MCV, hemoglobinization (MCH), hemoglobin subunit expression profiles, eosin-5-maleimide binding, pyruvate kinase activity and membrane protein expression (Figure 2D, S2A-D). Overall, shRNA-mediated reduction of *SH2B3* significantly improved the *in vitro* production of human red blood cells from primary human stem and progenitor cells. However, for large-scale production that could possibly be used for transfusion or other therapeutic purposes, the use of self-renewing cell lines may prove more feasible. Thus, we measured the effects of *SH2B3* ablation in human embryonic stem cells (hESC). Accordingly, isogenic lines with the intact wildtype allele or lines with CRISPR/Cas9 mediated deletions within *SH2B3* (Figure S2E,F) were differentiated *in vitro* to generate hematopoietic progenitor and ultimately erythroid cells (Figure 2E). Ablation of *SH2B3* expression resulted in a more than a 3-fold increase in the yield of erythroid cells compared to control cultures, but importantly showed similar morphology, surface marker and globin gene expression (Figure 2F-I, S2G,H). These experiments provide a proof of principle that *SH2B3* perturbation in self-renewing cells can allow improved erythroid differentiation and expansion.

5. Discussion

5.1 Cyclin A2 regulates the progression through cytokinesis in terminal erythropoiesis

Following up on results of GWAS, which associated noncoding variants in the cyclin A2 locus with red blood cell size [11], our findings illustrate how cyclin A2 regulates the progression through cytokinesis in erythroid cells. In the *in vitro* model utilized here, the reduction of cyclin A2 appeared to primarily affect the final passage through cytokinesis, while previous divisions were not ostensibly compromised. This led to the production of bi-nucleated erythroid cells, which continued to enucleate and ultimately produced larger red blood cells. Thus, regulatory variants modulating the expression of cyclin A2 may contribute to the inter-individual variation of red blood cell size observed in human populations. Previous work suggested A type cyclins to regulate entry into mitosis and faithful chromosome segregation [36, 66, 67]. Our findings illustrate a role for cyclin A2 in cytokinesis that has not been previously reported and may present a unique feature of erythroid cells, as earlier studies have also shown cell type dependent differences and requirements in cell division for cyclin A [36].

Previous work showed noncoding regulatory variants to modulate the expression of cyclin D3, the levels of which were critical to determine the number of cell divisions during terminal erythropoiesis, ultimately affecting red blood cell size and number [33]. With respect to the cyclin A2 locus, one study associated a regulatory variant within the human cyclin A2 promoter to be associated with an increased risk in diverse human cancers [68]. However, it is not clear whether the same variant may modulate the expression of cyclin A2 in erythroid cells. The associated locus encompasses a large number of variants in high linkage disequilibrium that hinders the reliable identification of the causal variant, which will require thorough testing of variants [11].

The bi-nucleated erythroid cells observed upon depletion of cyclin A2 are reminiscent of the congenital dyserythropoietic anemias (CDAs), many of which are characterized by the presence of bi- and multi-nucleated erythroblasts in the bone marrow [27, 69]. Causal genes have been implicated in DNA replication, transcription, chromatin assembly, the secretory pathway as well as cytokinesis [27], but a potentially unifying pathophysiologic mechanism has not been established and it remains speculative whether cyclin A2 may play a role in these

disorders. Mutations in cyclin A2 have not been associated with CDAs, but germline loss of function mutations of human cyclin A2 may not be viable given the lack of such reported mutations in humans (www.exac.broadinstitute.org, December 14, 2015) and early embryonic lethality in cyclin A2 knockout mice [35]. This also impeded our ability to study the function of this gene *in vivo*. Likewise, hematopoietic specific depletion of cyclin A2 resulted in stem cell loss and consequent pancytopenia [36]. Thus, inducible lineage specific deletion of cyclin A2, for example by means of utilizing mice expressing the Cre-recombinase under the control of the erythropoietin receptor, may be required to study its role *in vivo* [70]. Similarly, while both human and murine expression of cyclin A2 appear to be regulated by the erythroid transcription factor GATA1 and follow similar expression kinetics, the role of cyclin A2 has not been thoroughly studied applying *in vitro* culture models using primary human hematopoietic stem and progenitor cells. These and additional studies may prove helpful in elucidating the exact mechanism of how cyclin A2 regulates cytokinesis and whether there is a pathophysiologic relevant relationship to the CDAs.

5.2 An *ALAS2* mutation resulting in X-linked macrocytic dyserythropoietic anemia

Clinical whole-exome sequencing has been increasingly utilized for diagnostic purposes in patients with suspected rare genetic variants and disorders, especially when initial evaluation and targeted sequencing of implicated genes does not provide insight into the cause of the disease [14, 15]. This approach revealed an Y365C mutation in the X-chromosomal encoded gene *ALAS2*, significantly impairing binding of the cofactor pyridoxal 5'-phosphate (PLP) that plays a key role in stabilizing the encoded enzyme [63, 64]. Consequentially, the low affinity to PLP dramatically reduced stability of ALAS2 and likely explains the lack of improvement of the patient's anemia upon pyridoxine therapy, which has been shown to stabilize or promote the activity of mutant proteins [37, 71]. Accordingly, destabilization of the ALAS2 enzyme resulted in a severe loss of function phenotype, which may only be tolerated in female carriers, as no affected males were identified in the family. Typically, loss of function mutations in *ALAS2* are associated with microcytic sideroblastic anemia in male carriers and in females with highly skewed X-inactivation [37], and only rarely present with macrocytic anemia or dyserythropoiesis [39]. Accordingly, initial evaluation by targeted sequencing focused on genes associated with congenital dyserythropoietic anemia, although this was unrevealing [27, 69].

Highly skewed X chromosome inactivation (>90 %) was initially suspected in the affected probands [72], but only 62-82 % of active X chromosomes were shown to contain the *ALAS2* mutant. However, the most severely affected proband did show the greatest extent of skewing, suggesting that the ratio of early erythroid progenitor cells in the bone marrow expressing the wildtype to mutant allele influences clinical presentation. Still, there appears to be a clear selection against erythroid cells expressing the mutant *ALAS2* transcript over the course of erythroid differentiation, as revealed by analysis of *in vitro* cultured peripheral blood derived erythroid progenitor cells and blood reticulocyte mRNA samples from patients. These observations were consistent with the observed early block in erythropoiesis in progenitor cells in murine *Alas2* loss of function models [65]. However, despite the apparent loss of non-viable mutant expressing cells, the remaining cells expressing the wildtype *ALAS2* allele appear unable to fully maintain a normal erythroid output. It remains unclear whether the resulting pressure and stress within the wildtype *ALAS2* expressing erythroid compartment leads to dyserythropoiesis and skipped cell divisions, which may explain the observed macrocytosis in these patients, or if and whether the presence of a large number of apoptotic progenitor cells expressing the mutant transcript contributes to the observed cell-non-autonomous defects.

Closer examination of the *in vitro* differentiation process of erythroid progenitor cells may provide further insight, but unfortunately such studies were hampered by the limited availability of patient cells. Flow cytometry analysis investigating the expression of erythroid surface and apoptotic markers could potentially identify a bimodal distribution of cells showing variable differentiation kinetics and survival, reflecting expression of either the wildtype or the mutant *ALAS2* allele. Sorting of subpopulations followed by sequencing, HUMARA and related assays could help to identify the degree of X inactivation and of expression of the respective transcripts, respectively. Additionally and if possible, isolation and independent culture of progenitor cells predominantly expressing the mutant or wildtype *ALAS2* could provide additional information into how mutant-expressing cells may perturb and affect wildtype-expressing cells. Alternatively, targeted disruption of *ALAS2* expression by shRNAs in healthy donor-derived hematopoietic stem and progenitor cells and co-culture with control cells could present a possibility to mimic mutant-related effects. However, disruption of *ALAS2* expression may not reliably mimic the phenotype, especially given the variable clinical presentation of different mutations and potentially genome-editing techniques to specifically introduce the Y365C variant may need to be utilized. These and

additional approaches may be helpful to more fully comprehend the clinical manifestation as observed in the patients of this study.

5.3 Suppression of *SH2B3* results in improved proliferation and differentiation of erythroid progenitor cells

Previous studies aimed to improve human red blood cell production by optimizing culture conditions or utilizing ectopic expression of genes promoting self-renewal and proliferation and / or targeted depletion of tumor suppressor genes to expand erythroid progenitor cell populations [47, 49, 73]. In this study, we utilized insight from human genetic studies, which associated common and rare coding variants in *SH2B3* with increased red blood cell mass [11, 40, 41], to increase the efficiency of red blood cell production. Suppression of *SH2B3* expression in hematopoietic stem and progenitor as well as pluripotent embryonic cells resulted in improved erythroid maturation, enhanced proliferation and enucleation to ultimately increase erythroid output by 3-7 fold. Importantly, the produced red blood cells were of equivalent quality compared to cells derived from control cell cultures. As a negative regulator of cytokine signaling the targeted depletion of *SH2B3* augmented signaling downstream of the KIT- and erythropoietin-receptors to enhance erythroid gene expression [42]. A number of signaling factors downstream of these receptors are likely to be involved, but the increase in phospho-STAT5 levels may play an important role, as it has previously been established to be a crucial mediator of antiapoptotic signaling and survival in erythroid progenitor cells [74, 75]. Of note, although statistically not significant a small decrease in Annexin V positive cells has been observed at day 6 of differentiation and may be more pronounced at earlier stages, justifying further investigation and characterization with respect to elucidating a more detailed mechanism (Supplementary Figure S1C).

While erythroid culture protocols from pluripotent stem cells undergo continuous refinement to produce terminally mature erythroid cells [48, 76], it remains to be seen whether disruption of *SH2B3* in combination with ectopic expression and / or genetic perturbation of genes like *SOX2*, *MYC*, *BCL-XL*, and *TP53*, all of which have been associated with enhanced red blood cell production *in vitro*, may further potentiate the overall yield [47, 49]. The use and risk of such genetic alterations in progenitor cells appears feasible, as red blood cells enucleate and thus lose their genetic material. Genetic manipulation may otherwise contribute to leukemogenesis, be it mediated directly by the factors utilized or by

insertional mutagenesis [77, 78]. Nevertheless, safety and efficacy of such cell replacement approaches remain to be established thoroughly [79]. The effects of repeated transfusions of *in vitro* generated red blood cells remain unknown and alterations in the expression of blood group antigens, especially in the setting of genetic perturbations, may lead to sensitization after multiple administrations and thus present a significant concern [50]. Furthermore, genetic approaches will have to be complemented with refinements of culture conditions and tailored solutions for large-scale production in bioreactors to maximize output and cost efficiency to create a feasible alternative and overcome shortages in the blood supply [46, 48]. In addition, *in vitro* generated and engineered red blood cells in large quantities may have broader uses as vehicles for drugs and other therapeutic substances [50, 77, 80].

5.4 Conclusions

Numerous genetic studies have revealed a myriad of genetic variants associated with human traits and disease. In this thesis, we followed up on clues from various genetic studies to describe a role for cyclin A2 in cytokinesis during terminal erythropoiesis, resulting in the production of erythrocytes with variable size. Similarly, variants in the locus modulating the expression of cyclin A2 may contribute to the variation in red blood cells traits as observed in human populations. Additionally, the use of whole exome-sequencing identified a heterozygous loss of function mutation in *ALAS2* in a family presenting with macrocytic anemia and dyserythropoiesis. The mutation resulted in impaired binding of the cofactor pyridoxine, destabilization and loss of function of the ALAS2 enzyme and selected against erythroid cells expressing the mutant. Finally, the targeted disruption of the expression of *SH2B3*, encoding a negative regulator of cytokine signaling, in hematopoietic and pluripotent stem cells provides an avenue to increase erythroid output *in vitro* for potential cellular replacement therapies. Together, these studies provide examples of how functional investigation to elucidate the underlying molecular basis of genetic associations can provide new insight and a deeper understanding of human physiology and disease.

6. Bibliography

1. Lander ES. Initial impact of the sequencing of the human genome. Nature, 2011. 470(7333): p. 187-97.
2. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, et al. Initial sequencing and analysis of the human genome. Nature, 2001. 409(6822): p. 860-921.
3. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, et al. The sequence of the human genome. Science, 2001. 291(5507): p. 1304-51.
4. Bernstein BE, Meissner A, and Lander ES. The mammalian epigenome. Cell, 2007. 128(4): p. 669-81.
5. Roadmap Epigenomics C, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, et al. Integrative analysis of 111 reference human epigenomes. Nature, 2015. 518(7539): p. 317-30.
6. Consortium EP. An integrated encyclopedia of DNA elements in the human genome. Nature, 2012. 489(7414): p. 57-74.
7. Genomes Project C, Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, Hurles ME, and McVean GA. A map of human genome variation from population-scale sequencing. Nature, 2010. 467(7319): p. 1061-73.
8. Genomes Project C, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, and McVean GA. An integrated map of genetic variation from 1,092 human genomes. Nature, 2012. 491(7422): p. 56-65.
9. International HapMap C, Altshuler DM, Gibbs RA, Peltonen L, Altshuler DM, Gibbs RA, Peltonen L, Dermitzakis E, Schaffner SF, Yu F, et al. Integrating common and rare genetic variation in diverse human populations. Nature, 2010. 467(7311): p. 52-8.
10. Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, et al. A global reference for human genetic variation. Nature, 2015. 526(7571): p. 68-74.
11. van der Harst P, Zhang W, Mateo Leach I, Rendon A, Verweij N, Sehmi J, Paul DS, Elling U, Allayee H, Li X, et al. Seventy-five genetic loci influencing the human red blood cell. Nature, 2012. 492(7429): p. 369-75.
12. Uda M, Galanello R, Sanna S, Lettre G, Sankaran VG, Chen W, Usala G, Busonero F, Maschio A, Albai G, et al. Genome-wide association study shows BCL11A associated with persistent fetal hemoglobin and amelioration of the phenotype of beta-thalassemia. Proc Natl Acad Sci U S A, 2008. 105(5): p. 1620-5.
13. Consortium GT. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science, 2015. 348(6235): p. 648-60.
14. Lewis R. Exome sequencing comes to the clinic. JAMA, 2015. 313(13): p. 1301-3.
15. Yang Y, Muzny DM, Reid JG, Bainbridge MN, Willis A, Ward PA, Braxton A, Beuten J, Xia F, Niu Z, et al. Clinical whole-exome sequencing for the diagnosis of mendelian disorders. N Engl J Med, 2013. 369(16): p. 1502-11.
16. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, et al. Finding the missing heritability of complex diseases. Nature, 2009. 461(7265): p. 747-53.
17. Couzin-Frankel J. Major heart disease genes prove elusive. Science, 2010. 328(5983): p. 1220-1.
18. Fugger L, McVean G, and Bell JI. Genomewide association studies and common disease--realizing clinical utility. N Engl J Med, 2012. 367(25): p. 2370-1.
19. Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-Andre V, Sigova AA, Hoke HA, and Young RA. Super-enhancers in the control of cell identity and disease. Cell, 2013. 155(4): p. 934-47.
20. Claussnitzer M, Dankel SN, Kim KH, Quon G, Meuleman W, Haugen C, Glunk V, Sousa IS, Beaudry JL, Puviindran V, et al. FTO Obesity Variant Circuitry and Adipocyte Browning in Humans. N Engl J Med, 2015. 373(10): p. 895-907.
21. Kathiresan S and Srivastava D. Genetics of human cardiovascular disease. Cell, 2012. 148(6): p. 1242-57.
22. Sankaran VG, Menne TF, Xu J, Akie TE, Lettre G, Van Handel B, Mikkola HK, Hirschhorn JN, Cantor AB, and Orkin SH. Human fetal hemoglobin expression is regulated by the developmental stage-specific repressor BCL11A. Science, 2008. 322(5909): p. 1839-42.
23. Canver MC, Smith EC, Sher F, Pinello L, Sanjana NE, Shalem O, Chen DD, Schupp PG, Vinjamur DS, Garcia SP, et al. BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. Nature, 2015. 527(7577): p. 192-7.

24. Lenz TL, Deutsch AJ, Han B, Hu X, Okada Y, Eyre S, Knapp M, Zhernakova A, Huizinga TW, Abecasis G, et al. Widespread non-additive and interaction effects within HLA loci modulate the risk of autoimmune diseases. Nat Genet, 2015. 47(9): p. 1085-90.
25. Sankaran VG and Weiss MJ. Anemia: progress in molecular mechanisms and therapies. Nat Med, 2015. 21(3): p. 221-30.
26. Hattangadi SM, Wong P, Zhang L, Flygare J, and Lodish HF. From stem cell to red cell: regulation of erythropoiesis at multiple levels by multiple proteins, RNAs, and chromatin modifications. Blood, 2011. 118(24): p. 6258-68.
27. Iolascon A, Heimpel H, Wahlin A, and Tamary H. Congenital dyserythropoietic anemias: molecular insights and diagnostic approach. Blood, 2013. 122(13): p. 2162-6.
28. Hildick-Smith GJ, Cooney JD, Garone C, Kremer LS, Haack TB, Thon JN, Miyata N, Lieber DS, Calvo SE, Akman HO, et al. Macrocytic anemia and mitochondriopathy resulting from a defect in sideroflexin 4. Am J Hum Genet, 2013. 93(5): p. 906-14.
29. Gripp KW, Curry C, Olney AH, Sandoval C, Fisher J, Chong JX, Genomics UWCfM, Pilchman L, Sahraoui R, Stabley DL, et al. Diamond-Blackfan anemia with mandibulofacial dystostosis is heterogeneous, including the novel DBA genes TSR2 and RPS28. Am J Med Genet A, 2014. 164A(9): p. 2240-9.
30. Sankaran VG, Ghazvinian R, Do R, Thiru P, Vergilio JA, Beggs AH, Sieff CA, Orkin SH, Nathan DG, Lander ES, et al. Exome sequencing identifies GATA1 mutations resulting in Diamond-Blackfan anemia. J Clin Invest, 2012. 122(7): p. 2439-43.
31. Ludwig LS, Gazda HT, Eng JC, Eichhorn SW, Thiru P, Ghazvinian R, George TI, Gotlib JR, Beggs AH, Sieff CA, et al. Altered translation of GATA1 in Diamond-Blackfan anemia. Nat Med, 2014. 20(7): p. 748-53.
32. Wang R, Yoshida K, Toki T, Sawada T, Uechi T, Okuno Y, Sato-Otsubo A, Kudo K, Kamimaki I, Kanezaki R, et al. Loss of function mutations in RPL27 and RPS27 identified by whole-exome sequencing in Diamond-Blackfan anaemia. Br J Haematol, 2015. 168(6): p. 854-64.
33. Sankaran VG, Ludwig LS, Sicinska E, Xu J, Bauer DE, Eng JC, Patterson HC, Metcalf RA, Natkunam Y, Orkin SH, et al. Cyclin D3 coordinates the cell cycle during differentiation to regulate erythrocyte size and number. Genes Dev, 2012. 26(18): p. 2075-87.
34. Thom CS, Traxler EA, Khandros E, Nickas JM, Zhou OY, Lazarus JE, Silva AP, Prabhu D, Yao Y, Aribeana C, et al. Trim58 degrades Dynein and regulates terminal erythropoiesis. Dev Cell, 2014. 30(6): p. 688-700.
35. Murphy M, Stinnakre MG, Senamaud-Beaufort C, Winston NJ, Sweeney C, Kubelka M, Carrington M, Brechot C, and Sobczak-Thepot J. Delayed early embryonic lethality following disruption of the murine cyclin A2 gene. Nat Genet, 1997. 15(1): p. 83-6.
36. Kalaszczynska I, Geng Y, Iino T, Mizuno S, Choi Y, Kondratiuk I, Silver DP, Wolgemuth DJ, Akashi K, and Sicinski P. Cyclin A is redundant in fibroblasts but essential in hematopoietic and embryonic stem cells. Cell, 2009. 138(2): p. 352-65.
37. Fleming MD. Congenital sideroblastic anemias: iron and heme lost in mitochondrial translation. Hematology Am Soc Hematol Educ Program, 2011. 2011: p. 525-31.
38. Campagna DR, de Bie CI, Schmitz-Abe K, Sweeney M, Sendamarai AK, Schmidt PJ, Heeney MM, Yntema HG, Kannengiesser C, Grandchamp B, et al. X-linked sideroblastic anemia due to ALAS2 intron 1 enhancer element GATA-binding site mutations. Am J Hematol, 2014. 89(3): p. 315-9.
39. Aivado M, Gattermann N, Rong A, Giagounidis AA, Prall WC, Czibere A, Hildebrandt B, Haas R, and Bottomley SS. X-linked sideroblastic anemia associated with a novel ALAS2 mutation and unfortunate skewed X-chromosome inactivation patterns. Blood Cells Mol Dis, 2006. 37(1): p. 40-5.
40. Lasho TL, Pardanani A, and Tefferi A. LNK mutations in JAK2 mutation-negative erythrocytosis. N Engl J Med, 2010. 363(12): p. 1189-90.
41. Spolverini A, Pieri L, Guglielmelli P, Pancrazzi A, Fanelli T, Paoli C, Bosi A, Nichele I, Ruggeri M, and Vannucchi AM. Infrequent occurrence of mutations in the PH domain of LNK in patients with JAK2 mutation-negative 'idiopathic' erythrocytosis. Haematologica, 2013. 98(9): p. e101-2.
42. Bersenev A, Wu C, Balcerek J, and Tong W. Lnk controls mouse hematopoietic stem cell self-renewal and quiescence through direct interactions with JAK2. J Clin Invest, 2008. 118(8): p. 2832-44.
43. Oh ST, Simonds EF, Jones C, Hale MB, Goltsev Y, Gibbs KD, Jr., Merker JD, Zehnder JL, Nolan GP, and Gotlib J. Novel mutations in the inhibitory adaptor protein LNK drive JAK-STAT signaling in patients with myeloproliferative neoplasms. Blood, 2010. 116(6): p. 988-92.
44. Velazquez L, Cheng AM, Fleming HE, Furlonger C, Vesely S, Bernstein A, Paige CJ, and Pawson T. Cytokine signaling and hematopoietic homeostasis are disrupted in Lnk-deficient mice. J Exp Med, 2002. 195(12): p. 1599-611.

45. Gery S and Koeffler HP. Role of the adaptor protein LNK in normal and malignant hematopoiesis. Oncogene, 2013. 32(26): p. 3111-8.
46. Williamson LM and Devine DV. Challenges in the management of the blood supply. Lancet, 2013. 381(9880): p. 1866-75.
47. Huang X, Shah S, Wang J, Ye Z, Dowey SN, Tsang KM, Mendelsohn LG, Kato GJ, Kickler TS, and Cheng L. Extensive ex vivo expansion of functional human erythroid precursors established from umbilical cord blood cells by defined factors. Mol Ther, 2014. 22(2): p. 451-63.
48. Rousseau GF, Giarratana MC, and Douay L. Large-scale production of red blood cells from stem cells: what are the technical challenges ahead? Biotechnol J, 2014. 9(1): p. 28-38.
49. Hirose S, Takayama N, Nakamura S, Nagasawa K, Ochi K, Hirata S, Yamazaki S, Yamaguchi T, Otsu M, Sano S, et al. Immortalization of erythroblasts by c-MYC and BCL-XL enables large-scale erythrocyte production from human pluripotent stem cells. Stem Cell Reports, 2013. 1(6): p. 499-508.
50. Migliaccio AR, Whitsett C, Papayannopoulou T, and Sadelain M. The potential of stem cells as an in vitro source of red blood cells for transfusion. Cell Stem Cell, 2012. 10(2): p. 115-9.
51. Flygare J, Rayon Estrada V, Shin C, Gupta S, and Lodish HF. HIF1alpha synergizes with glucocorticoids to promote BFU-E progenitor self-renewal. Blood, 2011. 117(12): p. 3435-44.
52. Lee HY, Gao X, Barrasa MI, Li H, Elmes RR, Peters LL, and Lodish HF. PPAR-alpha and glucocorticoid receptor synergize to promote erythroid progenitor self-renewal. Nature, 2015.
53. Sankaran VG, Menne TF, Scepanovic D, Vergilio JA, Ji P, Kim J, Thiru P, Orkin SH, Lander ES, and Lodish HF. MicroRNA-15a and -16-1 act via MYB to elevate fetal hemoglobin expression in human trisomy 13. Proc Natl Acad Sci U S A, 2011. 108(4): p. 1519-24.
54. Ji P, Jayapal SR, and Lodish HF. Enucleation of cultured mouse fetal erythroblasts requires Rac GTPases and mDia2. Nat Cell Biol, 2008. 10(3): p. 314-21.
55. An X, Schulz VP, Li J, Wu K, Liu J, Xue F, Hu J, Mohandas N, and Gallagher PG. Global transcriptome analyses of human and murine terminal erythroid differentiation. Blood, 2014. 123(22): p. 3466-77.
56. Novershtern N, Subramanian A, Lawton LN, Mak RH, Haining WN, McConkey ME, Habib N, Yosef N, Chang CY, Shay T, et al. Densely interconnected transcriptional circuits control cell states in human hematopoiesis. Cell, 2011. 144(2): p. 296-309.
57. Wu W, Cheng Y, Keller CA, Ernst J, Kumar SA, Mishra T, Morrissey C, Dorman CM, Chen KB, Drautz D, et al. Dynamics of the epigenetic landscape during erythroid differentiation after GATA1 restoration. Genome Res, 2011. 21(10): p. 1659-71.
58. Wu C, Orozco C, Boyer J, Leglise M, Goodale J, Batalov S, Hodge CL, Haase J, Janes J, Huss JW, 3rd, et al. BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. Genome Biol, 2009. 10(11): p. R130.
59. Ulirsch JC, Lacy JN, An X, Mohandas N, Mikkelsen TS, and Sankaran VG. Altered chromatin occupancy of master regulators underlies evolutionary divergence in the transcriptional landscape of erythroid differentiation. PLoS Genet, 2014. 10(12): p. e1004890.
60. Kellis M, Wold B, Snyder MP, Bernstein BE, Kundaje A, Marinov GK, Ward LD, Birney E, Crawford GE, Dekker J, et al. Defining functional DNA elements in the human genome. Proc Natl Acad Sci U S A, 2014. 111(17): p. 6131-8.
61. Weiss MJ, Yu C, and Orkin SH. Erythroid-cell-specific properties of transcription factor GATA-1 revealed by phenotypic rescue of a gene-targeted cell line. Mol Cell Biol, 1997. 17(3): p. 1642-51.
62. Kadauke S and Blobel GA. Mitotic bookmarking by transcription factors. Epigenetics Chromatin, 2013. 6(1): p. 6.
63. Bishop DF, Tchaikovskii V, Hoffbrand AV, Fraser ME, and Margolis S. X-linked sideroblastic anemia due to carboxyl-terminal ALAS2 mutations that cause loss of binding to the beta-subunit of succinyl-CoA synthetase (SUCLA2). J Biol Chem, 2012. 287(34): p. 28943-55.
64. Cotter PD, Rucknagel DL, and Bishop DF. X-linked sideroblastic anemia: identification of the mutation in the erythroid-specific delta-aminolevulinate synthase gene (ALAS2) in the original family described by Cooley. Blood, 1994. 84(11): p. 3915-24.
65. Harigae H, Nakajima O, Suwabe N, Yokoyama H, Furuyama K, Sasaki T, Kaku M, Yamamoto M, and Sassa S. Aberrant iron accumulation and oxidized status of erythroid-specific delta-aminolevulinate synthase (ALAS2)-deficient definitive erythroblasts. Blood, 2003. 101(3): p. 1188-93.
66. Furuno N, den Elzen N, and Pines J. Human cyclin A is required for mitosis until mid prophase. J Cell Biol, 1999. 147(2): p. 295-306.
67. Kabeche L and Compton DA. Cyclin A regulates kinetochore microtubules to promote faithful chromosome segregation. Nature, 2013. 502(7469): p. 110-3.

68. Kim DH, Park SE, Kim M, Ji YI, Kang MY, Jung EH, Ko E, Kim Y, Kim S, Shim YM, et al. A functional single nucleotide polymorphism at the promoter region of cyclin A2 is associated with increased risk of colon, liver, and lung cancers. Cancer, 2011. 117(17): p. 4080-91.
69. Schwarz K, Iolascon A, Verissimo F, Trede NS, Horsley W, Chen W, Paw BH, Hopfner KP, Holzmann K, Russo R, et al. Mutations affecting the secretory COPII coat component SEC23B cause congenital dyserythropoietic anemia type II. Nat Genet, 2009. 41(8): p. 936-40.
70. Heinrich AC, Pelanda R, and Klingmuller U. A mouse model for visualization and conditional mutations in the erythroid lineage. Blood, 2004. 104(3): p. 659-66.
71. Astner I, Schulze JO, van den Heuvel J, Jahn D, Schubert WD, and Heinz DW. Crystal structure of 5-aminolevulinate synthase, the first enzyme of heme biosynthesis, and its link to XLSA in humans. EMBO J, 2005. 24(18): p. 3166-77.
72. Minks J, Robinson WP, and Brown CJ. A skewed view of X chromosome inactivation. J Clin Invest, 2008. 118(1): p. 20-3.
73. Miharada K, Hiroyama T, Sudo K, Nagasawa T, and Nakamura Y. Efficient enucleation of erythroblasts differentiated in vitro from hematopoietic stem and progenitor cells. Nat Biotechnol, 2006. 24(10): p. 1255-6.
74. Socolovsky M, Nam H, Fleming MD, Haase VH, Brugnara C, and Lodish HF. Ineffective erythropoiesis in Stat5a(-/-)5b(-/-) mice due to decreased survival of early erythroblasts. Blood, 2001. 98(12): p. 3261-73.
75. Socolovsky M, Fallon AE, Wang S, Brugnara C, and Lodish HF. Fetal anemia and apoptosis of red cell progenitors in Stat5a-/-5b-/- mice: a direct role for Stat5 in Bcl-X(L) induction. Cell, 1999. 98(2): p. 181-91.
76. Kobari L, Yates F, Oudrhiri N, Francina A, Kiger L, Mazurier C, Rouzbeh S, El-Nemer W, Hebert N, Giarratana MC, et al. Human induced pluripotent stem cells can reach complete terminal maturation: in vivo and in vitro evidence in the erythropoietic differentiation model. Haematologica, 2012. 97(12): p. 1795-803.
77. Shi J, Kundrat L, Pishesha N, Bilate A, Theile C, Maruyama T, Dougan SK, Ploegh HL, and Lodish HF. Engineered red blood cells as carriers for systemic delivery of a wide array of functional probes. Proc Natl Acad Sci U S A, 2014. 111(28): p. 10131-6.
78. Shaw A and Cornetta K. Design and potential of non-integrating lentiviral vectors. Biomedicines, 2014(2): p. 14-35.
79. Giarratana MC, Rouard H, Dumont A, Kiger L, Safeukui I, Le Pennec PY, Francois S, Trugnan G, Peyrard T, Marie T, et al. Proof of principle for transfusion of in vitro-generated red blood cells. Blood, 2011. 118(19): p. 5071-9.
80. Chang AH, Stephan MT, and Sadelain M. Stem cell-derived erythroid cells mediate long-term systemic protein delivery. Nat Biotechnol, 2006. 24(8): p. 1017-21.

7. Eidesstattliche Versicherung

„Ich, Leif Si-Hun Ludwig, versichere an Eides statt durch meine eigenhändige Unterschrift, dass ich die vorgelegte Dissertation mit dem Thema: „Functional studies of genetic variants in human erythropoiesis“ selbstständig und ohne nicht offengelegte Hilfe Dritter verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel genutzt habe.

Alle Stellen, die wörtlich oder dem Sinne nach auf Publikationen oder Vorträgen anderer Autoren beruhen, sind als solche in korrekter Zitierung (siehe „Uniform Requirements for Manuscripts (URM)“ des ICMJE -*www.**icmje**.org*) kenntlich gemacht. Die Abschnitte zu Methodik (insbesondere praktische Arbeiten, Laborbestimmungen, statistische Aufarbeitung) und Resultaten (insbesondere Abbildungen, Graphiken und Tabellen) entsprechen den URM (s.o.) und werden von mir verantwortet.

Meine Anteile an den ausgewählten Publikationen entsprechen denen, die in der untenstehenden gemeinsamen Erklärung mit der Betreuerin, angegeben sind. Sämtliche Publikationen, die aus dieser Dissertation hervorgegangen sind und bei denen ich Autor bin, entsprechen den URM (s.o.) und werden von mir verantwortet.

Die Bedeutung dieser eidesstattlichen Versicherung und die strafrechtlichen Folgen einer unwahren eidesstattlichen Versicherung (§156,161 des Strafgesetzbuches) sind mir bekannt und bewusst.“

Berlin, den 16. Dezember 2015

Unterschrift

8. Anteilserklärung / Declaration of contribution

Leif Si-Hun Ludwig hatte folgenden Anteil an den erfolgten Publikationen / contributed as follows to the individual publications:

Publication 1:

Ludwig LS*, Cho H*, Wakabayashi A, Eng JC, Ulirsch JC, Fleming MD, Lodish HF and Sankaran VG. *These authors contributed equally.

Genome-wide association study follow-up identifies cyclin A2 as a regulator of the transition through cytokinesis during terminal erythropoiesis.

American Journal of Hematology, 2015, May;90(5):386-91.

Beitrag im Einzelnen / Contribution

Leif Si-Hun Ludwig designed the study with Hyunjii Cho and Vijay Sankaran. He designed, performed and analyzed most experiments in the study together with Hyunjii Cho, including isolation of erythroid progenitor cells from murine fetal liver, retroviral infections, *in vitro* culture of erythroid cells, flow cytometric assays, qPCR, western blot experiments and coordinated bioinformatic analysis with Jacob Ulirsch. He wrote the manuscript with Hyunjii Cho, Harvey Lodish and Vijay Sankaran with input from all authors.

Publication 2:

Sankaran VG, Ulirsch JC, Tchaikovskii V, **Ludwig LS**, Wakabayashi A, Kadirvel S, Lindsley RC, Bejar R, Shi J, Lovitch SB, Bishop DF and Steensma DP.

X-linked macrocytic dyserythropoietic anemia in females with an ALAS2 mutation.

Journal of Clinical Investigation, 2015 Apr;125(4):1665-9.

Beitrag im Einzelnen / Contribution

Leif Si-Hun Ludwig isolated peripheral blood mononuclear cells, *in vitro* cultured erythroid progenitor cells and conducted sanger sequencing of genomic DNA (+/- HpaII digestion) and cDNA after reverse transcription of mRNA derived from peripheral blood mononuclear cells, reticulocytes and cultured erythroid cells with assistance from Aoi Wakabayashi. He reviewed and edited the manuscript.

Publication 3:

Giani FG, Fiorini C, Wakabayashi A, **Ludwig LS**, Jobaliya CD, Regan SN, Ulirsch JC, Liang G, Steinberg-Shemer O, Esko T, Hirschhorn JN, Tong W, Brugnara C, Weiss MJ, Zon LI, Chou ST, French DL, Musunuru K and Sankaran VG.

Targeted application of human genetic variation can improve red blood cell production from stem cells.

Cell Stem Cell, 2016, Jan 7;18(1):1-6.

Beitrag im Einzelnen / Contribution

Leif Si-Hun Ludwig infected and cultured primary human stem and progenitor cells and conducted flow cytometry analysis with a focus on expression of surface markers and blood antigens. He assisted in the design and completion of pyruvate kinase assays, PKH labeling and western blot experiments. He reviewed and edited the manuscript.

Unterschrift, Datum und Stempel der betreuenden Hochschullehrerin

_____

Unterschrift des Doktoranden

_____

9. Publications

The original articles, including supplementary material are included on the following pages or are available online:

Ludwig LS*, Cho H*, Wakabayashi A, Eng JC, Ulirsch JC, Fleming MD, Lodish HF and Sankaran VG. *These authors contributed equally.
Genome-wide association study follow-up identifies cyclin A2 as a regulator of the transition through cytokinesis during terminal erythropoiesis.
American Journal of Hematology, 2015, May;90(5):386-91.
http://dx.doi.org/10.1002/ajh.23952
http://onlinelibrary.wiley.com/doi/10.1002/ajh.23952/abstract

Sankaran VG, Ulirsch JC, Tchaikovskii V, **Ludwig LS**, Wakabayashi A, Kadirvel S, Lindsley RC, Bejar R, Shi J, Lovitch SB, Bishop DF and Steensma DP.
X-linked macrocytic dyserythropoietic anemia in females with an *ALAS2* mutation.
Journal of Clinical Investigation, 2015 Apr;125(4):1665-9.
http://dx.doi.org/10.1172/JCI78619
http://www.jci.org/articles/view/78619

Giani FG, Fiorini C, Wakabayashi A, **Ludwig LS**, Jobaliya CD, Regan SN, Ulirsch JC, Liang G, Steinberg-Shemer O, Esko T, Hirschhorn JN, Tong W, Brugnara C, Weiss MJ, Zon LI, Chou ST, French DL, Musunuru K and Sankaran VG.
Targeted application of human genetic variation can improve red blood cell production from stem cells.
Cell Stem Cell, 2016, Jan 7;18(1):1-6.
http://dx.doi.org/10.1016/j.stem.2015.09.015
http://www.cell.com/cell-stem-cell/abstract/S1934-5909%2815%2900422-1

Genome-wide association study follow-up identifies cyclin A2 as a regulator of the transition through cytokinesis during terminal erythropoiesis.

Ludwig LS*, Cho H*, Wakabayashi A, Eng JC, Ulirsch JC, Fleming MD, Lodish HF and Sankaran VG. *These authors contributed equally.

American Journal of Hematology, 2015, May;90(5):386-91.

The original article is available online:
http://dx.doi.org/10.1002/ajh.23952
http://onlinelibrary.wiley.com/doi/10.1002/ajh.23952/abstract

X-linked macrocytic dyserythropoietic anemia in females with an *ALAS2* mutation

Vijay G. Sankaran,[1,2] Jacob C. Ulirsch,[1,2] Vassili Tchaikovskii,[3] Leif S. Ludwig,[1,2,4,5] Aoi Wakabayashi,[1,2] Senkottuvelan Kadirvel,[3] R. Coleman Lindsley,[6] Rafael Bejar,[6] Jiahai Shi,[7] Scott B. Lovitch,[8] David F. Bishop,[3] and David P. Steensma[6]

[1]Division of Hematology/Oncology, Boston Children's Hospital and Department of Pediatric Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts, USA. [2]Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA. [3]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, Mount Sinai Medical Center, New York, New York, USA. [4]Institute for Chemistry and Biochemistry, Freie Universität, Berlin, Germany. [5]Charité-Universitätsmedizin, Berlin, Germany. [6]Division of Hematologic Malignancies, Dana-Farber Cancer Institute and Harvard Medical School, Boston, Massachusetts, USA. [7]Whitehead Institute for Biomedical Research, Cambridge, Massachusetts, USA. [8]Department of Pathology, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA.

Macrocytic anemia with abnormal erythropoiesis is a common feature of megaloblastic anemias, congenital dyserythropoietic anemias, and myelodysplastic syndromes. Here, we characterized a family with multiple female individuals who have macrocytic anemia. The proband was noted to have dyserythropoiesis and iron overload. After an extensive diagnostic evaluation that did not provide insight into the cause of the disease, whole-exome sequencing of multiple family members revealed the presence of a mutation in the X chromosomal gene *ALAS2*, which encodes 5′-aminolevulinate synthase 2, in the affected females. We determined that this mutation (Y365C) impairs binding of the essential cofactor pyridoxal 5′-phosphate to ALAS2, resulting in destabilization of the enzyme and consequent loss of function. X inactivation was not highly skewed in wbc from the affected individuals. In contrast, and consistent with the severity of the *ALAS2* mutation, there was a complete skewing toward expression of the WT allele in mRNA from reticulocytes that could be recapitulated in primary erythroid cultures. Together, the results of the X inactivation and mRNA studies illustrate how this X-linked dominant mutation in *ALAS2* can perturb normal erythropoiesis through cell-nonautonomous effects. Moreover, our findings highlight the value of whole-exome sequencing in diagnostically challenging cases for the identification of disease etiology and extension of the known phenotypic spectrum of disease.

Introduction

Macrocytic anemia and abnormal dysplastic erythropoiesis (herein referred to as dyserythropoiesis) are features found in a number of congenital and acquired conditions, including congenital dyserythropoietic anemias (CDAs), specific mitochondrial disorders, myelodysplastic syndromes (MDS), and megaloblastic anemias (1, 2). While numerous etiologies are known for this class of disorders, some cases lack a clear explanation (3). An improved understanding of the etiology of these disorders will provide additional insight into the mechanisms of abnormal blood cell production in humans (4).

We hypothesized that certain cases of this type of anemia either represent new genetic conditions or, alternatively, are due to the phenotypic variation in genes previously associated with anemia. We identified a female proband with a macrocytic dyserythropoietic anemia since childhood and evidence of iron overload, who had undergone an extensive, but unrevealing, clinical evaluation. The proband's sister and mother were also noted to have macrocytic anemia, supporting a potential genetic etiology. We therefore performed whole-exome sequencing and functional studies to identify and understand the basis of the anemia observed in this family.

Conflict of interest: The authors have delcared that no conflict of interest exists.
Submitted: August 26, 2014; **Accepted:** January 8, 2015.
Reference information: *J Clin Invest*. doi:10.1172/JCI78619.

Results and Discussion

The proband (family member II-2) (Figure 1E) was a 32-year-old woman with a history of atrial septal defect who had been noted to have an anemia in childhood of unclear etiology. She had come to clinical attention 7 years earlier, when a routine blood count during her first pregnancy revealed an anemia with a hematocrit of 24% to 26% and a mean corpuscular volume of 114 to 122 femtoliters (Supplemental Table 1; supplemental material available online with this article; doi:10.1172/JCI78619DS1) that was unresponsive to iron and vitamin B12 supplementation. Delivery occurred at full term and was complicated by placenta accreta with concomitant blood loss, and the proband required an rbc transfusion for the first time. Three years later, the proband delivered healthy twins and experienced a massive hemorrhage 2 weeks after the delivery, resulting in a drop of her hematocrit to 17% that required an additional transfusion.

At that time, it was noted that the proband had a macrocytic anemia and an elevated serum ferritin level (800 ng/ml), prompting referral to a hematologist. A BM biopsy revealed trilineage hematopoiesis with dyserythropoiesis (Figure 1, A–D). The proband had a normal BM karyotype and MDS/acute myeloid leukemia FISH panel (evaluating chromosomes 5, 7, 8, 13, 20, and X). According to the report, Prussian blue reaction on the initial BM aspirate showed no excess iron or sideroblasts. A liver biopsy revealed an iron index of 11 (normal, <1), and *HFE* genetic testing revealed no mutations. Therapeutic phlebotomy was attempted, but was limited by symptomatic anemia.

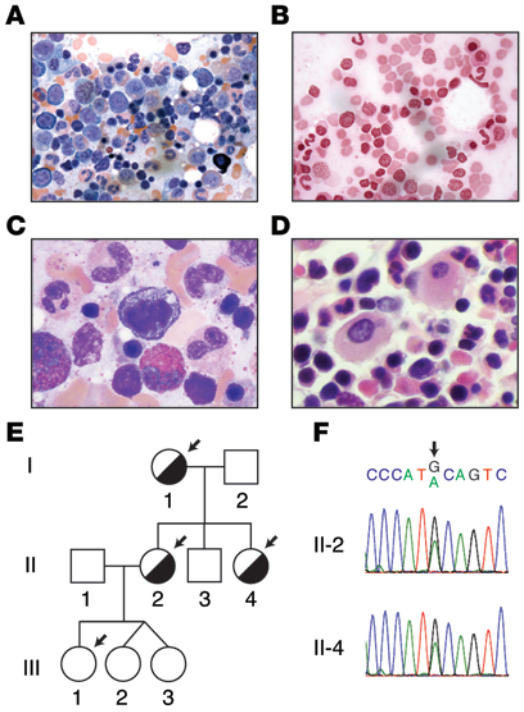


Figure 1. Identification of an *ALAS2* mutation in a family with macrocytic anemia and dyserythropoiesis. (**A**) BM aspirate from the proband (II-2) with erythroid hyperplasia. (**B**) Prussian blue staining revealed rare erythroblasts with siderotic granules. (**C**) BM aspirate with dyserythropoiesis. (**D**) Core biopsy demonstrating erythroid hyperplasia and dyserythropoiesis. (**E**) Pedigree of the family, with affected individuals highlighted by half-filled circles. Arrows highlight those individuals who underwent whole-exome sequencing. (**F**) Sanger sequencing traces around X chromosome position 55042086 (hg19 coordinates) in 2 affected individuals. Original magnification, ×100 (**A** and **B**) and ×2000 (**C** and **D**).

Laboratory testing of the proband, her sister (family member II-4), and mother (family member I-2) demonstrated rbc macrocytosis (Figure 1E and Supplemental Table 1). The brother and daughters of the proband had normal complete blood counts (Figure 1E). T2* MRI was performed on the proband and revealed an estimated liver iron concentration of greater than 350 μmol/g (normal, <35 μmol/g). In addition, the proband had a horseshoe kidney and a mildly enlarged spleen, measuring 14.8 cm. A trial with the iron chelator deferasirox was attempted but was not tolerated, and therefore another trial of phlebotomy with a target hemoglobin level of 10 g/dl was initiated. Since the clinical features of this patient suggested the possibility of CDA, targeted sequencing of *CDAN1*, *SEC23B*, and *KLF1* was performed, but no mutations were identified (5). The proband and family members then consented to participate in a whole-exome sequencing study (6, 7).

We used an automated pipeline to identify novel coding and the loss-of-function (LOF) variants found in the 3 affected individuals from this family (I-1, II-2, and II-4), but not in the unaffected daughter (III-1) (6). Each individual had a total of approximately 200 novel coding and LOF variants (Supplemental Table 2). A single notable variant fit the model of complete penetrance and was found to be a coding mutation in the X chromosomal gene *ALAS2*.

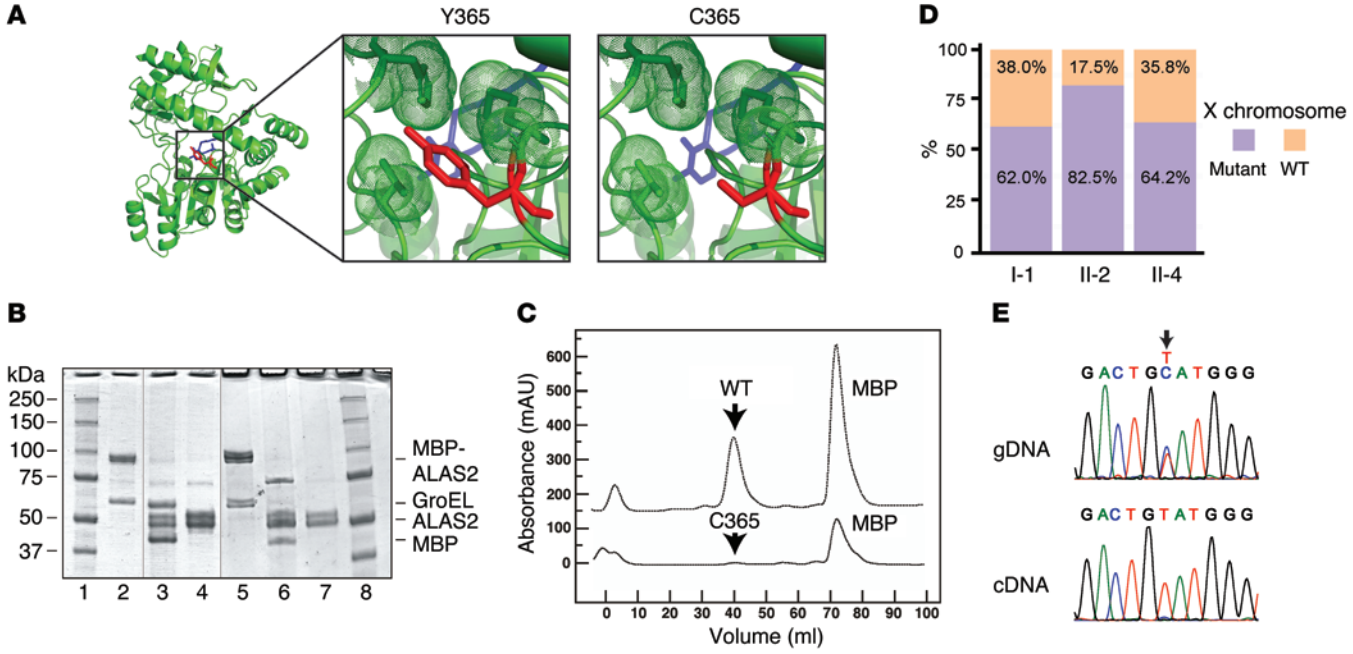


Figure 2. Severe LOF with the ALAS2 Y365C mutation and lack of highly skewed X inactivation in female mutation carriers. (**A**) Model of ALAS2 shows PLP highlighted in blue and the Y or C amino acid at position 365 highlighted in red. (**B**) SDS-PAGE gel of WT and mutant ALAS2. Lanes 1 and 8 contain the protein standards, while lanes 2–4 and 5–7 contain WT and mutant ALAS2 protein samples, respectively. Lanes 2 and 5 show partially purified samples after amylose affinity chromatography, lanes 3 and 6 show results after factor Xa digestion, and lanes 4 and 7 show results after gel filtration chromatography. Thin vertical lines in this composite figure separate noncontiguous lanes in the 2 original gels. (**C**) Chromatographic profiles for purification of WT and mutant ALAS2 by size exclusion (absorbance at 280 nm is shown in milliabsorbance units [mAU]). (**D**) Quantification of HUMARA results in all affected individuals showing WT and mutant X chromosomes. (**E**) Sanger sequencing traces of genomic DNA (gDNA) and cDNA derived from reticulocyte mRNA from the proband (II-2) for *ALAS2*, with an arrow highlighting the mutation.

Table 1. Properties of WT and mutant ALAS2

Property	WT (Y365)	C365 mutant
Yield (%)	17.7 ± 1.3 (n = 4)	5.5 ± 1.3 (n = 2)
Specific activity (U/mg)	90,700	62,100
V_{max} SCoA (U/mg)	96,200	78,800
V_{max} Gly (U/mg)	85,000	73,700
$Hill_n$ SCoA	1.4	1.3
K_m SCoA (μM)	60	33
K_m Gly (mM)	8.0	8.4
K_m PLP (nM)	11	737
$t_{1/2}$ in 10 uM PLP (min)	7.57 ± 0.96	1.24 ± 0.15
$t_{1/2}$ in 100 uM PLP (min)	15.6 ± 1.3	1.46 ± 0.15

Specific activities were for a single purification experiment, the kinetic values of V_{max}, $Hill_n$, and K_m were derived from Lineweaver-Burk plots of single kinetic experiments, and the half-life data at 50°C were the average and SD of 3 independent experiments for each enzyme at each PLP concentration. Gly, glycine; $Hill_n$, derived Hill coefficient; K_m, substrate concentration; SCoA, succinyl CoA; V_{max}, maximum reaction velocity.

This A-to-G variant was found at position 55042086 on the X chromosome (hg19 coordinates), resulting in a coding change of Y365C in the ALAS2 protein (Figure 1F and Supplemental Figure 1).

The enzyme encoded by the *ALAS2* gene plays a critical role in heme biosynthesis, and mutations in this gene are known to cause a microcytic sideroblastic anemia in male carriers and in females with skewed X inactivation (X-linked sideroblastic anemia [XLSA]; Online Mendelian Inheritance in Man [OMIM] entry no. 300571) (8–15). However, mutations in this gene have only rarely been associated with macrocytic anemia or dyserythropoiesis (14). We confirmed that this mutation was found in all affected individuals through Sanger sequencing (Figure 1F), while the unaffected individuals in this family lacked this mutation, further supporting its causality for the phenotype. Additionally, another BM aspirate from the proband revealed occasional cells with siderotic granules, while nearly all erythroid cells showed dyserythropoietic features, even in the absence of such granules (Figure 1, A–D).

By modeling this novel ALAS2 Y365C mutation in the structure of the *Rhodobacter capsulatus* homolog, we noted that Y365 fits within a hydrophobic core critical for binding the essential cofactor pyridoxal 5′-phosphate (PLP) (Figure 2A and ref. 16). The C365 mutation would disrupt this hydrophobic core and would therefore be predicted to reduce the ability of ALAS2 to bind to PLP (Figure 2A). Consistent with this idea, we found that the ALAS2 C365 mutant purified to homogeneity had a 70-fold lower affinity for PLP than did the WT enzyme (Table 1). The ALAS2 C365 mutant had enzyme kinetics similar to that of the WT enzyme (Table 1). However, the mutant enzyme was unstable, consistent with the notion that PLP binding plays a key role in stabilizing ALAS2 (17, 18). When the enzymes were expressed as maltose-binding protein (MBP) fusions, the yield of the mutant enzyme in the initial crude extract was only 3.3% of the WT enzyme yield (Figure 2B). After removal of the MBP and purification by liquid chromatography (Figure 2C), only 5.5% of the mutant enzyme activity was recovered compared with 17.9% of the WT enzyme activity (Table 1). Overall, the yield relative to that of WT was only 0.2% — lower than the residual activities in classic cases of XLSA (15). Furthermore, the mutant enzyme was 6-fold less stable than the WT enzyme at 50°C (Table 1). Under conditions of increased PLP, the mutant was 11-fold less stable because the PLP stabilized the WT, but not the mutant, enzyme (Table 1). Together, these findings indicate that the Y365C mutation markedly impairs PLP binding, which may account for some or all of the substantially reduced stability of the ALAS2 enzyme. Severe LOF due to this mutation may explain why no male carriers were identified in the family (Figure 1E) and why the proband's anemia did not improve with pyridoxine therapy.

As only female carriers were found to be affected, we reasoned that there may have been highly skewed X inactivation present in the individuals in this family. We were surprised to find that 62%–82% of the active X chromosomes in the proband and other affected family members contained the *ALAS2* mutant, as assessed by the human androgen receptor gene polymorphism assay (HUMARA) for X inactivation in whole-blood genomic DNA (Figure 2D). These values were all within 2 SD of the mean X inactivation ratios

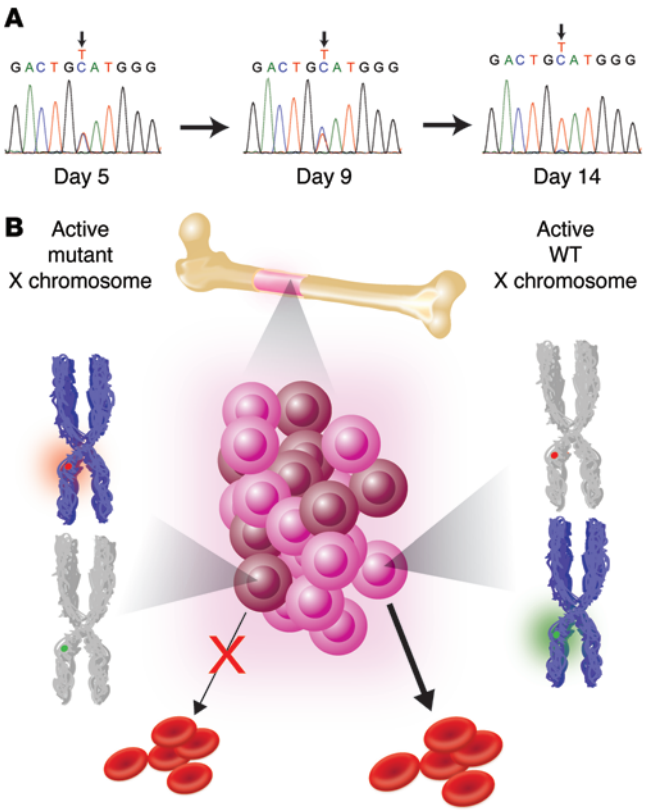


Figure 3. X-linked dominant LOF *ALAS2* mutations can result in macrocytosis and dyserythropoiesis through cell-nonautonomous effects. (**A**) Sequencing of cDNA derived from cultured erythroid cells from the proband showed skewing toward the WT *ALAS2* allele at the late stages of erythroid differentiation. Cells began as progenitors (day 5), became intermediate erythroblasts (day 9), and transitioned to orthochromatic erythroblasts (day 14). By late erythropoiesis, only a trace amount of the mutant allele was detectable. (**B**) Model showing how developing erythroid progenitors and precursors compete for space within the BM. The active X chromosome expressed in each group of cells (brown and pink for the mutant and WT, respectively) is shown in blue, with some cells expressing the mutant (red) and others expressing the WT (green) allele.

observed in healthy adult females (19). We obtained independent confirmation for a lack of highly skewed X inactivation by digesting genomic DNA from the affected family members with the methylation-sensitive restriction enzyme HpaII and noted no change in the intensity of the *ALAS2* mutation by Sanger sequencing (Figure 1F and Supplemental Figure 2). The proband, who was the most severely affected, had the greatest extent of skewing toward the mutant allele, at 82% (Figure 2D). Collectively, these findings suggest that the presence of some erythroid cells expressing WT ALAS2 in the patients' BM was insufficient to compensate for the presence of erythroid cells expressing the *ALAS2* mutation.

While most *ALAS2* mutations causing sideroblastic anemia are partial LOF alleles affecting heme biosynthesis (12, 13, 15), complete *ALAS2* LOF in mouse models results in an early block of erythropoiesis at approximately the proerythroblast or basophilic erythroblast stages (20). Consistent with these results, when we sequenced peripheral blood reticulocyte mRNA samples, we only detected the presence of transcripts with the WT allele, while the mutant allele was readily detectable in genomic DNA (Figure 2E). To confirm these findings, peripheral blood mononuclear cells from the proband were cultured in medium that promotes erythroid differentiation. We found that as the cells transitioned from the early progenitor stages to terminally differentiated erythroblasts, there was selection against cells expressing the mutant *ALAS2* mRNA, leaving only cells expressing mRNA from the WT allele (Figure 3A). Thus, the X inactivation and mRNA sequencing studies show that while there is initially a mixed population of erythroid progenitors and early precursors that express both *ALAS2* alleles, as cells undergo erythroid differentiation, there is a loss of nonviable cells expressing the mutant allele, with a survival advantage for cells expressing the WT allele (Figure 3B and ref. 14). This pressure to produce a normal output of erythroid cells may affect differentiation and result in skipped cell divisions to ensure adequate erythroid output, which would explain the macrocytosis and dyserythropoiesis observed among those cells expressing WT *ALAS2* alleles (ref. 21 and Figure 3B).

It is unclear whether there is a connection between the *ALAS2* mutation and the proband's congenital defects in other tissues (i.e., atrial septal defect, horseshoe kidney) and her abnormal placentation, which were not present in other family members. We examined known genetic causes of iron overload in the proband, including mutations in the *ALAS2*, *HAMP*, *HFE*, *HFE2*, *SLC40A1*, and *TFR2* genes, and only the mutation in *ALAS2* was identified. We also observed low levels of hepcidin in the blood of the proband (Supplemental Table 1), consistent with the expected reduction of hepcidin that occurs in the context of ineffective erythropoiesis (22).

Our findings illustrate how cell-nonautonomous defects could arise as a result of a mutant *ALAS2* allele on one X chromosome, even in the absence of significantly skewed X inactivation. Few examples exist of such mutations in human disease (23). We have only identified a single family with this phenotype, but as more cases of macrocytic dyserythropoietic anemia are studied, it is likely that other similar mutations in *ALAS2* will be identified. While XLSA is typically a dimorphic, microcytic anemia in females, sufficiently deleterious *ALAS2* mutations can result in macrocytic anemia without a dimorphic population in such females. We illustrate how defining such mutations using modern genomic sequencing can help lead to further insight into human disease and hematopoiesis (7).

Methods

Further information can be found in the Supplemental Methods.

The whole-exome sequencing data on the family members are available in the dbGaP database (http://www.ncbi.nlm.nih.gov/gap) under the accession number phs000474.v2.p1.

Statistics. All pairwise comparisons were performed using the 2-tailed Student's *t* test, unless otherwise indicated. Differences were considered significant if the *P* value was less than 0.05.

Study approval. All family members provided written informed consent to participate in this study. The IRBs of the Dana-Farber Cancer Institute, Boston Children's Hospital, and Massachusetts Institute of Technology approved the study protocols.

Acknowledgments

We are grateful to the family for their interest in this study. We thank D. Nathan, S. Lux, C. Brugnara, F. Briccetti, M. Westerman, J. Hirschhorn, G. Evrony, R. Do, D. Bennett, and M. Burns for their valuable advice and support. This work was supported by NIH grants R21 HL120791 and R01 DK103794 (to V.G. Sankaran) and by a grant from the Edward P. Evans Foundation (to D.P. Steensma).

Address correspondence to: Vijay G. Sankaran, Boston Children's Hospital, 3 Blackfan Circle, CLS 03001, Boston, Massachusetts 02115, USA. Phone: 617.919.6270; E-mail: sankaran@broadinstitute.org. Or to: David P. Steensma, Dana-Farber Cancer Institute, 450 Brookline Avenue, Boston, Massachusetts 02215, USA. Phone: 617.632.3712; E-mail: David_Steensma@dfci.harvard.edu.

Rafael Bejar's present address is: University of California, San Diego, La Jolla, California 92093, USA.

1. Hoffbrand V, Provan D. ABC of clinical haematology. Macrocytic anaemias. *BMJ*. 1997;314(7078):430–433.
2. Davenport J. Macrocytic anemia. *Am Fam Physician*. 1996;53(1):155–162.
3. Younes M, Dagher GA, Dulanto JV, Njeim M, Kuriakose P. Unexplained macrocytosis. *South Med J*. 2013;106(2):121–125.
4. Sankaran VG, Orkin SH. Genome-wide association studies of hematologic phenotypes: a window into human hematopoiesis. *Curr Opin Genet Dev*. 2013;23(3):339–344.
5. Iolascon A, Heimpel H, Wahlin A, Tamary H. Congenital dyserythropoietic anemias: molecular insights and diagnostic approach. *Blood*. 2013;122(13):2162–2166.
6. Sankaran VG, et al. Exome sequencing identifies GATA1 mutations resulting in Diamond-Blackfan anemia. *J Clin Invest*. 2012;122(7):2439–2443.
7. Sankaran VG, Gallagher PG. Applications of high-throughput DNA sequencing to benign hematology. *Blood*. 2013;122(22):3575–3582.
8. Cotter PD, Baumann M, Bishop DF. Enzymatic defect in "X-linked" sideroblastic anemia: molecular evidence for erythroid delta-aminolevulinate synthase deficiency. *Proc Natl Acad Sci U S A*. 1992;89(9):4028–4032.
9. Bottomley SS, May BK, Cox TC, Cotter PD, Bishop DF. Molecular defects of erythroid 5-aminolevulinate synthase in X-linked sideroblastic anemia. *J Bioenerg Biomembr*. 1995;27(2):161–168.
10. Cotter PD, et al. Four new mutations in the erythroid-specific 5-aminolevulinate synthase

(ALAS2) gene causing X-linked sideroblastic anemia: increased pyridoxine responsiveness after removal of iron overload by phlebotomy and coinheritance of hereditary hemochromatosis. *Blood*. 1999;93(5):1757–1769.
11. Cazzola M, May A, Bergamaschi G, Cerani P, Rosti V, Bishop DF. Familial-skewed X-chromosome inactivation as a predisposing factor for late-onset X-linked sideroblastic anemia in carrier females. *Blood*. 2000;96(13):4363–4365.
12. Bergmann AK, et al. Systematic molecular genetic analysis of congenital sideroblastic anemia: evidence for genetic heterogeneity and identification of novel mutations. *Pediatr Blood Cancer*. 2010;54(2):273–278.
13. Fleming MD. Congenital sideroblastic anemias: iron and heme lost in mitochondrial translation. *Hematology Am Soc Hematol Educ Program*. 2011;2011:525–531.
14. Aivado M, et al. X-linked sideroblastic anemia associated with a novel ALAS2 mutation and unfortunate skewed X-chromosome inactivation patterns. *Blood Cells Mol Dis*. 2006;37(1):40–45.
15. Ducamp S, et al. Sideroblastic anemia: molecular analysis of the ALAS2 gene in a series of 29 probands and functional studies of 10 missense mutations. *Hum Mutat*. 2011;32(6):590–597.
16. Astner I, Schulze JO, van den Heuvel J, Jahn D, Schubert WD, Heinz DW. Crystal structure of 5-aminolevulinate synthase, the first enzyme of heme biosynthesis, and its link to XLSA in humans. *EMBO J*. 2005;24(18):3166–3177.
17. Bishop DF, Tchaikovskii V, Hoffbrand AV, Fraser ME, Margolis S. X-linked sideroblastic anemia due to carboxyl-terminal ALAS2 mutations that cause loss of binding to the β-subunit of succinyl-CoA synthetase (SUCLA2). *J Biol Chem*. 2012;287(34):28943–28955.
18. Cotter PD, Rucknagel DL, Bishop DF. X-linked sideroblastic anemia: identification of the mutation in the erythroid-specific Δ-aminolevulinate synthase gene (ALAS2) in the original family described by Cooley. *Blood*. 1994;84(11):3915–3924.
19. Amos-Landgraf JM, et al. X chromosome-inactivation patterns of 1,005 phenotypically unaffected females. *Am J Hum Genet*. 2006;79(3):493–499.
20. Harigae H, et al. Aberrant iron accumulation and oxidized status of erythroid-specific Δ-aminolevulinate synthase (ALAS2)-deficient definitive erythroblasts. *Blood*. 2003;101(3):1188–1193.
21. Sankaran VG, et al. Cyclin D3 coordinates the cell cycle during differentiation to regulate erythrocyte size and number. *Genes Dev*. 2012;26(18):2075–2087.
22. Kautz L, Jung G, Valore EV, Rivella S, Nemeth E, Ganz T. Identification of erythroferrone as an erythroid regulator of iron metabolism. *Nat Genet*. 2014;46(7):678–684.
23. Migeon BR. The role of X inactivation and cellular mosaicism in women's health and sex-specific diseases. *JAMA*. 2006;295(12):1428–1433.

Supplemental Data:

X-Linked Macrocytic Dyserythropoietic Anemia in Females with an *ALAS2* Mutation

Vijay G. Sankaran, Jacob C. Ulirsch, Vassili Tchaikovskii, Leif S. Ludwig, Aoi Wakabayashi, Senkottuvelan Kadirvel, R. Coleman Lindsley, Rafael Bejar, Jiahai Shi, Scott B. Lovitch, David F. Bishop, David P. Steensma

Address correspondence to:
Vijay G. Sankaran, Boston Children's Hospital, 3 Blackfan Circle, CLS 03001, Boston, MA 02115; Phone: 617-919-6270; E-mail: sankaran@broadinstitute.org.
Or to: David P. Steensma, Dana-Farber Cancer Institute, 450 Brookline Avenue, Boston, MA 02215, Phone: 617-632-3712; E-mail: David_Steensma@dfci.harvard.edu

Supplemental Methods:

Whole-Exome Sequencing of Family Members

DNA was extracted and libraries were prepared for whole-exome sequencing (1). Sequencing reads were analyzed with an automated pipeline built around the Genome Analysis Toolkit. Mutation counts both before and after filtration of common variants are shown in Supplemental Table 2, and this analysis was performed as described previously (2). All data on polymorphisms from these individuals was deposited in the database of Genotypes and Phenotypes (dbGaP) from the National Center for Biotechnology Information.

HUMARA Assay

Analysis of X-inactivation was performed on genomic DNA extracted from peripheral blood mononuclear cells. Some DNA was digested with the methylation-sensitive enzyme HpaII and then PCR amplification of the androgen receptor (*AR*) locus was performed. The amount of each allele present in either sample with or without treatment with HpaII was quantitated on a DNA Analyzer 3730 (Applied Biosystems Inc.) using the fluorescein channel with analysis performed on GeneMapper software. The assay was carried out in a manner similar to that described previously (3).

Sanger Sequencing

Variants of interest were confirmed using PCR of regions of interest in the *ALAS2* gene. For analysis of samples after HpaII digestion, a similar PCR amplification of the regions of interest in the *ALAS2* gene was performed and subjected to Sanger sequencing as noted above. HpaII digestion results in cuts in the *ALAS2* gene between the mutation site and the PCR amplification sites, thus affecting the Sanger signal from DNA that is unmethylated.

ALAS2 Expression and Purification

The wild-type human ALAS2 expression construct, pMALc2-AE2 (4) was modified to introduce the Y365C mutation by site-directed mutagenesis using the Stratagene XL Site-directed mutagenesis kit. Transformation, expression, and purification to homogeneity were performed as previously described (5). For purification by gel filtration chromatography, a Superdex 200 (GE Healthcare) column was used. For size exclusion, proteins were separated by chromatography on two tandem HiLoad 16/600 Superdex 200 size-exclusion columns (GE Healthcare) at a flow rate of 0.2 ml/min.

Enzyme and Protein Assays

ALAS2 enzymatic activity was determined using succinyl CoA, glycine, and pyridoxal-5' phosphate in five minute endpoint assays and quantitated by chemical condensation with ethylacetoacetate followed by reaction with Ehrlich's reagent as described previously (5). Protein was assayed by the Fluorescamine method as previously described (6). For heat stability experiments, the enzymes were preincubated at 50 °C for up to 30 min in 50 mM Hepes, pH 7.4, 10 mM $MgCl_2$, 1 mM DTT, and 10 or 100 uM PLP, and then assayed with the addition of 0.5 mg/ml BSA to protect against inactivation upon dilution in the assay.

Protein Modeling

Modeling was performed using the crystal structure of the *Rhodobacter capsulatus* ALAS2 homologue (7). Modeling was performed using PyMOL (http://www.pymol.org/).

Mononuclear Cell Erythroid Cultures

Peripheral blood mononuclear cells were isolated using Ficoll-Paque Plus (GE Healthcare) and cultured in IMDM media containing human AB serum, human AB plasma, holo-transferrin, insulin, erythropoietin, stem cell factor, and IL-3 as previously described (8).

Supplemental Table 1: Hematological And Other Laboratory Parameters In Affected Family Members

Parameter	Proband (II-2)	Proband's Sister (II-4)	Proband's Mother (I-1)
Hemoglobin (g/dL) [normal, 12.0-15.0 g/dL]	11.0*	11.2*	12.0
Hematocrit (%) [normal, 37.0-47.0%]	32.1*	32.5*	36.1*
Mean cell volume (fL) [normal, 81-95 fL]	114*	108*	106*
Mean cell hemoglobin (pg) [normal, 27.6-33.9 pg]	38.9*	37.4*	35.0*
Red blood cell count ($x10^9$/L) [normal, 4.2-5.5 $x10^9$/L]	2.81*	3.04*	3.72*
Red cell distribution width [normal, 11-13.5%]	12.7%	11.8%	12.4%
Lactate dehydrogenase (U/L) [normal, 107-231 U/L]	143	ND	ND
Absolute and proportional reticulocyte count (M/µL; %) [normal, 0.043-0.095 M/µL; 0.6-2.8%]	0.0520; 1.8	ND	ND
Haptoglobin (mg/dL) [normal, 40-200 mg/dL]	75	ND	ND
Serum erythropoietin (U/L) [normal, 7-20 U/L]	86*	ND	ND
Serum ferritin (ng/mL) [normal, 10-170 ng/mL]	598-1030*	104	230*
Transferrin saturation (%) [normal, 12-45%]	96*, 83* on repeat testing	28	39
MRI LIC estimate (µmol/g dry weight) [normal, <36 µmol/g dry weight]	>350*	ND	ND
MRI cardiac T2* (ms) [normal, >20 ms]	49	ND	ND
Liver biopsy iron index [normal, <1.0]	11*	ND	ND
Serum hepcidin (ng/mL) [normal, 17-286 ng/ml]	32.3	ND	ND
HFE genotype	Wild-type	Wild-type	Wild-type
Other features	Horseshoe kidney; atrial septal defect (secundum type); placenta accreta; neonatal febrile seizures; recurrent childhood UTIs; adenocarcinoma of the uterine cervix	None	Recurrent miscarriages; brittle fingernails (telomere assay WNL)

Asterisks denote values outside the normal range. Proband's leukocyte count and differential, platelet count, hemoglobin electropheresis, routine coagulation tests (prothrombin time, activated partial thromboplastin time), vitamin B12 and folate levels, uric acid and general chemistry group (including total and direct bilirubin, transaminases, creatinine and blood urea nitrogen) were all within normal limits. Family is of mixed Irish, Polish, and German ancestry. Abbreviations: MRI=magnetic resonance imaging. LIC = liver iron concentration. WNL = within normal limits. UTI = urinary tract infection. ND = not done.

Supplemental Table 2: Classes of variants in a family with macrocytic anemia and dyserythropoiesis

	Proband (Individual II-2)	Proband's Mother (Individual I-1)	Proband's Sister (Individual II-4)	Proband's Daughter (Individual III-1) (Unaffected)
Novel				
CODON CHANGE PLUS CODON DELETION	28	19	15	23
CODON CHANGE PLUS CODON INSERTION	10	10	12	13
CODON DELETION	49	43	42	51
CODON INSERTION	6	10	7	11
DOWNSTREAM	215	192	202	226
FRAMESHIFT	24	21	28	31
INTRAGENIC	5	3	3	6
INTRON	403	403	371	382
NONSYNONYMOUS CODING	172	176	182	216
NONSYNONYMOUS START	1	0	0	0
SPLICE SITE ACCEPTOR	8	8	4	6
SPLICE SITE DONOR	2	4	4	3
START GAINED	1	1	1	3
START LOST	2	0	0	3
STOP GAINED	2	2	1	3
STOP LOST	0	1	0	0
SYNONYMOUS CODING	114	106	101	126
Total	1095	1044	1026	1146
TRANSCRIPT	15	15	15	14
UPSTREAM	18	6	15	7
UTR 3-PRIME	4	10	9	5
UTR 5-PRIME	16	14	14	17
Known				
CODON CHANGE PLUS CODON DELETION	93	87	82	89
CODON CHANGE PLUS CODON INSERTION	33	31	42	32
CODON DELETION	130	125	124	132
CODON INSERTION	88	88	91	90
DOWNSTREAM	8685	8284	8510	8515
FRAMESHIFT	282	268	280	267
INTERGENIC	52	61	55	51
INTRAGENIC	86	90	78	89
INTRON	15350	14982	15368	15297

NONSYNONYMOUS CODING	10075	9941	10017	10054
NONSYNONYMOUS START	1	3	2	1
SPLICE SITE ACCEPTOR	87	82	84	85
SPLICE SITE DONOR	63	57	61	56
START GAINED	149	147	160	156
START LOST	21	21	23	22
STOP GAINED	72	73	72	73
STOP LOST	16	18	18	16
SYNONYMOUS CODING	10988	10905	10927	11006
SYNONYMOUS STOP	12	10	9	8
Total	48133	47024	47794	47875
TRANSCRIPT	956	899	918	930
UPSTREAM	361	342	359	368
UTR 3-PRIME	324	319	318	340
UTR 5-PRIME	209	191	196	198

Supplemental References:

1. Sankaran, V.G., and Gallagher, P.G. 2013. Applications of high-throughput DNA sequencing to benign hematology. *Blood* 122:3575-3582.

2. Sankaran, V.G., Ghazvinian, R., Do, R., Thiru, P., Vergilio, J.A., Beggs, A.H., Sieff, C.A., Orkin, S.H., Nathan, D.G., Lander, E.S., et al. 2012. Exome sequencing identifies GATA1 mutations resulting in Diamond-Blackfan anemia. *The Journal of clinical investigation* 122:2439-2443.

3. Thouin, M.M., Giron, J.M., and Hoffman, E.P. 2003. Detection of nonrandom X chromosome inactivation. *Current protocols in human genetics / editorial board, Jonathan L. Haines ... [et al.]* Chapter 9:Unit9 7.

4. Cotter, P.D., Rucknagel, D.L., and Bishop, D.F. 1994. X-linked sideroblastic anemia: identification of the mutation in the erythroid-specific delta-aminolevulinate synthase gene (ALAS2) in the original family described by Cooley. *Blood* 84:3915-3924.

5. Bishop, D.F., Tchaikovskii, V., Hoffbrand, A.V., Fraser, M.E., and Margolis, S. 2012. X-linked sideroblastic anemia due to carboxyl-terminal ALAS2 mutations that cause loss of binding to the beta-subunit of succinyl-CoA synthetase (SUCLA2). *The Journal of biological chemistry* 287:28943-28955.

6. Bishop, D.F., Wampler, D.E., Sgouris, J.T., Bonefeld, R.J., Anderson, D.K., Hawley, M.C., and Sweeley, C.C. 1978. Pilot scale purification of alpha-galactosidase A from Cohn fraction IV-1 of human plasma. *Biochimica et biophysica acta* 524:109-120.

7. Astner, I., Schulze, J.O., van den Heuvel, J., Jahn, D., Schubert, W.D., and Heinz, D.W. 2005. Crystal structure of 5-aminolevulinate synthase, the first enzyme of heme biosynthesis, and its link to XLSA in humans. *The EMBO journal* 24:3166-3177.

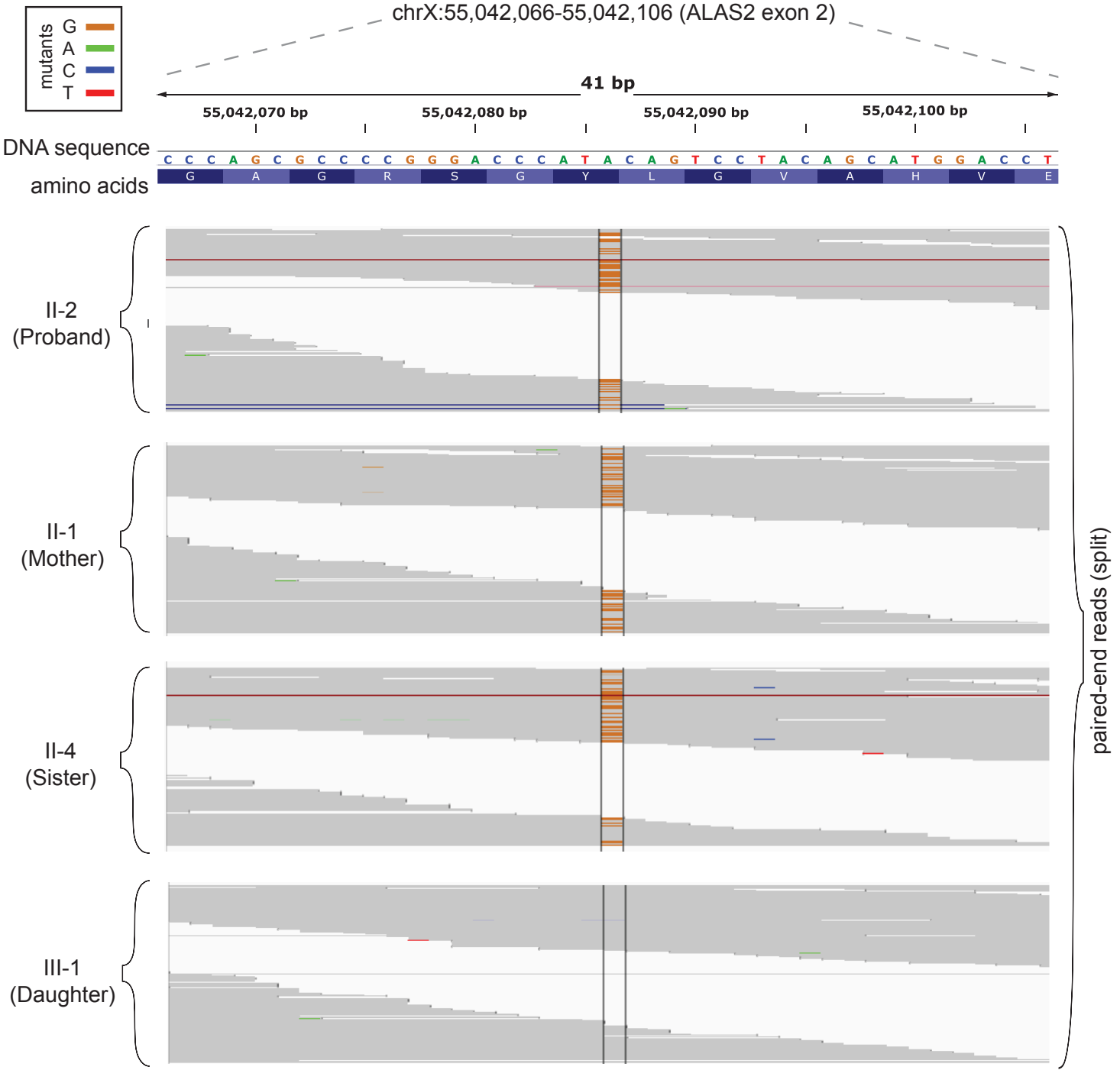
8. Ludwig, L.S., Gazda, H.T., Eng, J.C., Eichhorn, S.W., Thiru, P., Ghazvinian, R., George, T.I., Gotlib, J.R., Beggs, A.H., Sieff, C.A., et al. 2014. Altered translation of GATA1 in Diamond-Blackfan anemia. *Nature medicine* 20:748-753.

Supplemental Figure Legends:

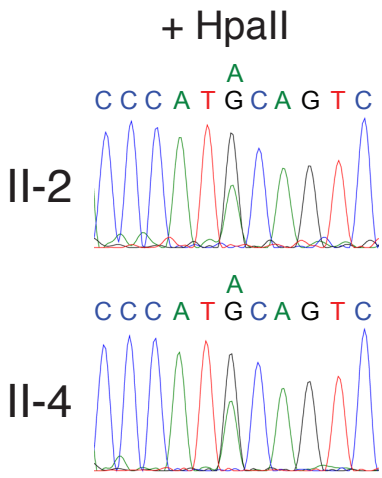
Supplemental Figure 1. Integrated Genomics Viewer was used to visualize sequence reads from all family members who underwent whole-exome sequencing to validate number of reads at position chrX: 55042086 (hg19 coordinates) and show those reads with the variant (mutant) allele present.

Supplemental Figure 2. Sanger sequencing traces following HpaII digestion of genomic DNA showing the region around 55042086 on the X chromosome (hg19 coordinates, trace is showing the positive strand on the X chromosome).

Supplemental Figure 1



Supplemental Figure 2



Targeted Application of Human Genetic Variation Can Improve Red Blood Cell Production from Stem Cells

Giani FG, Fiorini C, Wakabayashi A, Ludwig LS, Jobaliya CD, Regan SN, Ulirsch JC, Liang G, Steinberg-Shemer O, Esko T, Hirschhorn JN, Tong W, Brugnara C, Weiss MJ, Zon LI, Chou ST, French DL, Musunuru K and Sankaran VG.

Cell Stem Cell, 2016, Jan 7;18(1):1-6.

The original article is available online:
http://dx.doi.org/10.1016/j.stem.2015.09.015
http://www.cell.com/cell-stem-cell/abstract/S1934-5909%2815%2900422-1

Cell Stem Cell

Supplemental Information

Targeted Application of Human Genetic Variation Can Improve Red Blood Cell Production from Stem Cells

Felix C. Giani, Claudia Fiorini, Aoi Wakabayashi, Leif S. Ludwig, Rany M. Salem, Chintan D. Jobaliya, Stephanie N. Regan, Jacob C. Ulirsch, Ge Liang, Orna Steinberg-Shemer, Michael H. Guo, Tõnu Esko, Wei Tong, Carlo Brugnara, Joel N. Hirschhorn, Mitchell J. Weiss, Leonard I. Zon, Stella T. Chou, Deborah L. French, Kiran Musunuru, and Vijay G. Sankaran

Figure S1

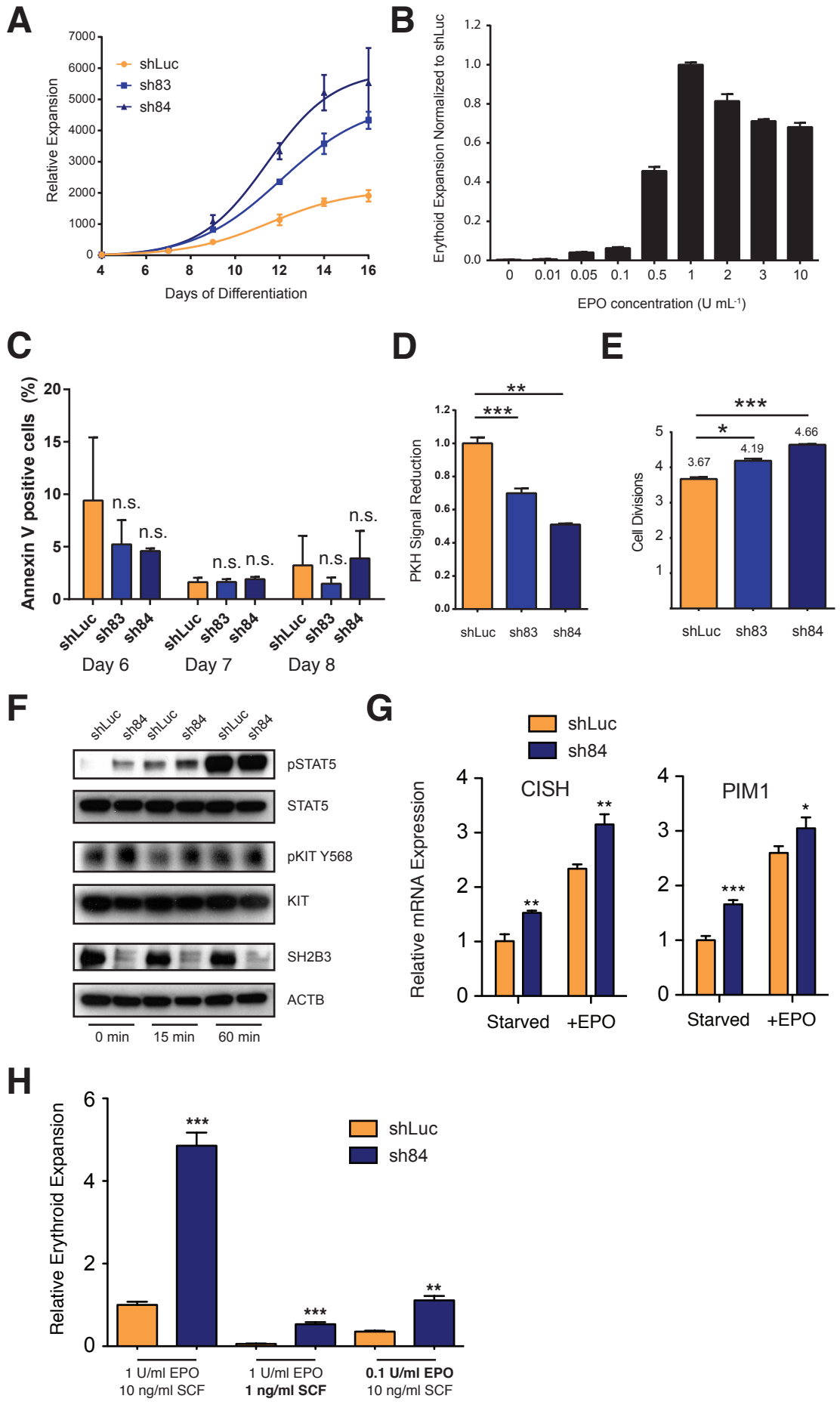

A
Relative Expansion
shLuc
sh83
sh84
Days of Differentiation
B
Erythoid Expansion Normalized to shLuc
EPO concentration (U mL-1)
C
Annexin V positive cells (%)
n.s.
shLuc
sh83
sh84
Day 6
Day 7
Day 8
D
PKH Signal Reduction
E
Cell Divisions
3.67
4.19
4.66
F
pSTAT5
STAT5
pKIT Y568
KIT
SH2B3
ACTB
0 min
15 min
60 min
G
Relative mRNA Expression
CISH
PIM1
Starved
+EPO
H
Relative Erythroid Expansion
1 U/ml EPO
10 ng/ml SCF
1 ng/ml SCF
0.1 U/ml EPO

Figure S1. Characterization of Erythroid Cells from HSPCs Following SH2B3 Suppression (Related to Figure 2).

(A) Mean amplification of control and SH2B3 KD peripheral blood-mobilized CD34+ HSPCs. Values shown are mean ± standard error of the mean (n = 3 independent experiments representative of > 3 experiments using different donors).

(B) Relative expansion of adult HSPCs observed for the indicated EPO concentrations. Erythroid expansion is calculated as total cell expansion until day 18 corrected for the percentage of CD235a positive cells observed by flow cytometry analysis on day 18. Values shown are mean ± the standard deviation and are normalized to the expansion observed at 1 U/mL (n = 3 per concentration).

(C) Bar graphs showing the frequency of annexin V+ apoptotic or dead control and SH2B3-KD samples on the indicated days of erythroid differentiation of adult HSPCs. Values show mean percentages ± the standard deviation. A two-tailed Student's t-test was performed (n = 3, n.s. = non-significant).

(D) Signal reduction of the fluorescent membrane dye PKH observed for control and SH2B3-KD adult HSPC derived erythroid cells. Ratios of PKH signal on day 9 compared with day 5, as determined by flow cytometry, are shown. Values are mean ± the standard deviation and were normalized to the control. The comparison was performed by a two-tailed Student's t-test (n = 3; ** $P < 0.01$; *** $P < 0.001$).

(E) Absolute number of cell divisions based on PKH signal reduction observed for adult HPSC-derived erythroid cells between days 5 and 9 of culture. Values represent mean ± the standard deviation. Comparisons were performed with a two-tailed Student's t-test (n = 3; * $P < 0.05$; ** $P < 0.01$, *** $P < 0.001$)

(F) Western blots showing activation of phosphorylated STAT5 (pSTAT5) and phosphorylated KIT Y568 (pKIT Y568) in the human TF-1 erythroid cell line transduced with control (shLuc) or *SH2B3* (sh84) shRNAs following cytokine starvation and stimulation with EPO and SCF. Cells were selected in puromycin for 48 hours before cytokine starvation. pKIT Y568 can be completely eliminated with serum starvation of cells, suggesting baseline stimulation with serum alone. STAT5 activation only

occurs downstream of specific cytokine receptors, such as the EPO receptor.

(G) Expression of EPO responsive genes CISH and PIM1 with SH2B3 suppression in TF-1 erythroid cells undergoing overnight cytokine starvation or following 2 hours of stimulation with EPO. Values represent the mean ± the standard deviation. Comparisons were performed with a two-tailed Student's t-test (n=3; * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$).

(H) Relative expansion of adult HSPCs corrected for CD235a positive cells at varying EPO and SCF concentrations as shown. Cultures were done using the more efficient differentiation protocol discussed in the Methods involving an expansion step prior to differentiation in the three-phase system, where EPO and SCF concentrations were varied in all phases. Values shown are the mean ± the standard error of the mean. Comparisons were performed with a two-tailed Student's t-test (n=3; ** $P < 0.01$; *** $P < 0.001$).

Figure S2

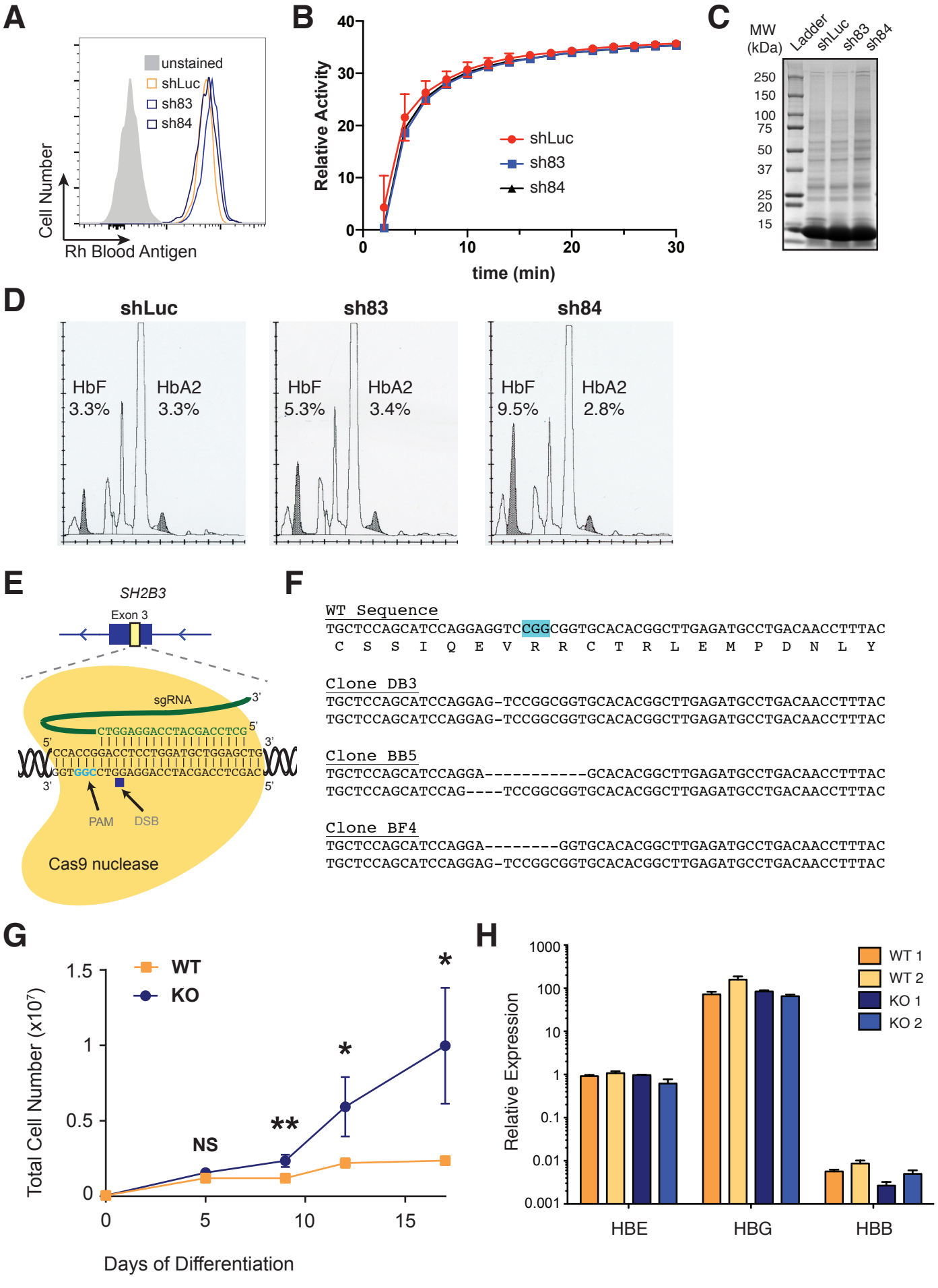

A
unstained
shLuc
sh83
sh84
Cell Number
Rh Blood Antigen
B
Relative Activity
time (min)
shLuc
sh83
sh84
C
MW (kDa)
Ladder
shLuc
sh83
sh84
250
150
100
75
50
37
25
20
15
D
shLuc
sh83
sh84
HbF 3.3%
HbA2 3.3%
HbF 5.3%
HbA2 3.4%
HbF 9.5%
HbA2 2.8%
E
SH2B3
Exon 3
sgRNA
CTGGAGGACCTACGACCTCG
CCACCGGACCTCCTGGATGCTGGAGCTG
GGTGGCCTGGAGGACCTACGACCTCGAC
PAM
DSB
Cas9 nuclease
F
WT Sequence
TGCTCCAGCATCCAGGAGGTCCGGCGGTGCACACGGCTTGAGATGCCTGACAACCTTTAC
C S S I Q E V R R C T R L E M P D N L Y
Clone DB3
TGCTCCAGCATCCAGGAG-TCCGGCGGTGCACACGGCTTGAGATGCCTGACAACCTTTAC
TGCTCCAGCATCCAGGAG-TCCGGCGGTGCACACGGCTTGAGATGCCTGACAACCTTTAC
Clone BB5
TGCTCCAGCATCCAGGA----------GCACACGGCTTGAGATGCCTGACAACCTTTAC
TGCTCCAGCATCCAG----TCCGGCGGTGCACACGGCTTGAGATGCCTGACAACCTTTAC
Clone BF4
TGCTCCAGCATCCAGGA--------GGTGCACACGGCTTGAGATGCCTGACAACCTTTAC
TGCTCCAGCATCCAGGAG-TCCGGCGGTGCACACGGCTTGAGATGCCTGACAACCTTTAC
G
WT
KO
Total Cell Number (x10^7)
NS
**
*
*
Days of Differentiation
H
Relative Expression
WT 1
WT 2
KO 1
KO 2
HBE
HBG
HBB

Figure S2. Characterization of Erythroid Cells from SH2B3 LoF HSPCs and Pluripotent Stem Cells (Related to Figure 2).

(A) Representative histogram plots showing Rh blood antigen expression on day 18 mature RBCs derived from adult CD34+ cells.

(B) Pyruvate kinase activity observed for control and SH2B3-KD mature RBCs generated from adult CD34+ cells at the indicated time points. Values shown are the mean ± the standard deviation (n = 2 independent experiments).

(C) Coomassie blue stained SDS-PAGE gels showing major membrane protein bands from RBC ghosts derived from adult RBCs.

(D) Hemoglobin high performance liquid chromatography analysis of mature hemoglobin subtypes in hemolysates from the RBCs produced in culture. The peaks for HbF and HbA2 are shaded in the chromatograms with labels above the corresponding peaks.

(E) Scheme illustrating the single guide RNA (gRNA) utilized to target the depicted *SH2B3* DNA sequence adjacent to the protospacer adjacent motif (PAM) to cause double-strand breaks (DSB) by the Cas9 protein in exon 3 of the *SH2B3* gene.

(F) Sequences for CRISPR/Cas9-targeted hESCs. The PAM sequence is highlighted in blue, and the predicted cleavage site is three bases upstream of the PAM. The allele sequences of one homozygous clone DB3 and two compound heterozygous clones BB5 and BF4 that harbor frameshift mutations in *SH2B3* are shown. Note that the sense DNA sequence is shown.

(G) Cell expansion of hESC-derived HPCs with SH2B3-KO and an isogenic control. HPCs were collected 9 days after initiation of differentiation for this experiment. The total cell number is shown as the mean ± the standard deviation at the various time points. Comparisons were performed with a two-tailed Student's t-test at each indicated day of culture (n = 3; NS = non-significant; * $P < 0.05$; ** $P < 0.01$).

(H) Globin gene expression in hESC-derived erythroid cells (with either *SH2B3* KO or WT alleles) on day 10 of HPC differentiation, as measured by quantitative RT-PCR. Expression relative to β-actin is shown on a $\log_{10}$ scale (n = 3 per group).

SUPPLEMENTAL TABLES

Table S1. Association with LoF and strict damaging missense mutations in the PH/ SH2 domains of *SH2B3* in European American subjects (significant values are shown in bold italic font, Related to Figure 1).

PHENOTYPE	BETA	SE	T	P value	Absent (N)[1]	Present (N) [2]	Total (N) [3]
Hemoglobin	***0.466***	***0.230***	***2.021***	***0.0435***	***1995***	***19***	***2014***
Hematocrit	***0.515***	***0.230***	***2.235***	***0.0255***	***1995***	***19***	***2014***
WBC	0.265	0.230	1.149	0.2508	2027	19	2046
Platelets	0.393	0.243	1.615	0.1064	1798	17	1815
RBC	0.701	0.379	1.847	0.0652	613	7	620
Neutrophils	-0.741	0.579	-1.281	0.2010	437	3	440
Lymphocytes	0.859	0.578	1.485	0.1382	477	3	480
Eosinophils	0.930	0.707	1.316	0.1890	445	2	447
Basophils	0.664	0.708	0.938	0.3489	420	2	422

1. Number of subjects without a LoF or strict damaging mutation in the PH/SH2 domain
2. Number of subjects with a LoF or strict damaging mutation in the PH/SH2 domain
3. Total number of subjects included in analysis for which mutation and phenotype information was available.

SE: standard error, T: T test statistic value,

Table S2. LoF and putative strict damaging missense mutations in the PH/SH2 domains of *SH2B3* in European American subjects (Related to Figure 1).

Chr	Pos(hg19)	Ref Allele	Alt Allele	Amino Acid Change	ESP AF	EXAC AF	SH2B3 Domain
12	111856571	G	C	E208Q	5.15E-04	3.83E-04	PH
12	111856588	C	A	S213R	0	1.43E-04	PH
12	111884594	T	C	I257R	2.54E-03	7.19E-04	PH
12	111884812	G	A	E301K	7.70E-05	3.30E-05	PH
12	111885221	C	G	S370C	7.70E-05	8.24E-06	SH2
12	111885292	A	G	S394G	0	0	SH2
12	111885295	G	A	E395K	6.15E-04	1.24E-04	SH2
12	111885310	G	A	E400K	1.23E-03	4.53E-04	SH2
12	111885496	C	T	R425C	7.70E-05	1.65E-05	SH2
12	111885559	A	G	I446V	7.70E-05	1.65E-05	SH2
12	111885930	C	T	R518*	7.70E-05	3.30E-05	

Alt Allele = Alternative Allele
Ref Allele = Reference Allele
Pos = Position
Chr = Chromosome
AF = Allele frequency

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Cell culture and lentiviral production

293T cells were kept in DMEM supplemented with 10% fetal bovine serum (FBS) and 1% penicillin/streptomycin at 30% - 90% confluency. For lentiviral production, 293T cells were transfected with the constructs described below along with the VSV-G envelope and pΔ8.9 packaging vector using the Fugene 6 reagent (Roche) according to the manufacturer's protocol. Medium was changed to phase 1 primary cell culture medium (without EPO, IL-3, and SCF) the following day, and viral supernatant was collected and filtered at 45 μm at 48 h post-transfection. TF-1 human erythroid cells were cultured in RPMI media supplemented with 10% FBS and 2 ng/ml GM-CSF for maintenance and 1% penicillin/streptomycin. For cytokine starvation, all cytokines were removed and cells were maintained in FBS only overnight. EPO and SCF were added at time 0 and then subsequent time points were collected for analysis.

Isolation and source of primary CD34+ cells.

CD34+ cells from G-CSF mobilized peripheral blood, bone marrow, or cord blood was purified by positive magnetic selection using an Ultrapure Microbead Kit (Miltenyi Biotech) according to the manufacturer's protocol following mononuclear cell purification on a Ficoll Density Gradient. At least 95% purity was reached as assessed by post-purification flow cytometry with a PE conjugated anti-human CD34 antibody (8012-0349, eBioscience), as described below.

Primary cell culture and lentiviral infection

Cells were differentiated into mature RBCs utilizing a three-phase culture protocol. In phase 1 (day 0 – 7), cells were cultured at a density of 10^5 - 10^6 cells per milliliter (mL) in IMDM supplemented with 2% human AB plasma, 3% human AB serum, 1% penicillin/streptomycin, 3 IU/mL heparin, 10 ug/mL insulin, 200 ug/mL holo-transferrin, 1 IU erythropoietin (Epo), 10 ng/mL stem cell factor (SCF) and 1 ng/mL IL-3. In phase 2 (day 7 – 12), IL-3 was omitted from the medium. In phase 3 (day 12 – 18), cells

were cultured at a density of 10^6 cells per milliliter, with both IL-3 and SCF omitted from the medium and the holo-transferrin concentration was increased to 1 mg/mL. Cells were cultured at 37°C and 5% CO_2. For lentiviral infection, medium was changed on day 1 to viral supernatant along with 8 ug/mL polybrene and EPO, IL-3, and SCF and spun at 2000 rpm for 90 min at room temperature. After 12 h of infection, cells were placed in 1 ug/mL puromycin for 30-48 h.

As an alternative approach, we also developed an adaptation of the erythroid culture method described above by adding an expansion phase lasting a total of 5 days prior to initiation of erythroid differentiation and starting phase 1 of the culture. Recent work has shown that such an approach can result in significant expansion of erythroid cells and can improve mature RBC production (Lee et al., 2015). Similar to those reported findings, we find that this allows for increased expansion and comparable differentiation as our standard 3-phase culture method. During the expansion phase >85% of cells remain CD34+. The expansion medium was composed of StemSpan II serum free expansion medium (Stem Cell Technologies) and 1X CC100 cytokine mix composed of FLT3L, IL-3, IL-6, and SCF (Stem Cell Technologies), as we have described previously (Sankaran et al., 2011; Sankaran et al., 2008). Media was changed every 2 – 3 days and cells were maintained at a density between 10^5 - 10^6 cells per milliliter (mL). Lentiviral infections were carried out on day 2 of this expansion phase and selection occurred, as described above.

Western Blotting

1– 2 x 10^6 cells were harvested, washed twice in PBS, and resuspended in RIPA buffer (50mM Tris-HCl, 150 mM NaCl, 0.1 % SDS, 1 % NP-40, 0.25% sodium deoxycholate) and lysed in the presence of sodium orthovanadate, protease inhibitor cocktail, and PMSF for 30 min on ice with intermittent mixing. After removal of cell debris by centrifugation, the supernatant was transferred to a new tube and the protein concentration was determined mixed with Loading Dye and incubated at 95°C for 5 min. SDS-PAGE analysis was performed using precast gels (BioRad) and 1% Tris-Glycine-SDS running buffer. Protein was transferred onto nitrocellulose or PVDF membranes in 10% methanol in

Tris/glycine. Membranes were blocked with TBS-T supplemented with 3% BSA for 1 h and incubated with the primary antibody for 1 h at room temperature or overnight at 4°C. Subsequent to washing four times in TBS-T for 10 min, the membrane was incubated with the secondary antibody for 1 h at room temperature. After another 40 minutes of washing in TBS-T and 5 min incubation in ECL-substrate (BioRad), proteins were visualized by scientific imaging film exposed to the treated membrane. LNK/SH2B3 sheep polyclonal antibody (AF5888, R&D Systems) and GAPDH mouse monoclonal antibody (6C5, sc-32233, Santa Cruz) served as primary, while goat anti-mouse (170-5047, BioRad) and donkey anti-sheep (713-035-147, Jackson Immuno Research) HRP-coupled antibodies served as secondary reagents. For the Western blots shown, cells were typically harvested at day 9 of culture.

Flow Cytometry

Cells were harvested and washed twice in FACS Buffer (PBS + 3 % FBS) for 5 min at 300 x g. For differentiation analysis, cells were subsequently stained for 10 - 15 min on ice with the following antibodies at a 1:20 dilution in FACS Buffer: APC conjugated anti-human CD235a (17-9987, eBioscience), FITC conjugated anti-human CD71 (11-0719, eBioscience), PE conjugated anti-human CD41a (12-0419, eBioscience), and PE conjugated anti-human CD11b (12-0118, eBioscience). Where indicated, cells were additionally stained with 1 ug/mL Hoechst 33342 (Life Technologies), and incubation was performed for 20 min at room temperature. Unbound antibodies were removed by washing twice with FACS Buffer; afterwards the pellet was resuspended in FACS Buffer supplemented with propidium iodide (PI) stain (00-6990, eBioscience) at a 1:20 dilution and transferred into FACS tubes for subsequent analysis on a FACS Canto II or LSR II (BD Bioscience) Flow Cytometer. Cells were analyzed with FlowJo (Tristar) and FACS Diva (BD Bioscience) software. PI positive cells, cell debris, and aggregates were gated out of the analysis.

May-Grünwald-Giemsa staining

Approximately 50,000 – 200,000 cells were harvested, washed once at 300 x g for 5 min, resuspended in 200 µl of FACS Buffer, and spun onto poly-L-lysine coated microscope slides with a

Shandon 4 (Thermo Scientific) cytocentrifuge at 300 rpm for 4 min. When visibly dry slides were transferred into May-Grünwald solution (Sigma-Aldrich) for 5 min, rinsed 4 times for each 30 s in deionized water, and transferred to Giemsa solution (Sigma-Aldrich) for 15 min. Slides were washed as described above, dry mounted with coverslips, and examined. All images shown were taken with AxioVision software (Zeiss) at 100X magnification.

Constructs

shRNA constructs targeting SH2B3 were used with the following sequences:
sh83 (CCGGCCTGACAACCTTTACACCTTTCTCGAGAAAGGTGTAAAGGTTGTCAGGTTTTTG) and sh84 (CCGGGCCTGACAACCTTTACACCTTCTCGAGAAGGTGTAAAGGTTGTCAGGCTTTTTG). Both constructs were in the pLKO.1-puro lentiviral vector. The lentiviral vectors pLKO-GFP and pLKO.1-puro shRNA targeting luciferase (shLuc) served as controls (the RNAi consortium of the Broad Institute of MIT and Harvard).

Cell Division Analysis

On day 5 of differentiation, an equal number of SH2B3-KD and control cells were labeled with PKH26 (PKH26GL, Sigma Aldrich) according to the manufacturer's protocol. In brief, cells were washed in IMDM (no addition of serum) and resuspended in diluent C. Immediately prior to staining, a 4 X 10^{-6} M PKH dye solution was prepared and rapidly added to the cell suspension at equal volumes. After 5 min of incubation, the staining reaction was stopped by addition of human AB serum, and unbound dye was removed by washing 3 times in phase 1 medium without EPO, SCF, or IL-3. Mean fluorescent intensities (MFI) of PKH26 were obtained immediately after labeling as well as on day 7 and 9 of differentiation on a FACSCanto II (BD Bioscience) flow cytometer. The number of cell divisions, x, was approximated as described previously: $x = (\log((MFI\ 0h)/(MFI\ yh)))/(\log 2)$ (Sankaran et al., 2012).

Eosin-5-maleimide (EMA) – Binding Test

2-4 x 10^6 cells were harvested washed with PBS and subsequently labeled with eosin-5-maleimide (Life Technologies), as described previously (King et al., 2000). In brief, pelleted mature RBCs were resuspended in 25 µl of 0.5 mg/ml EMA and incubated for 1 hour in the dark with careful intermittent mixing. Unbound dye was removed by washing 3 times in PBS + 3% FBS; afterwards, cells were incubated for another 10 min with APC-conjugated anti-human CD235a (eBioscience) at a 1:20 dilution at room temperature. After removal of unbound antibody as described above, pellets were resuspended in FACS-Buffer, and samples were analyzed on a FACSCanto II (BD Bioscience) flow cytometer, the emission for EMA was detected in the PE channel at 564 -606 nm.

Advia

Approximately 5 x 10^7 cells were harvested on day 17 of differentiation, spun, and resuspended in 300 µl PBS and analyzed on an Advia 2120i blood counter (Siemens Healthcare).

Gene Expression Analysis

Microarrays (GeneChip Human Gene 2.0 ST Arrays, Affymetrix) were performed on erythroblasts (day 7 of differentiation) for SH2B3 KD (sh83 and sh84) and control samples (shLuc). Raw files were processed and normalized using the RMA algorithm from the oligo package in R 3.2 (Carvalho and Irizarry, 2010). Differential expression analyses were conducted using limma (Ritchie et al., 2015). A heatmap is displayed for normalized expression values (Z-score) of genes that were significantly up-regulated in SH2B3 KD at a $\log_2$ fold change greater than 0.5 and a p-value threshold of 0.001. Gene set enrichment analysis (GSEA) was performed by comparing combined SH2B3 KD to control samples with standard options, except geneset permutation was used instead of sample permutation because there were n=3 samples for each group(Subramanian et al., 2005). We derived the erythroid differentiation signature gene set by identifying the top 200 genes that were expressed significantly higher in intermediate erythroblasts (CD71+/ CD235a+) compared to colony forming unit erythroid cells (CD71+/ CD235a-) (Merryweather-Clarke et al., 2011). All data has been deposited in the Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/) under accession GSE67219.

CRISPR/Cas9 genome editing in human ESCs

For *SH2B3* knockout in human embryonic stem cell line (hESCs; HUES-9), a plasmid coexpressing a codon-optimized *Streptococcus pyogenes* Cas9 and EGFP with the CAG promoter and a plasmid expressing the guide RNA (gRNA) with the human U6 promoter were co-transfected. To target *SH2B3* the following protospacer and protospacer-adjacent motif (PAM) were used:

5'- GCTCCAGCATCCAGGAGGTC-CGG -3'

hESCs were initially cultured in mTeSR1 supplemented with penicillin/streptomycin on Geltrex matrix coated tissue culture plates. Cells were dissociated using Accutase in the presence of ROCKi and subsequently 10 million single cells were electroporated with 25 ug of each plasmid in a single cuvette. Cells were replated and after 48-72 hours treated with Accutase, collected, and resuspended in PBS. EGFP-expressing cells were sorted by FACS (FACSAria II, BD Bioscience) and plated at 15,000 cells/plate in growth medium and allowed to recover for 7-10 days. Afterwards single colonies were manually picked and individually replated in wells of 96-well plates, grown near confluence for another 7 days, and treated with Accutase to create frozen stocks. PCR screening was employed to identify colonies with compound heterozygous or homozygous frameshift deletions in *SH2B3*. The lines BB5, DB3, and BF4 had homozygous or compound heterozygous frameshift mutations creating knockout of *SH2B3* and were used for the differentiation analyses. We used isogenic lines BA3, BA6, and BB1 with intact wild-type *SH2B3* alleles as controls for the differentiation analyses.

Maintenance and erythroid differentiation of hES cells

hESCs were maintained in six-well tissue culture plates containing 0.75-1.0 X 10^6 mouse embryonic fibroblasts (MEFs) in daily-exchanged HES medium containing DMEM supplemented with 20% KSR, 100 uL nonessential amino acids solution, 50 U/ml penicillin, 50 g/ml streptomycin, 2 mM glutamine, 1 mM sodium pyruvate, 0.075 mM sodium bicarbonate, and 0.1 mM beta-mercaptophenol at a maximum confluency of 90% at 37°C and 5% CO_2. For passaging, cells were treated with TrypLE, rinsed with medium, and after addition of 10 uM ROCKi to prevent apoptosis, were scraped into small

clumps, aspirated, mixed, and replated on MEF containing plates at an overall dilution of 1/7 in HES medium. For MEF depletion, hESCs were harvested as described above and replated onto Matrigel-coated dishes at a density of 1-1.5 X 10^5 per milliliter in HES medium containing 5 ng/mL bFGF and 10 uM ROCKi at 37°C, 5% CO_2, 5% O_2, and 90% N_2. Every 24 h cells were fed by replacing with fresh medium excluding ROCKi. Subsequent to reaching approximately 70% confluence 2-3 days following MEF depletion, differentiation to hematopoietic progenitor cells (HPCs) was performed as described(Mills et al., 2014). Briefly, cells were cultured on a daily feeding regime with sequential medium and cytokine changes depicted in the table below. Cells were kept in the Matrigel-coated dishes at 37°C, 5% CO_2, 5% O_2 and 90% N_2 for 8-9 days until appearing as bright round loosely adherent HPCs, upon which they were collected by carefully removing the supernatant. Subsequently an aliquot of cells was taken and flow cytometry was performed as described below to assess for the presence of CD41a and CD235a double-positive multipotent HPCs. For further differentiation into primitive erythoblasts, HPCs were spun at 335 X g and resuspended in erythroid expansion medium (EEM) containing SFD supplemented with 50 ug/ml ascorbic acid, 0.45 mM MTG, 50 ng/ml SCF, and 2 units/ml EPO. Medium was exchanged every 1-2 days. Cells were carefully counted with a hematocytometer and analyzed by flow cytometry at indicated time points.

Timetable of differentiation of PSCs into HPCs showing culture medium and supplements

Day	Medium*	BMP4***	VEGF	Wnt3a	BFGF	SCF	Flt3L	IL-6
0-1	RPMI	5	50	25				
2	RPMI	5	50		20			
3	SP34	5	50		20			
4-5	SP34		15		5			
6	IMDM**		50		50	50	5	
7-9	IMDM**		50		50	50	5	10

* All supplemented with 2 mM glutamine, 50 ug/mL ascorbic acid and 4X10^{-4} monothioglycerol

** Supplemented with 0.5% N2, 1% B27 without vitamin A, and 0.05% BSA

*** Cytokine concentrations in ng/mL

Statistical Analysis

All pairwise comparisons were assessed by an unpaired two-tailed Student's t-test. Results were considered significant if the P value was below 0.05.

Rare Variant Analyses

To test whether rare mutations in the *SH2B3* gene influence hematologic phenotypes (hemoglobin, hematocrit, etc.), we performed burden of rare variant analysis on 4,678 exome sequencing project (ESP) (Tennessen et al., 2012) samples from 7 cohorts (ARIC, CARDIA, CHS, FHS, JHS, MESA and WHI) retrieved from dbGaP (see dbGaP Sample Information/ Acknowledgements below).

Exome sequencing and variant calling for the NHLBI Exome Sequencing Project was performed as previously described (Tennessen et al., 2012). Variants were annotated using snpEff version 2.0.5 (Cingolani et al., 2012) and GATK VariantAnnotator version 2.5.2 (McKenna et al., 2010). Variants were filtered on global allele frequencies from the 1000 Genomes Project Phase 1 or ExAC database Version 0.3 (http://exac.broadinstitute.org/) less than 1% (Abecasis et al., 2010). Effects of missense variants were predicted using 5 algorithms: PolyPhen2 HumDiv and HumVar (Adzhubei et al., 2010), Mutation Taster (Schwarz et al., 2014), SIFT (Kumar et al., 2009), and LRT (Chun and Fay, 2009).

Association analysis of rare variants in the *SH2B3* gene were performed using a model based on the Combined Multivariate Collapsing test (Li and Leal, 2008), that groups the count of alleles by predicted damaging effect of the variant. Data from the studies was combined for analysis, each hematologic trait was regressed on age, gender and study variable, and inverse-normal transformed prior to testing. Linear regression was performed to test the association between variant burden grouping and hematologic phenotype performed using R (http://www.r-project.org/). We collapsed variants based on variant annotations from snpEff and computational predictions from PolyPhen2 HumDiv and HumVar, Mutation Taster, SIFT, and LRT. For our burden of rare variant association analysis, we use a MAF threshold of 1%. Furthermore, we use three different types of variant groupings when collapsing by

gene. These variant groups are: (1) disruptive and high impact (nonsense, indel frameshift, essential splice site) mutations only, (2) disruptive and high impact mutations plus a set of strict deleterious consisting of missense variants predicted to be damaging by all five prediction algorithms (Purcell et al., 2014), and (3) disruptive and high impact mutations plus a set of broad deleterious missense variants predicted to be damaging by at least one of the five algorithms. Since all known disease-associated somatic missense mutations in *SH2B3* occur in the PH or SH2 domains (Gery and Koeffler, 2013; Spolverini et al., 2013), we examined putative damaging variants and repeated analysis of strict and broad mutations in the PH and SH2 regions separately. Finally, analysis was performed combining disruptive and high impact mutations with strict and broad mutations from the PH and SH2 domains together. The analysis was done separately for individuals of European American and African American ethnicity. After initial analysis, we excluded the MESA cohort (n=404), since all the hemoglobin values appeared to be low and had a high ratio of hematocrit to hemoglobin where available, suggestive of aberrant values. Additionally, 3 subjects from the WHI cohort with hematocrit to hemoglobin ratios greater than 4 were excluded. Finally, we removed potential spurious outlier values for subjects with hemoglobin values greater than 33 g/dL (n=1), hematocrit value greater than 90 % (n=1), platelet value great than 1,000,000 per microliter (n=1), and white blood cell count greater 400,000 per microliter (n=1). All analyses were performed on the remaining individuals from the cohorts.

dbGaP Sample Information/ Acknowledgements

The datasets used for the analyses described in this manuscript were obtained from dbGaP at http://www.ncbi.nlm.nih.gov/sites/entrez?db=gap through dbGaP accession numbers: phs000281, phs000398, phs000399, phs000400, phs000401, phs000402, phs000403.

ARIC

The Atherosclerosis Risk in Communities Study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute contracts (HHSN268201100005C, HHSN268201100006C,

HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C). The authors thank the staff and participants of the ARIC study for their important contributions. This study is part of the NHLBI Grand Opportunity Exome Sequencing Project (GO-ESP). Funding for GO-ESP was provided by NHLBI grants RC2 HL103010 (HeartGO), RC2 HL102923 (LungGO) and RC2 HL102924 (WHISP). The exome sequencing was performed through NHLBI grants RC2 HL102925 (BroadGO) and RC2 HL102926 (SeattleGO). HeartGO gratefully acknowledges the following groups and individuals who provided biological samples or data for this study. DNA samples and phenotypic data were obtained from the following studies supported by the NHLBI: the Atherosclerosis Risk in Communities (ARIC) study, the Coronary Artery Risk Development in Young Adults (CARDIA) study, Cardiovascular Health Study (CHS), the Framingham Heart Study (FHS), the Jackson Heart Study (JHS) and the Multi-Ethnic Study of Atherosclerosis (MESA). This manuscript was not prepared in collaboration with investigators of the Atherosclerosis Risk in Communities Study and does not necessarily reflect the opinions or views of the ARIC, or NHLBI.

CARDIA

The Coronary Artery Risk Development in Young Adults Study (CARDIA) is conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with the University of Alabama at Birmingham (N01-HC95095 & N01-HC48047), University of Minnesota (N01-HC48048), Northwestern University (N01-HC48049), and Kaiser Foundation Research Institute (N01-HC48050). This study is part of the NHLBI Grand Opportunity Exome Sequencing Project (GO-ESP). Funding for GO-ESP was provided by NHLBI grants RC2 HL103010 (HeartGO), RC2 HL102923 (LungGO) and RC2 HL102924 (WHISP). The exome sequencing was performed through NHLBI grants RC2 HL102925 (BroadGO) and RC2 HL102926 (SeattleGO). HeartGO gratefully acknowledges the following groups and individuals who provided biological samples or data for this study. DNA samples and phenotypic data were obtained from the following studies supported by the NHLBI: the Atherosclerosis Risk in Communities (ARIC) study, the Coronary Artery Risk Development in Young

Adults (CARDIA) study, Cardiovascular Health Study (CHS), the Framingham Heart Study (FHS), the Jackson Heart Study (JHS) and the Multi-Ethnic Study of Atherosclerosis (MESA). This manuscript was not approved by CARDIA. The opinions and conclusions contained in this publication are solely those of the authors, and are not endorsed by CARDIA or the NHLBI and should not be assumed to reflect the opinions or conclusions of either.

CHS

The research reported in this article was supported by contract numbers N01-HC-85079, N01-HC-85080, N01-HC-85081, N01-HC-85082, N01-HC-85083, N01-HC-85084, N01-HC-85085, N01-HC-85086, N01-HC-35129, N01 HC-15103, N01 HC-55222, N01-HC-75150, N01-HC-45133, N01-HC-85239 and HHSN268201200036C; grant numbers U01 HL080295 from the National Heart, Lung, and Blood Institute and R01 AG-023629 from the National Institute on Aging, with additional contribution from the National Institute of Neurological Disorders and Stroke. A full list of principal CHS investigators and institutions can be found at http://www.chs-nhlbi.org/pi.htm. This study is part of the NHLBI Grand Opportunity Exome Sequencing Project (GO-ESP). Funding for GO-ESP was provided by NHLBI grants RC2 HL103010 (HeartGO), RC2 HL102923 (LungGO) and RC2 HL102924 (WHISP). The exome sequencing was performed through NHLBI grants RC2 HL102925 (BroadGO) and RC2 HL102926 (SeattleGO). HeartGO gratefully acknowledges the following groups and individuals who provided biological samples or data for this study. DNA samples and phenotypic data were obtained from the following studies supported by the NHLBI: the Atherosclerosis Risk in Communities (ARIC) study, the Coronary Artery Risk Development in Young Adults (CARDIA) study, Cardiovascular Health Study (CHS), the Framingham Heart Study (FHS), the Jackson Heart Study (JHS) and the Multi-Ethnic Study of Atherosclerosis (MESA). This manuscript was not prepared in collaboration with CHS investigators and does not necessarily reflect the opinions or views of CHS, or the NHLBI.

FHS

The Framingham Heart Study is conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with Boston University (Contract No. N01-HC-25195). The project utilized data from the Heart Cohorts Exome Sequencing Project (FHS), part of the NHLBI Grand Opportunity Exome Sequencing Project (GO-ESP). Funding for GO-ESP was provided by NHLBI grants RC2 HL103010 (HeartGO), RC2 HL102923 (LungGO) and RC2 HL102924 (WHISP). The exome sequencing was performed through NHLBI grants RC2 HL102925 (BroadGO) and RC2 HL102926 (SeattleGO). HeartGO gratefully acknowledges the following groups and individuals who provided biological samples or data for this study. DNA samples and phenotypic data were obtained from the following studies supported by the NHLBI: the Atherosclerosis Risk in Communities (ARIC) study, the Coronary Artery Risk Development in Young Adults (CARDIA) study, Cardiovascular Health Study (CHS), the Framingham Heart Study (FHS), the Jackson Heart Study (JHS) and the Multi-Ethnic Study of Atherosclerosis (MESA). This manuscript was not prepared in collaboration with investigators of the Framingham Heart Study and does not necessarily reflect the opinions or views of the Framingham Heart Study, Boston University, or NHLBI.

JHS

The Jackson Heart Study is supported and conducted in collaboration with Jackson State University (N01-HC-95170), University of Mississippi Medical Center (N01-HC-95171), and Touglaoo College (N01-HC-95172) contracts from the National Heart, Lung, and Blood Institute (NHLBI) and the National Institute for Minority Health and Health Disparities (NIMHD), with additional support from the National Institute on Biomedical Imaging and Bioengineering (NIBIB). This study is part of the NHLBI Grand Opportunity Exome Sequencing Project (GO-ESP). Funding for GO-ESP was provided by NHLBI grants RC2 HL103010 (HeartGO), RC2 HL102923 (LungGO) and RC2 HL102924 (WHISP). The exome sequencing was performed through NHLBI grants RC2 HL102925 (BroadGO) and RC2 HL102926 (SeattleGO). HeartGO gratefully acknowledges the following groups and individuals who provided biological samples or data for this study. DNA samples and phenotypic data were obtained from the following studies supported by the NHLBI: the Atherosclerosis Risk in Communities (ARIC)

study, the Coronary Artery Risk Development in Young Adults (CARDIA) study, Cardiovascular Health Study (CHS), the Framingham Heart Study (FHS), the Jackson Heart Study (JHS) and the Multi-Ethnic Study of Atherosclerosis (MESA). This manuscript was not prepared in collaboration with JHS investigators and does not necessarily reflect the opinions or views of JHS, or the NHLBI.

MESA

MESA and the MESA SHARe project are conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support for MESA is provided by contracts N01-HC-95159, N01-HC-95160, N01-HC-95161, N01-HC-95162, N01-HC-95163, N01-HC-95164, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169 and CTSA UL1-RR-024156. This study is part of the NHLBI Grand Opportunity Exome Sequencing Project (GO-ESP). Funding for GO-ESP was provided by NHLBI grants RC2 HL103010 (HeartGO), RC2 HL102923 (LungGO) and RC2 HL102924 (WHISP). The exome sequencing was performed through NHLBI grants RC2 HL102925 (BroadGO) and RC2 HL102926 (SeattleGO). HeartGO gratefully acknowledges the following groups and individuals who provided biological samples or data for this study. DNA samples and phenotypic data were obtained from the following studies supported by the NHLBI: the Atherosclerosis Risk in Communities (ARIC) study, the Coronary Artery Risk Development in Young Adults (CARDIA) study, Cardiovascular Health Study (CHS), the Framingham Heart Study (FHS), the Jackson Heart Study (JHS) and the Multi-Ethnic Study of Atherosclerosis (MESA). This manuscript was not prepared in collaboration with MESA investigators and does not necessarily reflect the opinions or views of MESA, or the NHLBI.

WHI

The WHI program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services through contracts N01WH22110, 24152, 32100-2, 32105-6, 32108-9, 32111-13, 32115, 32118-32119, 32122, 42107-26, 42129-32, and 44221. The Women's Health Initiative Sequencing Project (WHISP) is funded by Grant Number RC2

HL102924. This study is part of the NHLBI Grand Opportunity Exome Sequencing Project (GO-ESP). Funding for GO-ESP was provided by NHLBI grants RC2 HL103010 (HeartGO), RC2 HL102923 (LungGO) and RC2 HL102924 (WHISP). The exome sequencing was performed through NHLBI grants RC2 HL102925 (BroadGO) and RC2 HL102926 (SeattleGO). This manuscript was not prepared in collaboration with investigators of the WHI, has not been reviewed and/or approved by the Women's Health Initiative (WHI), and does not necessarily reflect the opinions of the WHI investigators or the NHLBI.

Supplemental References

Abecasis, G.R., Altshuler, D., Auton, A., Brooks, L.D., Durbin, R.M., Gibbs, R.A., Hurles, M.E., and McVean, G.A. (2010). A map of human genome variation from population-scale sequencing. Nature *467*, 1061-1073.

Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., and Sunyaev, S.R. (2010). A method and server for predicting damaging missense mutations. Nat Methods *7*, 248-249.

Carvalho, B.S., and Irizarry, R.A. (2010). A framework for oligonucleotide microarray preprocessing. Bioinformatics *26*, 2363-2367.

Chun, S., and Fay, J.C. (2009). Identification of deleterious mutations within three human genomes. Genome Res *19*, 1553-1561.

Cingolani, P., Platts, A., Wang le, L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and Ruden, D.M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin) *6*, 80-92.

Gery, S., and Koeffler, H.P. (2013). Role of the adaptor protein LNK in normal and malignant hematopoiesis. Oncogene *32*, 3111-3118.

King, M.J., Behrens, J., Rogers, C., Flynn, C., Greenwood, D., and Chambers, K. (2000). Rapid flow cytometric test for the diagnosis of membrane cytoskeleton-associated haemolytic anaemia. Br J Haematol *111*, 924-933.

Kumar, P., Henikoff, S., and Ng, P.C. (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc *4*, 1073-1081.

Lee, H.Y., Gao, X., Barrasa, M.I., Li, H., Elmes, R.R., Peters, L.L., and Lodish, H.F. (2015). PPAR-alpha and glucocorticoid receptor synergize to promote erythroid progenitor self-renewal. Nature.

Li, B., and Leal, S.M. (2008). Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. Am J Hum Genet *83*, 311-321.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M.*, et al.* (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res *20*, 1297-1303.

Merryweather-Clarke, A.T., Atzberger, A., Soneji, S., Gray, N., Clark, K., Waugh, C., McGowan, S.J., Taylor, S., Nandi, A.K., Wood, W.G.*, et al.* (2011). Global gene expression analysis of human erythroid progenitors. Blood *117*, e96-108.

Mills, J.A., Paluru, P., Weiss, M.J., Gadue, P., and French, D.L. (2014). Hematopoietic differentiation of pluripotent stem cells in culture. Methods Mol Biol *1185*, 181-194.

Purcell, S.M., Moran, J.L., Fromer, M., Ruderfer, D., Solovieff, N., Roussos, P., O'Dushlaine, C., Chambert, K., Bergen, S.E., Kahler, A.*, et al.* (2014). A polygenic burden of rare disruptive mutations in schizophrenia. Nature *506*, 185-190.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic acids research.

Sankaran, V.G., Ludwig, L.S., Sicinska, E., Xu, J., Bauer, D.E., Eng, J.C., Patterson, H.C., Metcalf, R.A., Natkunam, Y., Orkin, S.H.*, et al.* (2012). Cyclin D3 coordinates the cell cycle during differentiation to regulate erythrocyte size and number. Genes Dev *26*, 2075-2087.

Sankaran, V.G., Menne, T.F., Scepanovic, D., Vergilio, J.A., Ji, P., Kim, J., Thiru, P., Orkin, S.H., Lander, E.S., and Lodish, H.F. (2011). MicroRNA-15a and -16-1 act via MYB to elevate fetal hemoglobin expression in human trisomy 13. Proc Natl Acad Sci U S A *108*, 1519-1524.

Sankaran, V.G., Menne, T.F., Xu, J., Akie, T.E., Lettre, G., Van Handel, B., Mikkola, H.K., Hirschhorn, J.N., Cantor, A.B., and Orkin, S.H. (2008). Human fetal hemoglobin expression is regulated by the developmental stage-specific repressor BCL11A. Science *322*, 1839-1842.

Schwarz, J.M., Cooper, D.N., Schuelke, M., and Seelow, D. (2014). MutationTaster2: mutation prediction for the deep-sequencing age. Nat Methods *11*, 361-362.

Spolverini, A., Pieri, L., Guglielmelli, P., Pancrazzi, A., Fanelli, T., Paoli, C., Bosi, A., Nichele, I., Ruggeri, M., and Vannucchi, A.M. (2013). Infrequent occurrence of mutations in the PH domain of LNK in patients with JAK2 mutation-negative 'idiopathic' erythrocytosis. Haematologica *98*, e101-102.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S.*, et al.* (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A *102*, 15545-15550.

Tennessen, J.A., Bigham, A.W., O'Connor, T.D., Fu, W., Kenny, E.E., Gravel, S., McGee, S., Do, R., Liu, X., Jun, G.*, et al.* (2012). Evolution and functional impact of rare coding variation from deep sequencing of human exomes. Science *337*, 64-69.

10. Curriculum vitae

Mein Lebenslauf wird aus datenschutzrechtlichen Gründen in der elektronischen Version meiner Arbeit nicht veröffentlicht.

Mein Lebenslauf wird aus datenschutzrechtlichen Gründen in der elektronischen Version meiner Arbeit nicht veröffentlicht.

11. Complete publication list

Publications presented within the scope of this thesis

1. **Ludwig LS***, Cho H*, Wakabayashi A, Eng JC, Ulirsch JC, Fleming MD, Lodish HF, Sankaran VG. Genome-wide association study follow-up identifies cyclin A2 as a regulator of the transition through cytokinesis during terminal erythropoiesis. **American Journal of Hematology**. 2015 May;90(5):386-91. PMID: 25615569.
   * These authors contributed equally to this work
2. Sankaran VG, Ulirsch JC, Tchaikovskii V, **Ludwig LS**, Wakabayashi A, Kadirvel S, Lindsley RC, Bejar R, Shi J, Lovitch SB, Bishop DF, Steensma DP. Macrocytic Anemia and Dyserythropoiesis Resulting From an X-Linked Dominant *ALAS2* Mutation. **Journal of Clinical Investigation**. 2015 Apr 1;125(4):1665-9. PMID: 25705881
3. Giani FG, Fiorini C, Wakabayashi A, **Ludwig LS**, Jobaliya CD, Regan SN, Ulirsch JC, Liang G, Steinberg-Shemer O, Esko T, Hirschhorn JN, Tong W, Brugnara C, Weiss MJ, Zon LI, Chou ST, French DL, Musunuru K, Sankaran VG. Using human genetic variation to improve red blood cell production from stem cells. **Cell Stem Cell**. 2016, Jan 7;18(1):1-6. PMID: 26607381

Additional publications

4. **Ludwig LS**, Gazda HT, Eng JC, Eichhorn SW, Thiru P, Ghazvinian R, George TI, Gotlib JR, Beggs AH, Sieff CA, Lodish HF, Lander ES, Sankaran VG. Altered translation of GATA1 in Diamond-Blackfan anemia. **Nature Medicine**. 2014 Jul;20(7):748-53. PMID: 24952648
5. Sankaran VG*, **Ludwig LS***, Sicinska E, Xu J, Bauer DE, Eng JC, Patterson HC, Metcalf RA, Natkunam Y, Orkin SH, Sicinski P, Lander ES, Lodish HF. Cyclin D3 coordinates the cell cycle during differentiation to regulate erythrocyte size and number. **Genes and Development**. 2012 Sep 15; 26(18):2075-87. PMID: 22929040.
   * These authors contributed equally to this work
6. Polansky JK, Schreiber L, Thelemann C, **Ludwig L**, Krüger M, Baumgrass R, Cording S, Floess S, Hamann A, Huehn J. Methylation matters: binding of Ets-1 to the demethylated Foxp3 gene contributes to the stabilization of Foxp3 expression in regulatory T cells. **Journal of Molecular Medicine**. 2010 Oct;88(10):1029-40. PMID: 20574810

In preparation

Wakabayashi A, Ulirsch JC, **Ludwig LS**, Yasuda M, Choudhuri A, McDonel P, Zon LI, Sankaran VG. Insight into GATA1 transcriptional activity through interrogation of *cis*-elements disrupted in human erythroid disorders. (*under review*)

12. Acknowledgements

I am deeply grateful for having had the opportunity to work with Vijay Sankaran, Harvey Lodish and Petra Knaus in an absolutely fascinating research environment. I am incredibly thankful for all their inspiration, unlimited enthusiasm, guidance, mentorship and support in and outside of the laboratory over the past years.

I would like to thank my friends and labmates, my mentors and students all over the world for all their time, help and fun we shared, for their motivation, diligence, dedication and commitment that have made my time an unforgettable experience. I thank Felix Giani, Rajiv Khajuria, Jacob Ulirsch, Aoi Wakabayashi, Bill Wong, Jiahai Shi, Jennifer Eng, Anjali Nirmalan, Justina Cho, Novalia Pishesha, Marko Knoll, Heide Christine Patterson, Johan Flygare, Hojun Li, Claire Mitrokostas, Katrin Heindl, Prathapan Thiru and Fatima Cabral.

I have been very fortunate and incredibly privileged for having had the chance to work and spend time with all of you.

Once again, and I do not seem to ever run short for praising and thanking my parents, Young-Cha Bae and Klaus-Peter Ludwig, for all their support and belief in my abilities and who have ultimately enabled all of my small and huge achievements.