Altered Translation of GATA1 in Diamond-Blackfan Anemia and the Role of Cyclin D3 in Erythropoiesis

Inaugural-Dissertation

to obtain the academic degree
Doctor rerum naturalium (Dr. rer. nat.)
submitted to the Department of Biology, Chemistry and Pharmacy
of Freie Universität Berlin

by
Dipl.-Biochem. Leif Si-Hun Ludwig
born in Berlin, Germany

2015

Acknowledgements

I am deeply grateful to Harvey Lodish for inviting me to his laboratory and giving me the opportunity to work in one of the most vibrant research environments. He would always question the very fundamentals of one's project and it certainly doesn't take long to understand how Harvey trained so many great scientists over the years. With his track record and scientific as well as personal input he was a constant source of motivation and inspiration.

I am incredibly thankful to Vijay Sankaran, who has been one of the most dazzling and determined clinician-scientists I have had the chance to work with. I don't think I have ever learned as much in such a short time and he has allowed me to go above and beyond of what I thought is possible. His constant and unlimited enthusiasm for science and medicine has made the past years an unforgettable experience.

None of my work abroad would have been possible without the commitment of Petra Knaus. I am very thankful for all her advice and support and the opportunities to present my work in Berlin over the last years.

I would like to thank Jennifer Eng for helping me with numerous experiments, always pointing out the best food in town and together with Anjali Nirmalan for giving my friends and me shelter when needed. Special thanks to Bill Wong, who took me under his wing the first couple of months, and Jiahai Shi for being the best bay mates one could ask for!

I co-supervised a number of undergraduate students from MIT during my stay at the Whitehead Institute and more than once I have been surprised by their work ethic and brilliance. One certainly learns a lot from teaching and mentoring. Thanks to Paula Trepman, Katherine Luo, Evelyn Wang, Cynthia Chen and super special thanks to Justina Cho for all her dedication and commitment!

Thank you Novalia Pishesha for just being who you are and teaching me what real life fairy tales can look like. I have to especially thank Marko Knoll. I am not sure why, but he insisted! I thank Heide Christine Patterson for help with everything related to histology and pathology and for discussions from a German perspective. I would also like to thank all current and former members of the Lodish laboratory for scientific discussions, motivation and a great

Table of contents

1. Introduction

1.1 Hematopoiesis and Erythropoiesis

Hematopoiesis describes the process of blood cell formation and has served as a paradigm for the studies and understanding of a complex developmental system of mammalian biology (Orkin and Zon 2008; Doulatov et al. 2012). At the apex of the hematopoietic hierarchy lies the hematopoietic stem cell (HSC) that ultimately forms all the cell types of the mature blood (Figure 1) (Sankaran and Weiss 2015). Its capability to self-renew makes the blood one of the most highly regenerative tissues, and this quality is widely exploited clinically in the form of hematopoietic stem cell transplantation to reconstitute the hematopoietic system after high-dose chemotherapy or to cure genetic disorders of hematopoiesis (Doulatov et al. 2012; Fagioli et al. 2014).

One of the demanding duties of hematopoiesis lies in the sufficient production of red blood cells (erythrocytes). The process of erythropoiesis describes the proliferation and differentiation of hematopoietic stem and progenitor cells toward the erythroid lineage (Hattangadi et al. 2011; Sankaran and Weiss 2015). Different developmental stages of erythropoiesis can be characterized (McGrath and Palis 2008; Palis 2014). Primitive erythroid cells derive from the yolk sac and circulate transiently in the early embryo (Palis et al. 2010). Definitive erythropoiesis occurs in the fetal liver and the adult bone marrow. During fetal development, definitive erythropoiesis expands the red blood cell compartment to meet the growing demands of the developing organism. In contrast, adult definitive erythropoiesis is largely characterized by homeostatic maintenance of a steady state of erythrocyte numbers in the peripheral blood (McGrath and Palis 2008; Palis 2014). In humans, two million red blood cells are produced per second, making erythrocytes the most abundant cell type in the human body (Palis 2014). In the peripheral blood, erythrocytes comprise 40-50 % of the blood volume and are the key transporters of oxygen and carbon dioxide for cellular respiration.
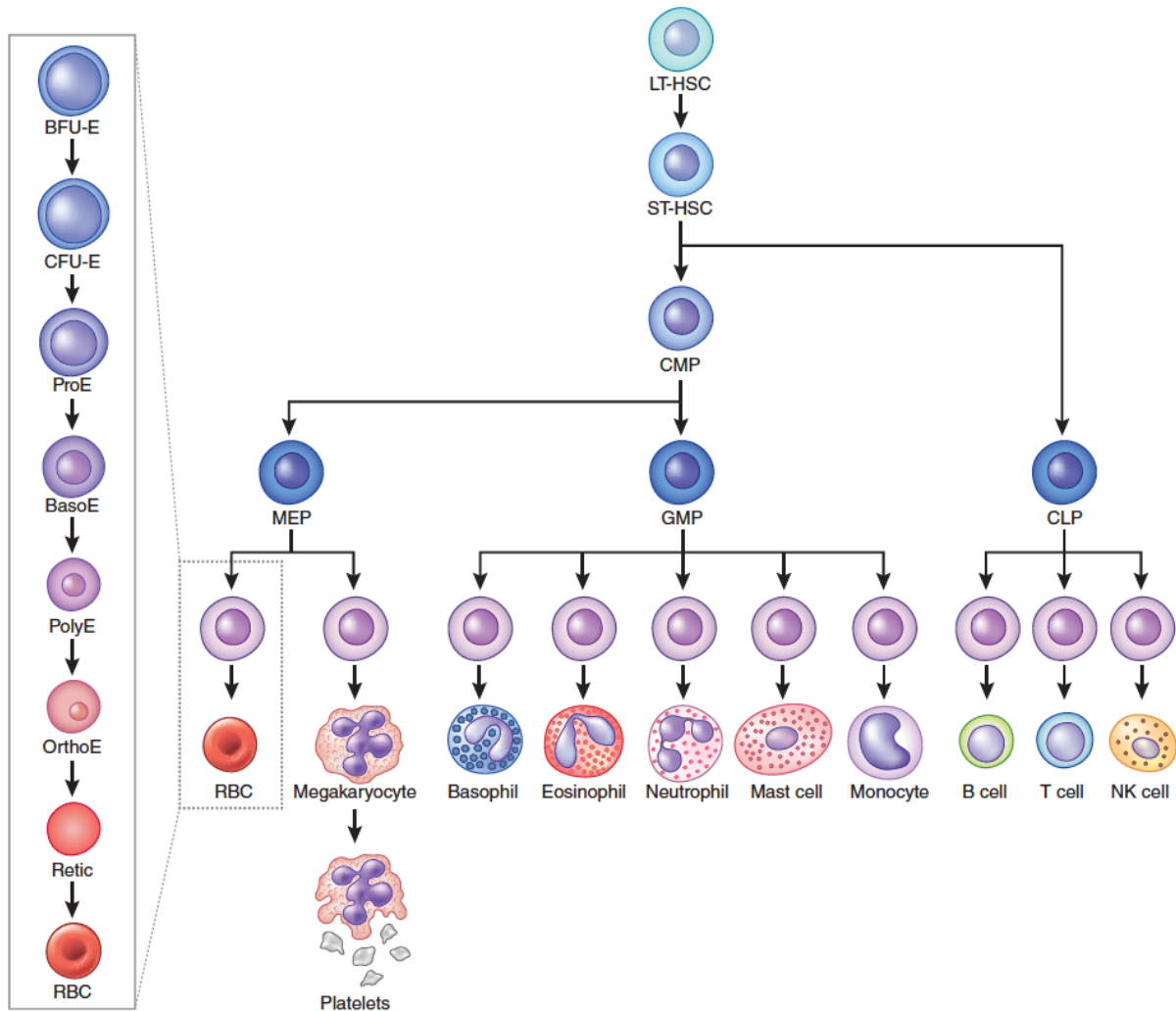
Figure 1 A model for hematopoiesis. The long-term hematopoietic stem cell (LT-HSC) gives rise to short-term hematopoietic stem cells (ST-HSCs) that then give rise to the multipotent common myeloid progenitor (CMP) and common lymphoid progenitor (CLP). The CMP then gives rise to megakaryocyte progenitors (MEP) and granulocyte-macrophage progenitors (GMP). The maturation of lineage committed erythroid progenitors is shown on the left side. The earliest progenitor, burst-forming unit erythroid (BFU-E), gives rise to the colony-forming unit erythroid (CFU-E). The CFU-E differentiates into morphologically distinct precursors that undergo progressive maturation: proerythroblasts (ProE), basophilic erythroblasts (BasoE), polychromatic erythroblasts (PolyE), and orthochromatic erythroblasts (OrthoE). OrthoE enucleate to produce reticulocytes (Retic) that are released into circulation and mature further into RBCs. Figure from Sankaran and Weiss 2015.

The earliest committed progenitor derives from a bipotential megakaryocyte and erythroid progenitor cell (MEP) and is the burst forming unit erythroid (BFU-E), which gives rise to the colony-forming unit erythroid (CFU-E) (Figure 1). Both cell types have traditionally been characterized by their ability to form colonies in semisolid culture *in vitro* (Palis 2014). The CFU-E further matures into distinct precursors, which can be distinguished by their morphology (Figure 1, left) (Palis 2014; Sankaran and Weiss 2015). These cells progressively differentiate and form the orthochromatic erythroblasts, which enucleate to produce reticulocytes that are released into circulation to ultimately mature into erythrocytes.

Fundamental changes in gene expression govern the course of differentiation in erythropoiesis, leading to characteristic changes including a progressive reduction in proliferative capacity, a decrease in cell size, upregulation of heme metabolism and iron absorption and utilization that enable the massive production of hemoglobin as the major transporter of blood oxygen (Hattangadi et al. 2011; An et al. 2014; Palis 2014; Yien et al. 2014).

A number of cytokines regulate erythropoiesis and may allow for a substantial increase in red blood cell production (i.e. in response to acute blood loss). The major cytokine controlling erythroid output is erythropoietin (EPO) which is regulated at the transcriptional level by hypoxia in a specialized population of cells in the kidney (Jelkmann 2013). EPO binds to the EPO receptor (EPOR) on erythroid progenitor cells promoting erythroid differentiation, proliferation and survival (D'Andrea et al. 1989; Socolovsky et al. 1999; Socolovsky et al. 2001; Palis 2014). Both $Epo^{-/-}$ and $EpoR^{-/-}$ mice die around embryonic day 13 illustrating the essential role of EPO signaling in erythropoiesis (Wu et al. 1995). In addition to EPO a number of cytokines and hormones, including stem cell factor (SCF), interleukin 3 (IL-3), interleukin 6, glucocorticoids and growth differentiation factor 11 are known to act especially at the early stages of erythropoiesis (Hattangadi et al. 2011; Dussiot et al. 2014; Suragani et al. 2014).

1.1.1 Transcriptional Regulators of Erythropoiesis and GATA1

Genes encoding transcription factors play a vital role in the regulation of erythropoiesis. These include *STAT5* downstream of the erythropoietin receptor and a number of other lineage restricted factors including *GATA1*, *KLF1*, *FOG1*, *TAL1*, *GFI1B*, *LMO2*, *LDB1*, *MYB*, *SOX6* and *BCL11A* (Orkin and Zon 2008; Kerenyi and Orkin 2010; Hattangadi et al. 2011; Palis 2014). The hematopoietic factor GATA1 is one of the best studied transcriptional regulators and has been discovered as the first member of the GATA family to recognize the DNA consensus sequence (A/T)GATA(A/G) (Wall et al. 1988). GATA1 is composed of three functional domains: the N-terminal transactivation domain and the N-terminal and C-terminal zinc fingers responsible for DNA binding (Ferreira et al. 2005). The *GATA1* gene is encoded on the X chromosome and consists of 6 exons. The mRNA transcript can be alternatively spliced resulting in the presence of a long and short isoform (Figure 2) (Sankaran et al. 2012a). The long transcript is translated to a full-length form of GATA1

(GATA1 FL) and a short form (GATA1s), resulting from an alternative translation initiation site in exon 3 (Calligaris et al. 1995). The short transcript lacks exon 2 and only produces GATA1s, which lacks amino acids 1-83 encoding the N-terminal transactivation domain (TD) (Ferreira et al. 2005).
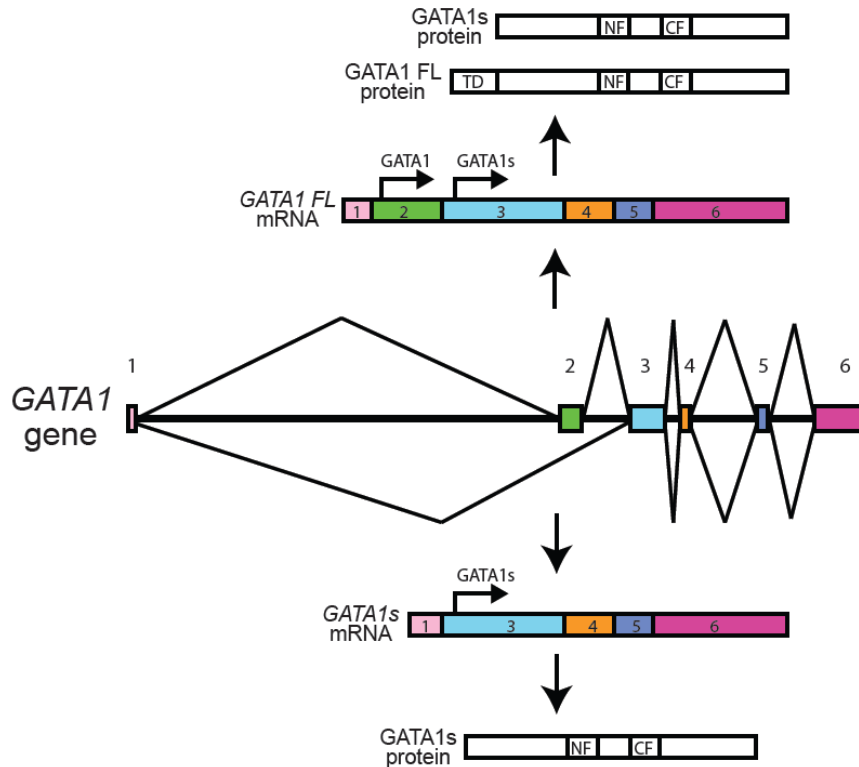
Figure 2 Normal splicing pattern of *GATA1* in humans and the resulting mRNA and protein products. For the protein products, the transactivation domain (TD), N-terminal zinc finger (NF), and C-terminal zinc finger (CF) are shown. From Sankaran et al. 2012.

GATA1 is essential for erythropoiesis as its disruption leads to arrested development at the proerythroblast stage (Pevny et al. 1995; Fujiwara et al. 1996). A variety of the aforementioned transcriptional regulators are known to complex with GATA1, and the targeted disruption of these factors leads to pronounced defects in erythropoiesis (Warren et al. 1994; Robb et al. 1995; Shivdasani et al. 1995; Tsang et al. 1998; Ferreira et al. 2005; Li et al. 2010; Stonestrom et al. 2015). The fundamental function of GATA1 in governing erythroid gene expression is further ensured by mitotic bookmarking to preserve the transcriptional profile through cell divisions and differentiation. While GATA1 coregulators FOG1 and TAL1 dissociate from mitotic chromatin, GATA1 binding is retained at a subset of target genes through mitosis, ensuring their rapid reactivation and restoration of cell-specific transcription in the daughter cells (Kadauke et al. 2012; Kadauke and Blobel 2013). Moreover, GATA1 may undergo a number of posttranslational modifications, including

acetylation, phosphorylation and sumoylation, which modulate its activity (Ferreira et al. 2005; Yu et al. 2010). During human erythropoiesis, HSP70 has been suggested to stabilize GATA1, preventing its caspase-3-mediated degradation (De Maria et al. 1999; Ribeil et al. 2007; Arlet et al. 2014).

GATA1 is also expressed in the testes and required in other hematopoietic cell types including megakaryocytes, mast cells, eosinophils and basophils (Shivdasani et al. 1997; Cantor et al. 2008; Nei et al. 2013). While most of the studies of GATA1 have been conducted in mice and other model systems, its role appears less well understood in human hematopoiesis. Interestingly, somatic mutations preventing synthesis of full length GATA1, but not synthesis of GATA1 short, have been identified in virtually all cases of the megakaryoblastic leukemia of Down syndrome (Wechsler et al. 2002; Greene et al. 2003; Yoshida et al. 2013). However, congenital mutations in the human gene *GATA1* appear to predominantly present with thrombocytopenia (low platelet counts) and/or anemia (Sankaran et al. 2012a; Campbell et al. 2013; Crispino and Weiss 2014).

1.2 Anemia and Mendelian Genetics

In healthy individuals the lifespan of red blood cells is 120 days and steady-state erythropoiesis maintains stable numbers of red blood cells and hemoglobin levels (Hattangadi et al. 2011). Anemia is defined as a decrease in the amount of red blood cells or hemoglobin in the peripheral blood and may ensue when homeostatic mechanisms of erythropoiesis are temporarily or permanently unable to maintain normal hemoglobin levels (Kassebaum et al. 2014; Sankaran and Weiss 2015). Much of the global burden of anemia is attributable to nutritional deficiencies such as lack of iron, folic acid, or vitamin B12 that are essential for hemoglobin production and proliferation of erythroid cells. Other causal factors include infections, such as malaria or HIV/AIDS, chronic illnesses, such as chronic kidney disease that results in a lack of EPO production, or hereditary lesions in the human genome (Calis et al. 2008; Balarajan et al. 2011; Jelkmann 2013; Sankaran and Weiss 2015).

The discovery of an abnormal electrophoretic mobility of sickle cell hemoglobin in sickle cell anemia is an example of one of the first human diseases where an underlying molecular defect was recognized (Pauling et al. 1949; Bunn 1997). The genetic defect in sickle cell anemia has subsequently been mapped to the beta-globin gene and disorders

caused by mutations in a single locus or gene are termed Mendelian or monogenic diseases. Over the last few decades, advances in genetic technology have lead to a significant increase in the number of genetic lesions that have been identified as causal or associated with numerous human diseases, including anemia. The Online Mendelian Inheritance in Man database (OMIM) contains information on all known Mendelian disorders and now contains over 5,400 phenotypes with a known molecular basis. The number of total entries has increased from 1,500 in 1966 to about 7,000 in 1994 and has since tripled to over 22,900 in May 2015 (http://www.omim.org/) (Pearson et al. 1994). A milestone in the discovery of genetic lesions has come from the improvement in DNA sequencing techniques. Since the initial sequencing and analysis of the human genome, technical refinement and a significant reduction in costs have made DNA sequencing approaches a robust strategy to detect genetic causes of human disease (Lander et al. 2001; Lander 2011).

One approach, known as whole-exome sequencing (WES), is a technique that allows for the targeted sequencing of all the protein-coding genes or exons in a genome (termed the exome) (Gnirke et al. 2009). The exome represents about 1 % of the human genome, but ~86 % of mutations listed in the human genome mutation database can be found within the coding region, making WES a cost-effective approach to search for sequence variants in patients with unknown molecular etiology (http://www.hgmd.cf.ac.uk/ac/index.php) (Cooper et al. 2010). In 2009, a proof-of concept study showed that WES can successfully identify candidate genes for Mendelian disorders and it has first been successful in identifying the underlying genetic defect in a Mendelian disorder of unknown cause in 2010 (Ng et al. 2009; Ng et al. 2010). However, interpreting the results of WES in the context of Mendelian disease can be challenging (Yang et al. 2013; Casanova et al. 2014; Yang et al. 2014). Each patient carries numerous unique or singleton variants, many of which are predicted to result in loss of function for various proteins. If no causal genetic variant has yet been identified for a specific Mendelian disease, WES of a single patient may not be sufficient to identify the causal gene. Usually, WES is performed on a cohort of patients with similar disorders and compared to an otherwise healthy control group. Identified genes that are repeatedly mutated or may act in synergistic pathways, potentially causing a shared clinical phenotype, may then be subjected to additional functional analyses.

With respect to erythropoiesis, a significant number of genetic lesions have been identified as causal in numerous anemias (Balarajan et al. 2011; Sankaran and Weiss 2015).

These include mutations that affect a variety of erythrocyte components, such as the enzyme glucose-6-phosphate dehydrogenase, or mutations in structural proteins of the red blood cell cytoskeleton in hereditary spherocytosis, both of which result in non-immune hemolytic anemia (van Wijk and van Solinge 2005; Perrotta et al. 2008). Disorders of the hemoglobin genes are among the most common causes of anemia and include sickle cell disease and thalassemia (Modell and Darlison 2008; Piel and Weatherall 2014). Estimates suggest that at least 5 % of the global population are carriers of a significant hemoglobin variant (Balarajan et al. 2011). Some variants are particularly common, as carriers are less likely to die from malaria (Modell and Darlison 2008). A variety of rare mutations have also been identified in the aforementioned erythroid-essential gene *GATA1* leading to hereditary forms of anemia (Campbell et al. 2013; Crispino and Weiss 2014). Likewise, mutations in GATA1 binding sites may result in erythroid phenotypes and further underline the importance of this factor in human erythropoiesis (Tournamille et al. 1995; Solis et al. 2001; Campagna et al. 2014; Kaneko et al. 2014). Recently, mutations in *GATA1* have been identified in patients with Diamond-Blackfan anemia (DBA), which have not been previously linked to this disease (Sankaran et al. 2012a; Weiss et al. 2012; Klar et al. 2014; Parrella et al. 2014).

1.3 Mysterious Ribosomopathies and Diamond-Blackfan Anemia

DBA is a member of a group of inherited diseases collectively known as the ribosomopathies, including Schwachman-Diamond syndrome, dyskeratosis congenita, cartilage hair hypoplasia and Treacher Collins syndrome (Narla and Ebert 2010; Ruggero and Shimamura 2014; De Keersmaecker et al. 2015). An acquired ribosomopathy, the chromosome 5q deletion (5q-) syndrome leads to hematologic defects that may in part be due to the loss of one copy of *RPS14*, a gene encoding for a protein of the small ribosomal subunit (Ebert et al. 2008). It presents as a subtype of the myelodysplastic syndrome and shares clinical features with DBA. The unifying characteristic of all ribosomopathies is that each associated genetic defect has been suggested to affect aspects of ribosome biogenesis and function.

DBA was first reported by Josephs in 1936 and more comprehensively described by Louis Diamond and Kenneth Blackfan in 1938 as a potentially lethal congenital hypoplastic anemia characterized by impaired erythropoiesis (Josephs 1936; Diamond and Blackfan 1938). The first genetic lesion was identified in 1999 in a child with a reciprocal X;19

chromosomal translocation (Gustavsson et al. 1997; Draptchinskaia et al. 1999). The breakpoint occurred in the gene encoding ribosomal protein S19 (RPS19). Subsequently, heterozygous-loss of function mutations in *RPS19* were identified in approximately 25 % of patients with DBA. The single intact copy of the normal *RPS19* gene appeared unable to compensate for the loss of the mutant allele, resulting in a state of haploinsufficiency (Gazda et al. 2004). Sixteen different ribosomal protein genes with heterozygous loss of function mutations inherited in an autosomal dominant fashion have so far been associated to DBA; these include mutations in ribosomal proteins (RP) of the small 40S (*RPS7, RPS10, RPS17, RPS19, RPS24, RPS26, RPS27, RPS28, RPS29*) and the large 60S subunit (*RPL5, RPL11, RPL15, RPL26, RPL27, RPL31, RPL35A*) of the ribosome. Together, these mutations account for 50-70 % of the identified genetic lesions in patients with DBA (Weiss et al. 2012; Farrar et al. 2014; Gripp et al. 2014; Wang et al. 2015).

DBA is usually recognized within the first year of life and the diagnosis relies primarily on clinical findings but may be supported by the identification of potentially pathogenic RP mutations (Vlachos et al. 2008). The anemia is characterized by macrocytic (large) erythrocytes and a lack of reticulocytes (reticulocytopenia). The low frequency of these immature red blood cells is indicative of impaired erythropoiesis, and while examination of the bone marrow shows normal cellularity, a decrease or complete absence of red blood cell precursors is observed (Vlachos et al. 2008). Usually no other significant cytopenias (reduction in the number of blood cells) are observed in other hematopoietic lineages, although variable neutrophil and/or platelet counts, including thrombocytosis (high platelet counts) have been reported (Willig et al. 1999; Vlachos et al. 2008).

DBA patients may present with small stature and a variety of malformations are reported in up to 35-45 % of patients, including anomalies of the head, upper limbs, heart and genitourinary system (Ball et al. 1996; Ramenghi et al. 1999; Lipton et al. 2006; Vlachos et al. 2008). Cleft palate and thumb anomalies have been more frequently associated with mutations in *RPL5* and *RPL11* (Gazda et al. 2008). There is a wide range with respect to the congenital defects and the hematological manifestations. Identical mutations may present with various clinical phenotypes and both mild and severe forms may present within the same family, indicating variable penetrance and expressivity (Gripp et al. 2001; Vlachos et al. 2008).

1.3.1 Treatment of Diamond-Blackfan Anemia

The major drugs in the treatment of DBA are corticosteroids and up to 80 % of patients respond to an initial course of treatment (Vlachos et al. 2008). At first it remained unclear how steroid hormones improve the anemia, and DBA was in fact suspected to be an autoimmune disorder, but no evidence for suppression by patient lymphocytes or other immune cells could be identified (Nathan et al. 1978a; Nathan et al. 1978b). Corticosteroids do not appear to alter expression of ribosomal protein genes (Ebert et al. 2005), but may rather promote self-renewal of BFU-E progenitor cells (Flygare et al. 2011; Zhang et al. 2013a). The subsequent increase in the total numbers of CFU-E and erythroid cells may thus ameliorate the anemia. However, a substantial fraction of patients do not respond, become refractory, or experience significant side effects due to corticosteroids (Vlachos et al. 2008). Moreover, cytokines such as EPO, SCF and IL-3, which have all been trialed, are not clinically helpful in DBA patients (Sjogren and Flygare 2012; Lee et al. 2015). Chronic transfusion therapy is started in many patients until one year of age or in patients not responsive to corticosteroids and requires the careful assessment of iron, which accumulates in tissues after breakdown of red blood cells. With chronic transfusions, intensive chelation therapy must be started, since the body has no physiologic pathway for removal of excess iron (Vlachos et al. 2008). Hematopoietic stem cell transplantation presents a curative option in DBA, but substantial morbidity and mortality present significant concerns (Roy et al. 2005; Mugishima et al. 2007; Fagioli et al. 2014). Gene therapy approaches have been successfully tested in animal models, but significant challenges remain to translate these approaches to the clinic (Flygare et al. 2008; Jaako et al. 2014). Moreover, in a significant fraction of patients, the genetic lesion is not known and thus targeted reconstitution is hindered. The clinical understanding of DBA is also confounded by the observation that about 20% of patients will experience a period of independence from therapy and often have entirely normal blood counts for periods of time. The basis by which this occurs and why some individuals with RP mutations in families with DBA individuals have variable phenotypes remains unknown (Vlachos et al. 2008; Lipton and Ellis 2009).

1.3.2 Pathophysiology of Diamond-Blackfan Anemia

The ribosome is the central unit of translation and is expressed in every cell of the human body. The central question for all ribosomopathies is how the different genetic alterations lead to dysregulation of such a central molecular process like translation and result

in specific clinical phenotypes, often involving bone marrow failure and/or craniofacial or other skeletal defects (Narla and Ebert 2010; Weiss et al. 2012; McCann and Baserga 2013; Boultwood and Pellagatti 2014; Ruggero and Shimamura 2014). Structural studies of the ribosome have provided no profound insight as the ribosomal proteins associated with DBA appear to be scattered all over the ribosomal surface (Ben-Shem et al. 2011; Sulima et al. 2014; De Keersmaecker et al. 2015). Furthermore, many mutations have been shown to cause DBA due to ribosomal haploinsufficiency without changing the protein structure (i.e. deletion of one copy of a ribosomal protein gene) (Vlachos et al. 2008).

Early studies with patient-derived erythroid cells suggested qualitative and quantitative defects of BFU-E and CFU-E progenitor cells and were characterized by apoptosis (Nathan et al. 1978a; Perdahl et al. 1994). RNA interference strategies targeting ribosomal proteins resulted in the block of erythroid development and mimicked defects observed in DBA (Ebert et al. 2005; Flygare et al. 2005; Ebert et al. 2008). Ribosomal proteins mutated in DBA appear to be required for the maturation of individual ribosomal RNAs (rRNA) and subunits. Different steps in ribosomal RNA processing are impaired depending upon whether a ribosomal protein of the small or large subunit is affected (Flygare et al. 2007; Narla and Ebert 2010; Garcon et al. 2013; Farrar et al. 2014). Depletion of a large number of individual ribosomal proteins leads to decreased quantities of the other ribosomal proteins of the same subunit and the resulting 40S or 60S subunits. Thus, the reduced availability of any ribosomal protein may change the rate of ribosome biogenesis. This negatively affects the rate of ribosome production and ultimately reduces the number of functional ribosomal complexes, which is illustrated by altered polysome profiles in DBA (Robledo et al. 2008; Horos and von Lindern 2012; Garcon et al. 2013). Consistent with a reduced number of ribosomes, lymphocytes derived from patients have an impaired protein synthesis capacity (48-73 % of healthy controls) (Cmejlova et al. 2006). Impaired cell growth is observed in DBA patient-derived fibroblasts or induced pluripotent stem cells and is also a common phenomenon observed after ribosomal protein knockdown of a variety of cell types (Ebert et al. 2005; Flygare et al. 2005; Badhai et al. 2009; Garcon et al. 2013; Teng et al. 2013). Slowed cell cycle progression appears to result from impaired translational capacity, which delays synthesis and accumulation of cyclins (Teng et al. 2013). Both reduced translation and cellular proliferation may thus contribute to the frequently observed low birth weight and growth retardation in patients (Ball et al. 1996; Willig et al. 1999). However, short

stature is difficult to evaluate in the context of a severe anemia and corticosteroid use during childhood in patients with DBA (Vlachos et al. 2008; Lipton and Ellis 2009).

A number of animal models exist for DBA, but do not appear to reliably recapitulate all aspects of the human disease. Mice with heterozygous depletion of *Rps19* do not present with an obvious hematopoietic phenotype while complete deletion leads to early embryonic lethality (Matsson et al. 2004; Matsson et al. 2006). Zebrafish deficient for *rps19* present with developmental abnormalities and defective erythropoiesis (Danilova et al. 2008), and mice with inducible knockdown of *Rps19* develop macrocytic anemia (Jaako et al. 2011). However, unlike most patients with DBA, these mice also show defects in multiple hematopoietic lineages, including a significant reduction in lymphocyte counts, presumably due to general cell cycle defects. Interestingly, concomitant p53 deletion ameliorated the phenotypes in both instances (Danilova et al. 2008; Jaako et al. 2011) and in a mouse model of 5q- syndrome (Barlow et al. 2010). Subsequent work proposed selective activation of p53 in erythroid cells to provide a basis for the failure of erythropoiesis in DBA (Dutt et al. 2011).

A prevalent model of DBA suggests that the specific loss of a ribosomal protein leads to an imbalance of the remaining ribosomal proteins, resulting in p53 activation causing cell cycle arrest and apoptosis (Zhang and Lu 2009; Narla and Ebert 2010; Weiss et al. 2012). Mouse double minute 2 (MDM2) is an ubiquitin ligase regulating the degradation of p53 (Fang et al. 2000) and has been shown to interact with 5S rRNA and RPL5 (Marechal et al. 1994). Several free ribosomal proteins have subsequently been shown to interact with MDM2, including RPL5, RPL23, RPL11, RPS7, and RPL26 (Zhang and Lu 2009; Narla and Ebert 2010; Bursac et al. 2012). Excessive ribosomal proteins bind and inhibit MDM2, resulting in the accumulation of p53 and subsequent cell cycle arrest and apoptosis (Figure 3) (Weiss et al. 2012). Consistent with this hypothesis, a MDM2 point mutant disrupting the interaction with RPL5 and RPL11 reverses the p53 response upon induction of ribosomal protein deficiency (Macias et al. 2010) and ameliorates the anemia in a mouse model of DBA (Jaako et al. 2015).

While this model is widely accepted, its role with respect to the specificity of the erythroid phenotype in DBA remains controversial. RPL5 and RPL11 have been shown to protect each other from degradation and are both required and essential for p53 activation (Bursac et al. 2012). Consistent with this finding, the knockdown of either RPL11 or RPL5

does not lead to p53 activation in a number of cell types (Bursac et al. 2012; Teng et al. 2013). Since both *RPL5* and *RPL11* mutations have been identified in DBA (Gazda et al. 2008), it remains unclear to what extent p53 activation generally contributes to the underlying erythropoietic defect. Moreover, recent work in zebrafish models have suggested that only the morphological defects are alleviated by co-inhibition of p53 (Torihara et al. 2011; Yadav et al. 2014) and thus suggest p53-independent cell cycle and erythroid differentiation defects as also identified by other studies (Jia et al. 2013; Singh et al. 2014).
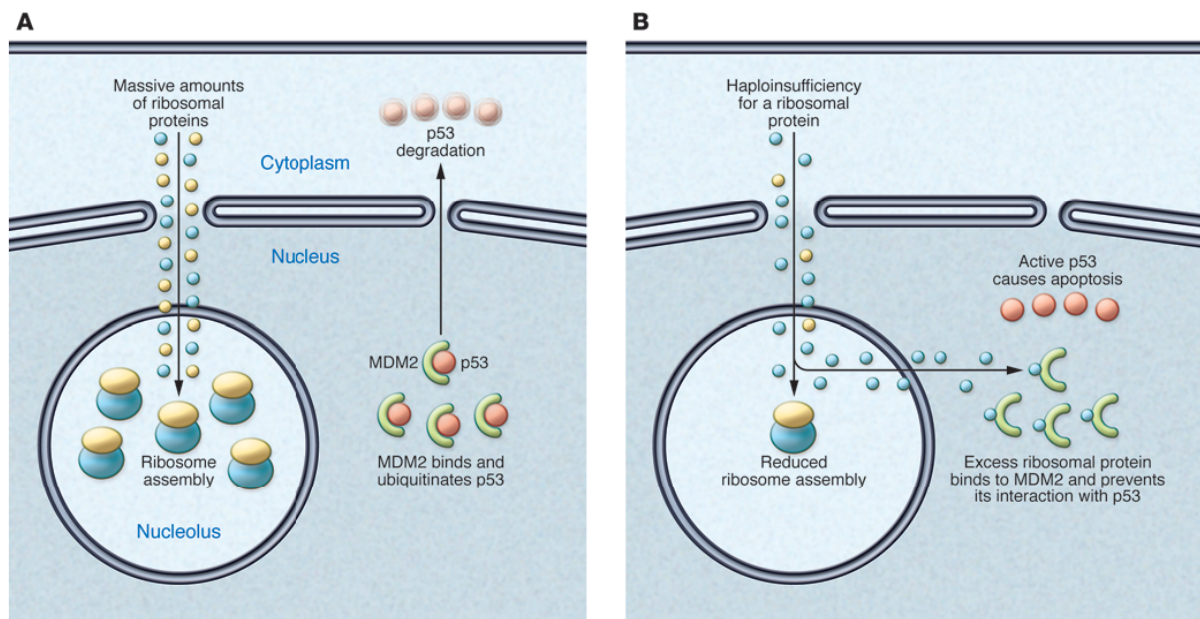
Figure 3 A model for how RP haploinsufficiency causes DBA. **(A)** Normal erythroblasts produce large numbers of ribosomes for protein synthesis. Levels of p53 remain low via a feedback loop whereby MDM2, a transcriptional p53 target, ubiquitinates p53 to promote its degradation by proteasomes. **(B)** Haploinsufficiency for specific RPs causes accumulation of other RPs, which bind to MDM2, thereby inhibiting its ability to promote p53 degradation. Consequently, p53 accumulates and triggers cell cycle arrest and apoptosis. From Weiss et al. 2012.

1.3.3 Translation in Diamond-Blackfan Anemia and Disease

A role for translation in the pathogenesis of DBA has long been discussed and a few studies have investigated the role of altered translation in DBA. Deregulated translation initiation has been identified for selected transcripts, including *BAG1*, *CSDE1* and *BCAT1* (Horos et al. 2012; Pereboom et al. 2014). *Bag1*$^{-/-}$ mice die of anemia during embryonic development (Gotz et al. 2005) and *CSDE1* is essential for cell cycle progression (Horos et al. 2012; Horos and von Lindern 2012). However, their expression was only moderately reduced in patient samples and may not be sufficient to explain the anemia (Horos et al. 2012). *BCAT1* is involved in the metabolism of branched-chain amino acids leucine, isoleucine and valine,

but its contribution to the pathophysiology underlying DBA remains unclear (Pereboom et al. 2014). None of these genes have been found mutated in patients with DBA and reduced translation of a set of transcripts may thus contribute to the anemia (Horos et al. 2012). Long 5' untranslated regions (UTR) and internal ribosomal entry site (IRES) elements have been suggested to play a role in the impaired translation of these transcripts, but no common regulatory element or specific concluding mechanism has yet been identified with respect to translation deficits in DBA (Horos et al. 2012; Pereboom et al. 2014).

Leucine has been suggested in some instances to improve translational efficiency in lymphocytes from DBA patients (Cmejlova et al. 2006) and a case report described the successful treatment of a DBA patient with leucine (Pospisilova et al. 2007). Leucine has subsequently been shown to improve the anemia in cellular models of DBA and 5q-syndrome by modulating mammalian target of rapamycin (mTOR) signaling and appeared to act independent of p53 (Jaako et al. 2012; Payne et al. 2012; Boultwood et al. 2013; Yip et al. 2013; Narla et al. 2014; Yadav et al. 2014; Zhang et al. 2014). The mTOR pathway impacts fundamental cellular functions and integrates a variety of environmental cues to regulate translation, growth and homeostasis (Laplante and Sabatini 2012). A critical role of mTOR signaling has recently been described in murine erythropoiesis and hemoglobin production, and genetic activation or inhibition of this pathway may result in macrocytic or microcytic anemia, respectively (Knight et al. 2014; Chung et al. 2015). Activation of mTOR in mice results in macrocytic anemia, a feature of DBA, but presents with reticulocytosis (high reticulocyte count) and signs of compensatory extramedullary erythropoiesis like splenomegaly (Socolovsky et al. 2001). Both characteristics appear distinct from human patients and the relevance of mTOR signaling remains to be more thoroughly evaluated as clinical trials with leucine in DBA are under way (https://clinicaltrials.gov/, Identifier: NCT01362595 and NCT02386267).

Beyond DBA, translational defects leading to specific phenotypes have been implicated in a variety of conditions, including in disorders of autism (Gkogkas et al. 2013; Santini et al. 2013), and heterozygous loss of function mutations in *RPSA*, encoding for ribosomal protein SA, have recently been recognized as the cause of isolated congenital asplenia (without anemia or other cytopenias) (Bolze et al. 2013). Repressed synthesis of individual ribosomal proteins results in protein-specific and morphological phenotypes in yeast (Thapa et al. 2013) and ribosomal protein paralogs appear to play essential and

antagonistic roles in hematopoiesis (Zhang et al. 2013b). Mutations in the *Rpl38* gene have been associated with impaired translation of *Hox* mRNAs in murine skeletal patterning (Kondrashov et al. 2011). IRES elements within the 5'UTR of *Hox* mRNAs that are regulated by Rpl38 have been suggested to be required for proper expression during development (Xue et al. 2015). Ribosomal protein mutations in *RPL5* and *RPL11* have recurrently been identified in T-cell acute lymphoblastic leukemia (T-ALL) and when expressed in yeast and lymphoid cells the respective mutants resulted in ribosome biogenesis defects (De Keersmaecker et al. 2013; Tzoneva et al. 2013; De Keersmaecker et al. 2015). Both *RPL5* and *RPL11* are also frequently mutated in DBA (Gazda et al. 2008), and some of the mutations appear identical to mutations described in DBA patients (De Keersmaecker et al. 2013). Somatic mutations have also been described for two non-DBA genes *RPL10* and *RPL22* in T-ALL, gastric and ovarian cancer (Wang et al. 2011; Rao et al. 2012; De Keersmaecker et al. 2015). Interestingly, DBA patients do appear to have a predisposition to specific forms of cancer, but if and how ribosomal haploinsufficiency contributes to oncogenesis remains an outstanding question (Lipton and Ellis 2009; Vlachos et al. 2012; De Keersmaecker et al. 2015).

1.3.4 Translational Control and Specialized Ribosomes

These and a myriad of additional studies emphasize the role of translational control in development and disease (Sonenberg and Hinnebusch 2009; Kong and Lasko 2012; McCann and Baserga 2013; Buszczak et al. 2014), and the importance of highly regulated protein synthesis has recently been highlighted in hematopoiesis (Signer et al. 2014). Specific mechanisms have been proposed for selected transcripts (Kondrashov et al. 2011; Horos et al. 2012; Xue et al. 2015), but a profound understanding of the extent to which translational deficits play a role with respect to these ribosomal-associated disorders is still lacking (McCann and Baserga 2013; De Keersmaecker et al. 2015).

A variety of parameters of translation have been scrutinized. Protein production is typically limited by the availability of free ribosomes and translation initiation is one of the most dominant predictors of mRNA abundance on ribosomes (Shah et al. 2013). Cap-dependent translation presents the most general mechanism of translation initiation and requires eukaryotic initiation factors (eIF) to recruit the 5' end (cap) of the mRNA to the 40S ribosomal subunit and subsequently scan the 5'UTR for an AUG start codon (Ruggero and

Shimamura 2014). As a result, structural barriers of the mRNA impairing the ability of the ribosome and associated factors to bind or scan the 5'UTR may reduce the efficiency of initiation (Sonenberg and Hinnebusch 2009). Multiple elements and characteristics of the 5'UTR of mRNAs have been shown to modulate the initiation rate of transcripts by constraining mRNA-ribosome binding, including length, location and folding energies of RNA structures, IRES elements, codon usage, start codon context of the Kozak sequence, non-AUG start codons and upstream open reading frames (uORFs) (Kozak 1991; Babendure et al. 2006; Calvo et al. 2009; Dvir et al. 2013; Shah et al. 2013; Xue et al. 2015). However, even systematic analysis of a library of sequence variants of a 10 bp UTR preceding the translational start site and application of computational modeling only allow prediction of around 70 % of the expression variability (Dvir et al. 2013). A comprehensive understanding of the regulatory code as embedded within the 5'UTR is thus still lacking and will require more rigorous experimental approaches (Dvir et al. 2013). Similarly, the contribution of the poly(A)-tail of mRNAs to translational control has not been completely appreciated (Subtelny et al. 2014). Tail lengths can be coupled to translational efficiencies in early zebrafish, but not in non-embryonic samples and the mechanistic nature of this embryonic switch remains to be elucidated (Subtelny et al. 2014).

Current evidence points to a reduced number of functional ribosomes in DBA and suggests a quantitative defect (Robledo et al. 2008; Garcon et al. 2013). However, qualitative defects have also been proposed in the context of specialized ribosomes (Xue and Barna 2012). *RPL5* and *RPL11* mutations have been noted to be more frequently associated with cleft palate and abnormal thumbs (Gazda et al. 2008), which may suggest that specific ribosomal proteins have unique functions in different cell types (Xue and Barna 2012). Tissue specific expression of ribosomal protein transcripts may support this hypothesis (Kondrashov et al. 2011), but mRNA profiles do no reliably represent protein expression (Schwanhausser et al. 2011) and quantitative differences on the protein level have not been demonstrated. In fact, comprehensive analysis of the riboproteome shows individual ribosomal proteins to be evenly represented across a panel of cell types (Reschke et al. 2013). Various examples of ribosome-associated factors with specific functions may strengthen the notion of a specialized ribosome, while extra-ribosomal roles have also been described for select ribosomal proteins (Mukhopadhyay et al. 2008; Warner and McIntosh 2009; Xue and Barna 2012; Xue et al. 2015). However, aside from mitochondrial ribosomes and specific subcellular localizations, the presence of distinct ribosomes with specialized sub-composition and function has not

been completely clarified and will certainly fuel further research of translation in physiology and disease.

With respect to DBA, decreased protein synthesis and cell cycle progression appear to be general features not restricted to erythroid cells (Ebert et al. 2005; Flygare et al. 2005; Cmejlova et al. 2006; Badhai et al. 2009). It remains unclear to what extent the congenital anomalies can be attributed to these basic defects in cellular function. Morphological defects are not consistently observed (Vlachos et al. 2008), and may arise in a stochastic or unspecific fashion due to altered synthesis of hormones and transcription factors or impaired growth of specific cell populations during embryogenesis and pattern development. In contrast, virtually all patients may present with signs of anemia and this remains the clinical hallmark of this syndrome.

1.3.5 Diamond-Blackfan Anemia and the "GATA1" Controversy

The specificity of the defect in DBA remains an enigma. Treatment options appear limited and are correlated with a number of adverse effects, especially in childhood (Vlachos et al. 2008). Since a genetic defect could only be identified in 50-70 % of patients, this created an incentive to uncover the remaining genetic lesions to gain deeper insight into the underlying mechanism of disease. For this purpose, a cohort of DBA patients that screened negative for ribosomal protein mutations were subject to whole-exome sequencing, which lead to the identification of the first non-ribosomal protein gene mutation in DBA (Sankaran et al. 2012a). The human mutation reported in the DBA patients disrupted a splice site impairing the production of the full-length form of GATA1 (Figure 4). The identical mutation has also been previously reported in a family with congenital anemia and mild dysplasia of the megakaryocytes and neutrophils (Hollanda et al. 2006). Although not diagnosed with DBA, these patients shared clinical characteristics with the patients described by Sankaran and colleagues (Weiss et al. 2012). In 2014, the identical mutation was recognized in an independent family with DBA (Klar et al. 2014), and Parrella et al. identified a mutation in the translation initiation codon also leading to the loss of full-length GATA1 in a patient with DBA (Parrella et al. 2014).
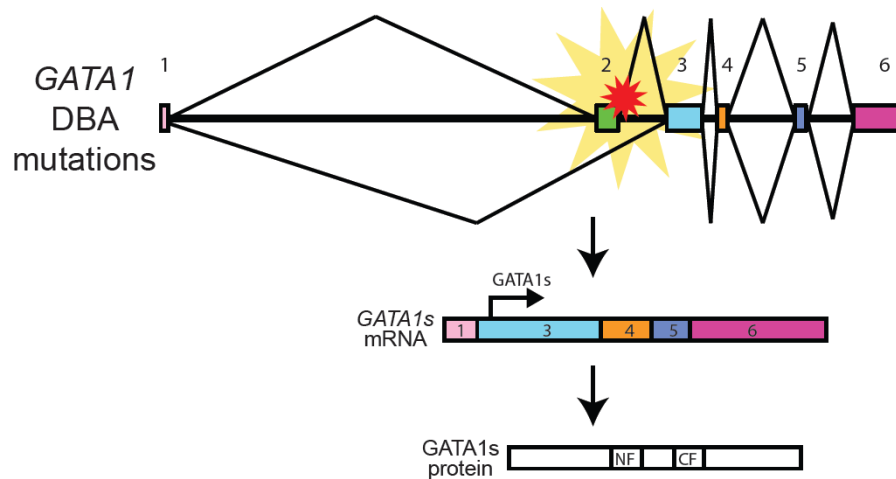
Figure 4 A model of how *GATA1* mutations in DBA favor production of GATA1s alone. Aberrant splicing of exons 2 and 3 (highlighted by red star at mutation sites) leads to production of GATA1s alone. From Sankaran et al. 2012.

Apart from these studies, only ribosomal protein mutations have been described in DBA, making it a classical ribosomopathy. With no known link of GATA1 to the ribosome, the initial finding triggered a controversy about whether the reported mutations in *GATA1* are in fact a "different type of DBA" or "not DBA" (Weiss et al. 2012). Additionally, *GATA1* mutations appear to be quite rare and can only be identified in a small fraction of patients with DBA (Sankaran et al. 2012a; Klar et al. 2014). However, *GATA1* mutations remain attractive to explain or cause anemia and many clinical features are shared between DBA patients with *GATA1* and ribosomal protein mutations (Weiss et al. 2012). Both ribosomal protein mutations and loss of GATA1 are characterized by defects in early progenitor differentiation and apoptosis (Perdahl et al. 1994; Weiss and Orkin 1995). Furthermore, GATA1 associates with and inhibits p53, which may be essential for erythroid cell development and survival (Trainor et al. 2009). Both GATA1 and the ribosome may thus control apoptotic regulators, but could still function by independent pathways resulting in the same phenotype (Weiss et al. 2012). Whether ribosomal protein haploinsufficiency may alter translation of *GATA1* transcripts or its cofactors has not been previously investigated (Weiss et al. 2012). Moreover, similar or identical *GATA1* mutations may present with varying clinical phenotypes (Hollanda et al. 2006; Sankaran et al. 2012a; Klar et al. 2014; Parrella et al. 2014). While these findings clearly underline the power of genetic approaches to uncover disease genes, they also emphasize the importance of functional molecular studies to fully understand the underlying mechanisms.

1.4 Common Variants and Genome-wide Association Studies

Unlike in Mendelian disorders such as DBA, many genetic and environmental factors may influence and predispose individuals to the development of a complex disease including diabetes, cardiovascular disease, cancers or psychiatric disorders (Lupski et al. 2011; Kathiresan and Srivastava 2012). In recent years, advances in next generation sequencing and genotyping have revolutionized the genetics of complex diseases by enabling the discovery of previously unrecognized rare and common variants in the human population (McCarthy et al. 2008; Lupski et al. 2011; Raychaudhuri 2011). The effect or contribution of a specific common allele or variant on the individual developing a phenotype is substantially smaller than that of a Mendelian disease allele, where the presence of a single variant may be sufficient to phenotypically cause a disease (Figure 5). However, due to their deleterious effects, genetic variants causing Mendelian disease are usually found at a much lower frequency, as they are more likely to be selected against over the course of evolution. On the other hand, variants with an intermediate or low effect on a specific phenotype will be more likely to be propagated over generations and are thus more prevalent in the human population (McCarthy et al. 2008; Manolio et al. 2009; Lander 2011).
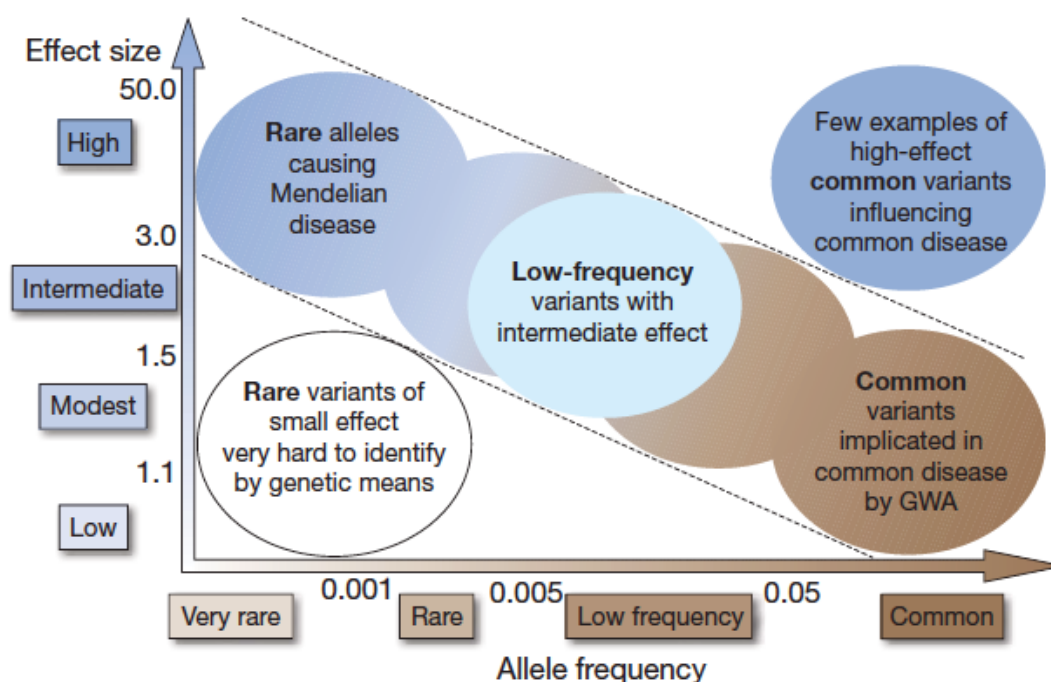

Figure 5 – Scheme illustrating the frequency of genetic alleles and their effect. From: Finding the missing heritability of complex diseases. From Manolio et al. 2009.

A major impact on the studies and discoveries of genetic variants influencing specific diseases or phenotypes has come from genome-wide association studies (GWAS) (McCarthy

et al. 2008; Lander 2011). These studies and functional follow-up thereof have provided new insights into basic biology or pathophysiology and revealed new unsuspected etiologic pathways for common diseases and human traits (McCarthy et al. 2008; Musunuru et al. 2010; Stranger et al. 2011). GWAS examine common genetic variants or single-nucleotide polymorphisms (SNPs) in a population and their association with a specific disease or quantitative trait (e.g. height, cholesterol levels or red blood cell count). The first GWAS was published in 2005 and identified two SNPs associated with an increased likelihood of developing age-related macular degeneration (Klein et al. 2005). A large number of GWAS have investigated the association of genetic variants with a myriad of traits and diseases. As of February 2015, 2,111 publications associating 15,396 SNPs with various traits and phenotypes have been documented (http://www.genome.gov/gwastudies, https://www.ebi.ac.uk/gwas/) (Welter et al. 2014).

The majority of SNPs identified from GWAS have been mapped to non-coding regions, many of which overlap with a specific cell type enhancer or transcription factor binding site (Hindorff et al. 2009; Consortium 2012). Hnisz and colleagues analyzed 5,303 SNPs from 1,675 GWAS studies and found 93 % of variants to be non-coding SNPs (Hnisz et al. 2013). 64 % of the trait-associated variants occurred within enhancer regions and were more enriched in super-enhancers than in typical enhancers (Hnisz et al. 2013; Whyte et al. 2013). Thus, many GWAS variants may modulate disease phenotypes or quantitative traits by influencing gene expression. Some variants appear to occur in cell-type specific regulatory regions of transcription. GWAS mapped variants associated with fetal hemoglobin levels to an erythroid-specific enhancer region modulating expression of the hematopoietic transcription factor BCL11A (Bauer et al. 2013), an essential regulator of globin switching (Sankaran et al. 2008; Sankaran et al. 2009). Genome editing strategies revealed that the enhancer was essential for erythroid, but not B-lymphoid specific expression of BCL11A. The impact of the associated variants was considered modest, but aided in the identification of an important regulatory region that was proposed as a target for therapeutic genome engineering approaches (Bauer et al. 2013). Overall, the Encyclopedia of DNA elements (ENCODE) project indicates over 300,000 regions with enhancer-like features and many genetic variants may thus modulate gene expression in a cell type specific fashion (Consortium 2012). Future studies will also aid to elucidate the role of genetic variants and their effect on the expression or function of noncoding RNAs (Edwards et al. 2013).

1.4.1 Limitations of Genome-wide Association Studies

Despite their power to detect genetic associations, GWAS have also been criticized because of their assumption that common genetic variants play fundamental roles in contributing to the heritability of complex disease (Couzin-Frankel 2010; Fugger et al. 2012; Visscher et al. 2012). More than 10,000 to 100,000 of individuals may have to be genotyped to identify meaningful associations (Teslovich et al. 2010; Gieger et al. 2011; van der Harst et al. 2012; Wood et al. 2014). Most variants have only a small effect size (Figure 5) and thus have a marginal role contributing to the studied disease, questioning the costs and expenditure to identify such variants (Couzin-Frankel 2010; Visscher et al. 2012). Yet, there may be little correlation with the phenotypic effect of one variant with its ultimate biological or therapeutic significance (Fugger et al. 2012; Kathiresan and Srivastava 2012). Common genetic variants within the genes *HMGCR* and *NPC1L1*, encoding for 3-hydroxy-3-methyl-glutaryl-CoA reductase and Niemann-Pick C1-like protein 1, appear to have rather small effects on plasma LDL cholesterol levels (Teslovich et al. 2010). However, both of these genes are important regulators in cholesterol metabolism and successfully targeted by major drugs like statins and ezetemibe (Kathiresan and Srivastava 2012). Furthermore, no rare large-effect Mendelian mutations have so far been identified in *HMGCR*, presumably due to their incompatibility with life. Thus, the relevance of a certain gene may only be recognized through common genetic variants with a small effect size (Fugger et al. 2012; Kathiresan and Srivastava 2012).

However, GWAS by themselves only provide an association between a genetic locus and a specific trait with little direct insight into the mechanistic biology underlying this association (McCarthy et al. 2008; Couzin-Frankel 2010). Modern DNA-microarrays detect about one million SNPs across the human genome. Each SNP on the array is referred to as a sentinel-SNP and locates to a specific locus of the genome. Within that locus, a number of SNPs may be in linkage disequilibrium (LD) with the sentinel-SNP as a result of co-segregation after homologous recombination during meiosis (Jeffreys et al. 2001; Service et al. 2006; Slatkin 2008). Thus, each SNP in linkage is co-associated and may be causal. This significantly complicates the identification of the causal variant within the associated locus. Additionally, many genes may lie in vicinity of the respective variants and thus it is not immediately evident which gene and by what mechanism it is regulated or affected by one or more variants in the associated locus (McCarthy et al. 2008). Furthermore, both the functions of those candidate genes and the cell type(s) where this association is in fact of relevance for the quantified trait are often unclear. Overall, in only a few instances has a non-coding causal

variant, its targeted gene, the function of the gene, and its association with the original phenotype been described (Edwards et al. 2013).

Many of the outlined challenges may be overcome by consolidating a combination of methods, including fine mapping of causal variants, analysis of gene expression data, reporter assays, loss of function studies, transcription factor binding patterns or chromatin conformation capture assays to allow identification of the causal variant and the associated gene (Sankaran et al. 2008; Musunuru et al. 2010; Bauer et al. 2013; Edwards et al. 2013; Smemo et al. 2014; Farh et al. 2015). Yet, exact identification of the functional variant remains challenging (Edwards et al. 2013; Bielczyk-Maczynska et al. 2014; Thom et al. 2014; Ludwig et al. 2015), and failure to properly associate the variant and the candidate gene may lead to misleading results and interpretations. Association studies identified variants within introns of the gene *FTO* associated with obesity and diabetes (Dina et al. 2007; Frayling et al. 2007). Initial studies demonstrated a role of the *FTO* gene in body mass and composition, but were unable to identify a direct correlation between the associated variants and FTO expression (Fischer et al. 2009; Church et al. 2010). Eventually, a subsequent study revealed the obesity-associated variants within the *FTO* gene to be functionally connected, at megabase distances, to the gene *IRX3* (Smemo et al. 2014). The proper identification of the causal variant is thus an important and challenging task when studying the functional biology of GWAS derived genetic variants.

1.4.2 Genome-wide Association Studies and Erythropoiesis
A variety of GWAS have revealed common genetic variants that are associated with hematologic traits and diseases (Uda et al. 2008; Gieger et al. 2011; Sankaran and Orkin 2013). With respect to erythropoiesis, a meta-analysis of multiple GWASs has uncovered at least 75 loci that affect several quantitative erythrocyte traits, i.e. hemoglobin concentration (Hb), hematocrit (Hct), mean corpuscular volume (MCV), mean corpuscular hemoglobin (MCH), mean corpuscular hemoglobin concentration (MCHC) and red blood cell count (RBC) (Ganesh et al. 2009; Soranzo et al. 2009; Kamatani et al. 2010; van der Harst et al. 2012). Principal among these are the RBC and MCV, which are heritable traits that show natural variation in humans. Moreover, variation in these traits is also characteristic of specific pathologic conditions affecting erythropoiesis. For example, the megaloblastic anemias and certain congenital anemias like Diamond-Blackfan anemia are characterized by a

markedly elevated MCV (Vlachos et al. 2008). Many variants and their neighboring candidate genes are known to be involved in processes affecting erythropoiesis, including globin synthesis, erythrocyte membrane function and iron metabolism (van der Harst et al. 2012). However, for the majority of candidate genes a specific role in red blood cell production has not been previously recognized. Many non-coding variants associated with erythroid traits appear to act intrinsically (Paul et al. 2013; Slowikowski et al. 2014), and most studies that have followed up on results of GWAS identified intrinsic regulators that modulate erythroid traits and have revealed new aspects of erythroid biology (Sankaran et al. 2008; Bielczyk-Maczynska et al. 2014; Thom et al. 2014; Ludwig et al. 2015). In contrast, a smaller number of candidate genes identified by GWAS, including *TMPRSS6*, have been implicated in iron metabolism. These are more likely to be extrinsic regulators of erythropoiesis that act by controlling systemic iron homeostasis, thus indirectly affecting red blood cell production (Du et al. 2008; Benyamin et al. 2009; Chambers et al. 2009; van der Harst et al. 2012).

1.4.3 GWAS identify a Variant near *CCND3* associated with RBC and MCV

Among variants identified by GWAS of erythroid traits, the reference SNP (rs) rs9349205 was the most significantly associated with RBC and MCV. Both traits show extensive inverse variation across species (Boylan et al. 1991) and a similar inverse association was identified in GWAS for rs9349205 and a number of other loci (Ganesh et al. 2009; Soranzo et al. 2009; Kamatani et al. 2010; van der Harst et al. 2012). The variant was found upstream of the gene *CCND3*, which encodes cyclin D3 and is one of three members of the D cyclin family. The D cyclins – cyclin D1, D2 and D3 – are core proteins of the mammalian cell cycle machinery (Ciemerych et al. 2002). Upon induction, D cyclins associate with cyclin dependent kinases (CDKs) CDK4 and CDK6, which phosphorylate and subsequently inactivate the retinoblastoma tumor suppressor and related proteins. This causes release and derepression of E2F transcription factors to promote progression from the G1 to S phase of the cell cycle (Malumbres and Barbacid 2009). Consistent with their role in promoting cell growth, the abnormal expression of D cyclins is involved in the progression of human cancers (Sicinski et al. 1995; Sicinski et al. 1996; Sicinska et al. 2003; Deshpande et al. 2005; Choi et al. 2012), and inhibition of the associated CDKs may produce therapeutic benefits (Shapiro 2006; Malumbres and Barbacid 2009).

The role of the individual cyclins has been examined through the studies of knockout mice of one or multiple D-type cyclins (Sicinski et al. 1995; Sicinski et al. 1996; Ciemerych et al. 2002; Sicinska et al. 2003; Kozar et al. 2004; Cooper et al. 2006; Sicinska et al. 2006). Their expression appears highly orchestrated during embryogenesis, and they are expressed in an overlapping fashion in all proliferating cells (Wianny et al. 1998; Sherr and Roberts 1999). Similar to mice lacking individual cyclin-dependent kinases (Malumbres and Barbacid 2009), genetic deletion of individual D cyclins leads to distinct phenotypes, suggesting individual roles of these cell cycle regulators in the respective cell types. Cyclin D1$^{-/-}$ mice show retinal defects, reduced body size, reduced viability and neurological impairment (Sicinski et al. 1995). Cyclin D2$^{-/-}$ mice present with hypoplasia in ovaries and testes (Sicinski et al. 1996), and mice lacking cyclin D3 show defects in T-lymphocyte development, B-lymphocyte maturation and neutrophil production (Sicinska et al. 2003; Cooper et al. 2006; Sicinska et al. 2006).

Mice expressing only a single D-type cyclin develop normally until late gestation but present with focus abnormalities leading to premature mortality later on (Ciemerych et al. 2002). Cyclin D1$^{-/-}$D2$^{-/-}$ mice lack normal cerebella, cyclin D1$^{-/-}$D3$^{-/-}$ mice present with neurological abnormalities, and cyclin D2$^{-/-}$D3$^{-/-}$ mice develop severe anemia pointing to a role of these cyclins in erythropoiesis. Interestingly, cyclin D1$^{-/-}$D2$^{-/-}$D3$^{-/-}$ triple knockout embryos develop normally as late as embryonic day 13.5 (E13.5), indicating proper cell proliferation during early embryogenesis in the absence of all D cyclins (Kozar et al. 2004). Eventually, cyclin D1$^{-/-}$D2$^{-/-}$D3$^{-/-}$ embryos die prior to day E17.5 due to cardiac failure and severe anemia, as the D cyclins are critical for the expansion of hematopoietic stem cells (Kozar et al. 2004). Together these studies emphasize a role for D cyclins and cyclin D3 in hematopoiesis. However, a specific role for cyclin D3 in erythropoiesis has not been established and the exact mechanism underlying the GWAS association remained elusive.

1.5 Aims of the Thesis

The general aim of this thesis was to functionally follow up on the results of two genetic studies, one of rare variation and one of common variation, relevant to erythropoiesis to gain deeper insight into the underlying biology and pathophysiology of this process. Both studies were independent of each other and thus two independent published manuscripts will be presented and discussed in the scope of this thesis.

The first manuscript follows up on the results of a whole-exome sequencing study of a cohort of patients with Diamond-Blackfan anemia and will discuss the role of GATA1 in the pathophysiology of this disease. The second manuscript follows up on results of recent GWASs, which have identified common DNA variants associated with erythroid traits and will illustrate the role of cyclin D3 in terminal erythropoiesis.

Specific Aim 1

Most cases of Diamond-Blackfan anemia are classically caused by haploinsufficiency of ubiquitously expressed ribosomal proteins. The disease is characterized by a reduction in red blood cell counts, and current models of the underlying pathophysiology are unable to explain the exquisite specificity of this defect. Recent work has identified specific mutations in the transcription factor GATA1 in patients with Diamond-Blackfan anemia. GATA1 is essential for erythropoiesis, and a variety of mutations have been recognized to cause anemia in humans. Since no mechanistic link exists between GATA1 and the ribosome, the relationship between these patients and those with the more commonly observed ribosomal protein mutations in classical DBA remains unclear. The major aims of this project were to

1. Investigate whether translation of *GATA1* mRNA is impaired upon ribosomal protein knockdown or haploinsufficiency.
2. Examine how altered translation of *GATA1* mRNA may contribute to the pathophysiology in Diamond-Blackfan anemia.
3. Identify a possible mechanism of altered translation of *GATA1* mRNA.

Specific Aim 2

Genome-wide association studies implicate the SNP rs9349205 as inversely associated with mean corpuscular volume and red blood cell count. The SNP is located upstream of the gene *CCND3*. Although *Ccnd2$^{-/-}$Ccnd3$^{-/-}$* double knockout mice have been noted to be severely anemic and die in mid-gestation, a role of cyclin D3 in erythropoiesis has not been previously explored. The major aims of this project were to

1. Determine if cyclin D3 is regulated by rs9349205.
2. Identify the mechanism of action by which rs9349205 may influence expression of cyclin D3.
3. Explore the role of cyclin D3 in erythropoiesis and its relevance to MCV and RBC.

2. Manuscript I — Altered translation of GATA1 in Diamond-Blackfan anemia

Leif S. Ludwig, Hanna T. Gazda, Jennifer C. Eng, Stephen W. Eichhorn, Prathapan Thiru, Roxanne Ghazvinian, Tracy I. George, Jason R. Gotlib, Alan H. Beggs, Colin A. Sieff, Harvey F. Lodish, Eric S. Lander and Vijay G. Sankaran

Experimental contribution

I conceived the project together with Vijay Sankaran. I planned, performed and evaluated most experiments with assistance from Vijay Sankaran and Jennifer Eng. Hanna Gazda and Colin Sieff obtained patient and healthy donor samples for genetic studies, gene expression analysis and additional experiments requiring human cells. Stephen Eichhorn aided in polysome profiling and reporter analysis. Prathapan Thiru conducted bioinformatic analysis. Roxane Ghazvinian performed genetic sequencing studies. Tracy George and Jason Gotlib provided clinical assessments. Vijay Sankaran supervised all experimental and analytic aspects of this project. Alan Beggs, Harvey Lodish and Eric Lander provided intellectual contributions to the project. I wrote the manuscript together with Vijay Sankaran, Harvey Lodish and Eric Lander with input from all authors.

The following figures present my work:

Figures 1c-d, 2a-c, 3a-f, 4e-i.

Supplementary Figures 3, 4, 6, 7, 8, 9, 10, 11, 12, 13d, 14c, 17, 18, 20.

The original article and Supplemental tables 1-7 and the original article are available online:

http://dx.doi.org/10.1038/nm.3557

http://www.nature.com/nm/journal/v20/n7/full/nm.3557.html

Supplementary figures are included on the following pages.

SUPPLEMENTAL MATERIAL FOR:

Altered translation of GATA1 in Diamond-Blackfan anemia

Leif S. Ludwig, Hanna T. Gazda, Jennifer C. Eng, Stephen W. Eichhorn, Prathapan Thiru, Roxanne Ghazvinian, Tracy I. George, Jason R. Gotlib, Alan H. Beggs, Colin A. Sieff, Harvey F. Lodish, Eric S. Lander, and Vijay G. Sankaran*

*Correspondence should be addressed to V.G.S. (sankaran@broadinstitute.org)

Supplementary Figure Legends

Figure 1. Bone marrow aspirate (above at 100X objective magnification) and biopsy (below at 50X objective magnification) from a healthy individual. Morphologically identifiable erythroblasts are indicated by red arrows.

Figure 2. Scheme illustrating the production of full-length and short protein forms of GATA1 from the full-length mRNA.

Figure 3. Chromatograms of *GATA1* cDNA wild-type and T>C mutant vectors fail to reveal signs of contamination accounting for full-length protein production. Note the arrow below the C nucleotide in the mutant cDNA trace with no signs of even trace amounts of the wild-type T allele present. The cDNAs were cloned into the HMD lentiviral vector.

Figure 4. Reduction of GATA1 and RPS19 protein expression 4-6 days after infection with RPS19 shRNAs. (a) This time course reveals the decreased amount of these proteins present after even 4 days following infection. Additionally, the protein expression of erythroid-important factors CD71/TFRC, EPOR, JAK2 (both bands correspond to this protein), STAT5A (doublet), and TAL1 is shown. GAPDH and β-Actin loading controls are shown. No changes in these other proteins are noted over this time course. Arrowheads indicate the major protein forms of GATA1 full length and GATA1 short. (b) Growth curves for RPS19 shRNA cells. Growth after infection is measured in absolute cell numbers for the control and RPS19 shRNAs, sh916 and sh913. The data are shown as the mean ± the standard deviation (n=3).

Figure 5. Relative *GATA1* mRNA levels in comparison with β-actin mRNA levels in the monosome (80 S) fraction from the polysome gradient fractionation. The data are shown as the mean ± the standard deviation (n=3).

Figure 6. Ribosomal protein haploinsufficiency results in reduced translation of GATA1. (a) K562 erythroid cells show similar GATA1 protein levels with shRNA-mediated haploinsufficiency of RPS19 at an early time point three days post-infection. Arrowheads indicate the major protein forms of GATA1 full length and GATA1 short. (b) Western blot of GATA1 protein levels after immunoprecipitation of GATA1 using the M20 antibody. Similar total levels of proteins are immunoprecipitated in control and RPS19 knockdown cells. (c) Detection of L-azidohomoalanine levels after 4 hours of labeling and GATA1 immunoprecipitation shows a dramatic reduction in newly synthesized GATA1 protein.

Figure 7. Relative L-azidohomoalanine incorporation over 4 hours of labeling to monitor the rates of de novo protein synthesis in K562 erythroid cells for the control (pLKO1.s) and RPS19 shRNAs, sh916 and sh913.

Figure 8. Knockdown of ribosomal proteins correlates with protein expression levels of GATA1. The expression level of ribosomal proteins were quantified for each set of shRNAs targeting RPL5, RPL11, RPS19 or RPS24 and were plotted against the respective expression level of GATA1 relative to the control (pLKO1.s). A linear regression was drawn for each individual set of shRNAs. R^2 values are shown for each linear regression.

Figure 9. Knockdown of GATA1 increases apoptosis and inhibits cell growth in erythroid K562 cells. (a) Increased Annexin V staining with RPS19 or GATA1 Knockdown. Results are shown as the percentage of Annexin V positive cells. Cells were infected with shRNAs targeting RPS19 or GATA1 (as labeled). The data are shown as the mean ± the standard deviation (n=3 per group; ** $p < 0.01$, *** $p < 0.001$). (b) Growth curves are shown for control vector transduced cells (GFP and pLKO1.s) or for GATA1 shRNA transduced cells (sh19-sh23). The data are shown as the mean ± the standard deviation (n=3). (c) Western blot of GATA1 levels following knockdown with shRNAs is shown. Arrowheads indicate the major protein forms of GATA1 full length and GATA1 short.

Figure 10. Treatment of *in vitro* differentiated primary cells with shRNAs against RPS19, RPL11, RPL5, and GATA1 impair erythroid differentiation and favor differentiation of non-erythroid cells. Erythroid cells are marked by CD235a expression. Non-erythroid cells are marked by the expression of CD41a, CD11b, or expression of no markers.

Figure 11. 4EGI-treatment of *in vitro* differentiated primary cells impairs erythroid differentiation indicated by reduced CD235a expression and increase in CD41a and CD11b positive cells and double negative cells. FACS plots for controls and various concentrations of 4EGI-1 are shown.

Figure 12. GATA1 expression rescues erythroid differentiation of 4EGI-1 treated primary CD34+ cells. Representative FACS plots are shown of cells treated for 48h with DMSO and 40µM 4EGI-1 in cells transduced with GATA1 or empty HMD control lentivirus. 4EGI-1 was added to cultures on day 2 post infection, when infected cells start to coexpress GFP. The ratio of erythroid (CD235a+) to non-erythroid (CD235a-) -cells is shown on the right. The data are shown as the mean ± the standard deviation (n=3).

Figure 13. The 5' UTR of GATA1 mRNA is highly self-complementary. (a) Chromatogram showing the 5' end of the 5' UTR of GATA1 mRNA as determined by 5' RACE. (b) Sequence of the complete 113bp 5' UTR of GATA1 mRNA. The ATG start codon is marked in red. (c) Predicted secondary structure of the 5' UTR of GATA1 mRNA. (d) The 5' UTR of GATA1 restricts translation as measured by a luciferase reporter assay. The 5' UTRs of GAPDH, ACTB, and GATA1 were cloned downstream of the major transcription initiation start site of the SV40 promoter replacing the 5' UTR of the pGL3-SV40 promoter Firefly luciferase vector. Data are shown as the mean ± the standard error of the mean (n = 6 per group) and are normalized to Renilla luciferase activity, and the activity of the construct carrying the GAPDH UTR (p-values are shown above respective bars in comparison with GAPDH; N.S. = not significant). Free energy analysis for the different 5' UTRs is indicated below, based on analysis by the RNAfold web server (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi).

Figure 14. Transcripts with structured 5' UTRs show decreased association with larger polysomes in the setting of reduced ribosomal protein levels. (a) Reduced association with larger polysomes (> 4 ribosomes) compared to monosomes for genes with highly-structured 5' UTR elements. Data are plotted on a log10 scale (n=3 per group; *** $p < 0.001$). (b) Predicted RNA folding pattern of the 5' UTR of these genes. (c) Reduction of RPS20 protein expression 4-6 days after infection with RPS19 shRNAs confirms reduced translation as noted from association with larger polysomes.

Figure 15. Selection of GATA1 target genes from GSE628. (a) Here the 0 hour and 21 hour time points after GATA1 induction are highlighted (corresponding to the late BFU-E and polychromatophilic erythroblast stages, respectively). By 21 hours, the majority of validated GATA1 targets are induced and therefore this time point was chosen to create a gene set of experimentally derived GATA1 up-regulated target genes for GSEA analysis (226 genes with log2 fold change > 1 with p-values < 0.05). (b) Here the 0 hour and 30 hour time points are highlighted (corresponding to the late BFU-E and orthochromatophilic erythroblast stages, respectively). 260 such genes with log2 fold change > 1 with p-values < 0.05 are shown. GSEA analysis from this gene set is shown on the right side, as in Figure 4.

Figure 16. GSEA analysis of erythroid-important transcription factor targets fails to show global changes in DBA patient erythroid progenitors compared with controls. (a) The genes in each plot that are upregulated in DBA patients relative to controls are shown in the left side, while those that are downregulated in DBA patients relative to controls are shown on the right side of each plot. (b) GSEA analysis of p53 transcriptional targets do not show

global upregulation of p53 targets in DBA patient erythroid progenitors compared with controls. The genes in each plot that are upregulated in DBA patients relative to controls are shown in the left side, while those that are downregulated in DBA patients relative to controls are shown on the right side of each plot.
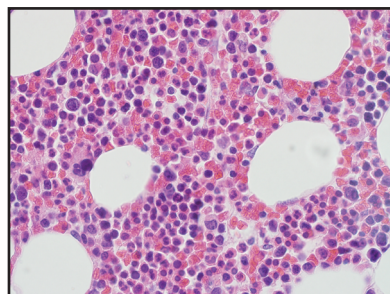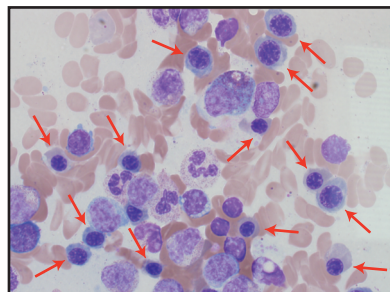
Figure 17. Rescue of impairments in erythroid cells due to reduced ribosomal protein levels by GATA1. (**a**) Representative western blot showing GATA1, RPS19, and GAPDH protein expression for control and sh916 transduced cells (either with empty HMD, HMD-GATA1, or HMD-GATA1 start codon mutant (G1 Mut) which favors production of GATA1s. (**b**) Annexin V staining (percentage of cells are Annexin V positive cells) indicates increased apoptosis of erythroid K562 cells with a reduction of RPS19 levels that can be rescued with exogenous GATA1 expression. Data are shown as the mean ± the standard error of the mean (n ≥ 3 per group; ** $p < 0.01$, *** $p < 0.001$, comparing the different transduced erythroid cell populations to the empty HMD control). (**c**) Representative FACS plots showing increased erythropoiesis on transduction of CD34+ cultured cells with an shRNA targeting RPL11 or RPL5 and either with empty HMD control lentiviruses, GATA1 WT, or the GATA1 mutant. The ratio of erythroid (CD235a$^+$) to non-erythroid (CD235a$^-$) cells is shown below the plots.

Figure 18. Representative cytospin images of cultured DBA patient cells transduced with control or GATA1 show increased differentiation and maturation of GATA1 transduced cells. Transduced cells were sorted based on GFP expression and stained with May-Grünwald-Giemsa staining. Scale bar = 10µm.

Figure 19. A model showing how several select mRNAs, including GATA1, are downregulated in the setting of ribosomal protein haploinsufficiency in DBA while the majority of other mRNAs are relatively less affected. Reduction in GATA1 protein results in decreased activity of this transcription factor, which is necessary for normal erythropoiesis and therefore results in impaired production of red blood cells from hematopoietic progenitors.

Figure 20. Knockdown of RPS19 or GATA1 increases p53 levels in primary erythroid cells. Western blot showing GATA1 and p53 levels of CD34+ cells transduced with shRNAs targeting RPS19 or GATA1. β-Actin was used as a loading control. Arrowheads indicate the major protein forms of GATA1 full length and GATA1 short.

Supplementary Figure 1

Supplementary Figure 3
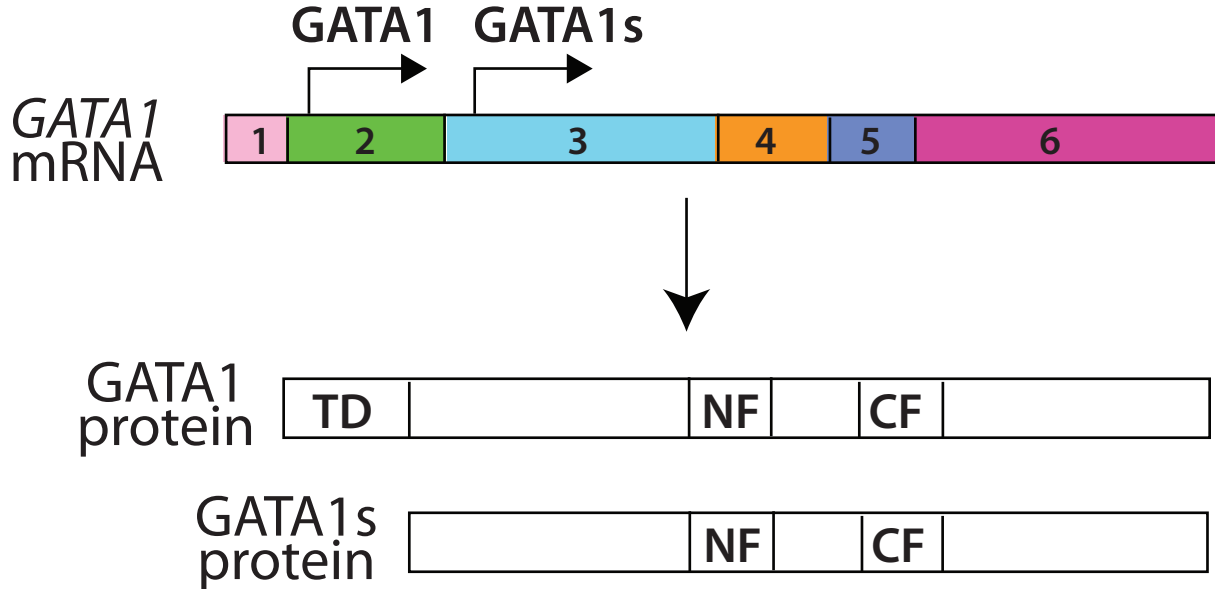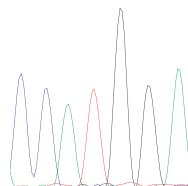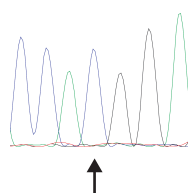
Supplementary Figure 4

Supplementary Figure 5
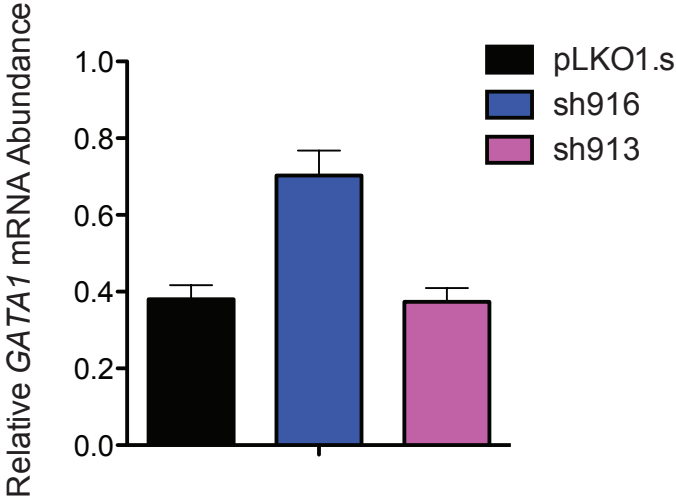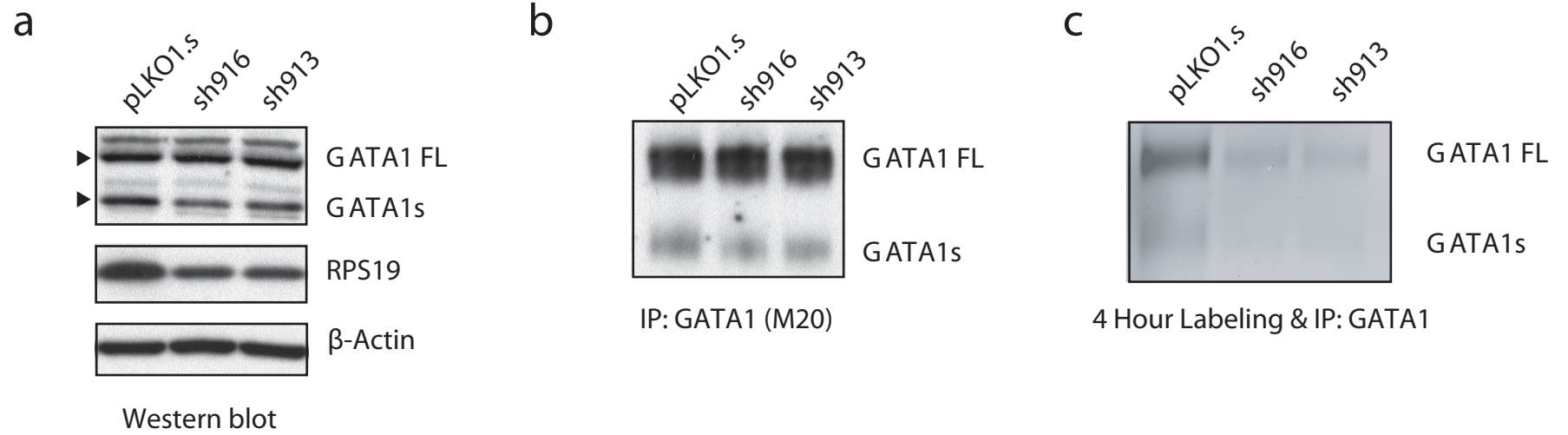
Supplementary Figure 6

a

| | pLKO1.s | sh916 | sh913 |

GATA1 FL
GATA1s

RPS19

β-Actin

Western blot

b

GATA1 FL

GATA1s

IP: GATA1 (M20)

c

GATA1 FL

GATA1s

4 Hour Labeling & IP: GATA1

Supplementary Figure 7

Supplementary Figure 8

a

b

c
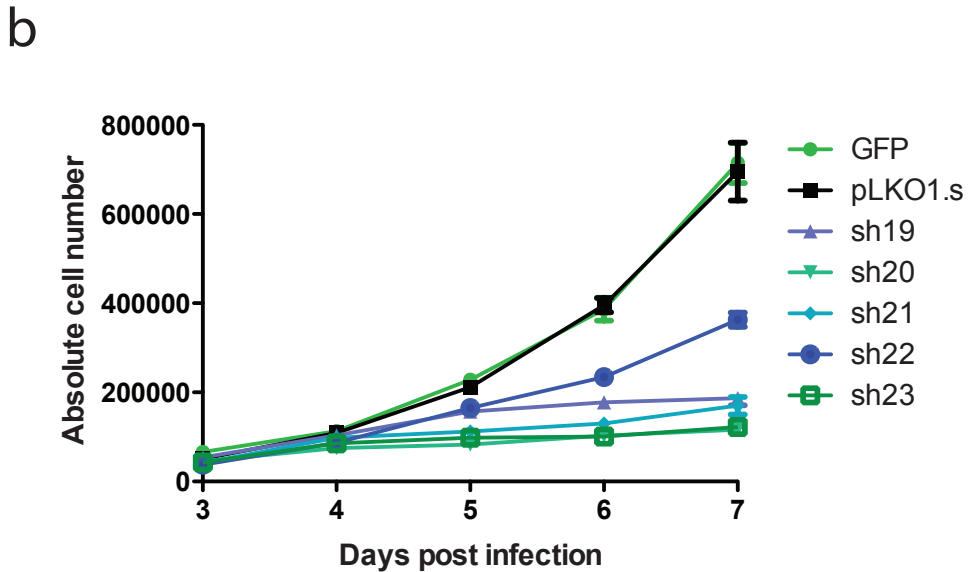
Supplementary Figure 10

Supplementary Figure 11

Supplementary Figure 12

Supplementary Figure 13

a

Adapter   *GATA1* 5' UTR

G A T G A A A C T C A A A G G C C A A G G C C A G C C A G G A C A C C C C C T G G G
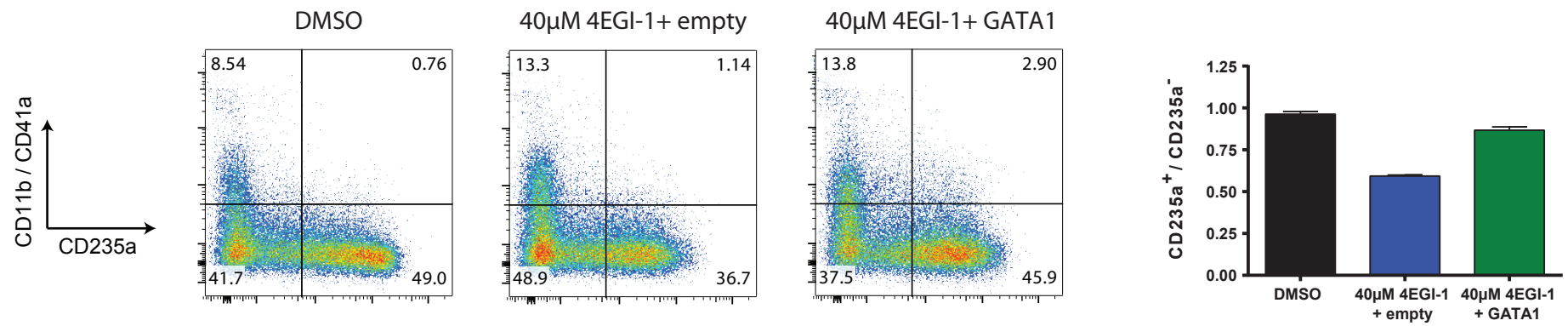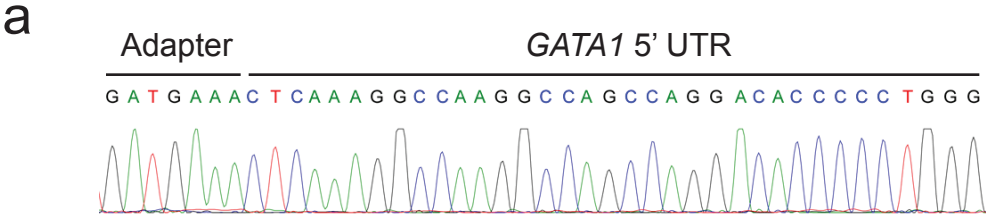
b

CTCAAAGGCCAAGGCCAGCCAGGACACCCCCTGGGATCACA
CTGAGCTTGCCACATCCCCAAGGCGGCCGAACCCTCCGCAA
CCACCAGCCCAGGTTAATCCCCAGAGGCTCCATG

c

d

	GAPDH	ACTB	GATA1
Free energy [kcal/mol]	-21.61	-15.52	-30.23

Supplementary Figure 14

a

b

RPS20　　　　　RBM38　　　　　MTA1

c

Supplementary Figure 15

Supplementary Figure 16

a

E2F4 Targets

FOXO3 Targets

LMO2 Targets

STAT5A Targets

YY1 Targets

Up in DBA

Down in DBA

P-value = 0.264

P-value = 0.283

P-value = 0.346

P-value = 0.886

P-value = 0.874

b

p53 Transcriptional Targets

Enrichment Score

0.15
0.10
0.05
-0.05
-0.10
-0.15

Up in DBA

Down in DBA

Enrichment Score = -0.155
p-value = 0.983

Supplementary Figure 17

a

+ sh916

Control HMD GATA1 G1 Mut

◄ GATA1 FL
◄ GATA1s
RPS19
GAPDH

b

Percentage of Annexin V Positive Cells

40 30 20 10 0

HMD   sh916   sh916 + HMD   sh916 + GATA1   sh916 + G1 Mut

*** *** NS **

c

RPL11 sh4 + empty

29.1   3.32
31.4   36.2

CD11b / CD41a
CD235a

RPL11 sh4 + GATA1

21.5   3.08
26.1   49.4

RPL11 sh4 + GATA1 Mutant

22.6   2.57
27.4   47.4

Erythroid / Non-Erythroid → 0.60   1.04   0.95

RPL5 sh2 + empty

27.2   1.84
36   35

CD11b / CD41a
CD235a

RPL5 sh2 + GATA1

23.3   2.09
28.6   46

RPL5 sh2 + GATA1 Mutant

24.4   1.98
32.9   40.7

Erythroid / Non-Erythroid → 0.55   0.89   0.71

Supplementary Figure 18

+ empty

+ GATA1

Supplementary Figure 19

Supplementary Figure 20
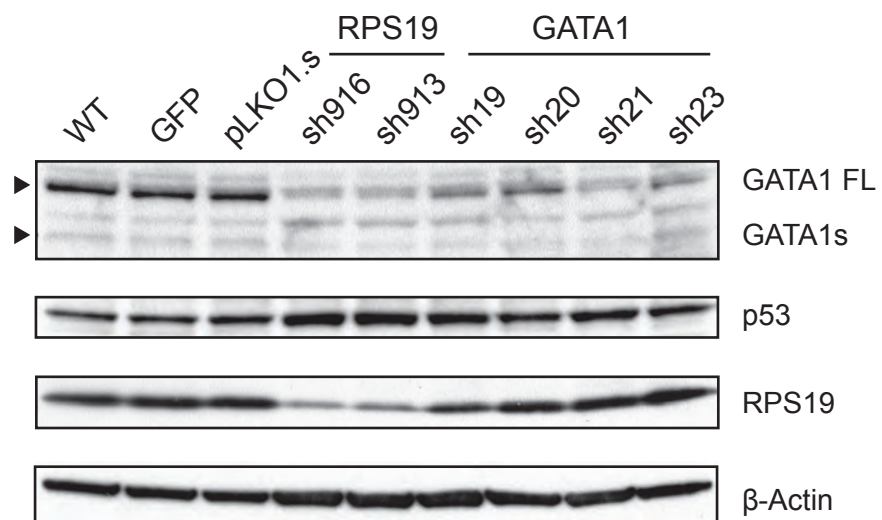
2.1 Additional information not published in the original manuscript

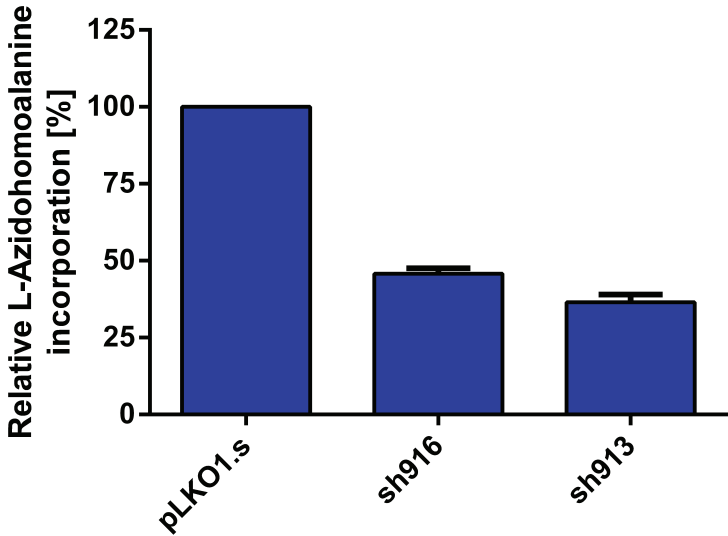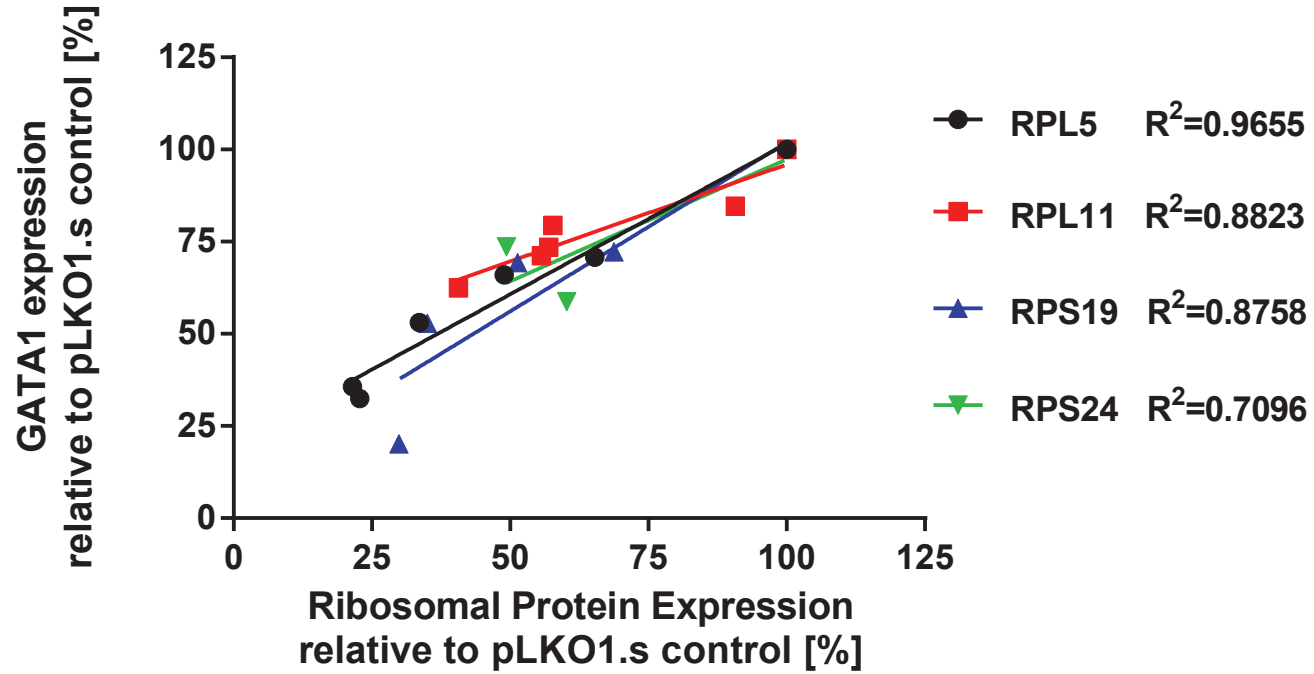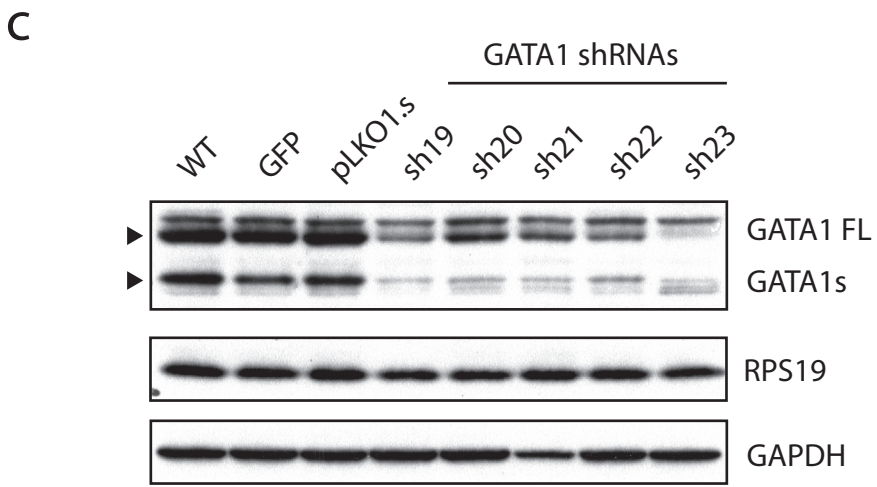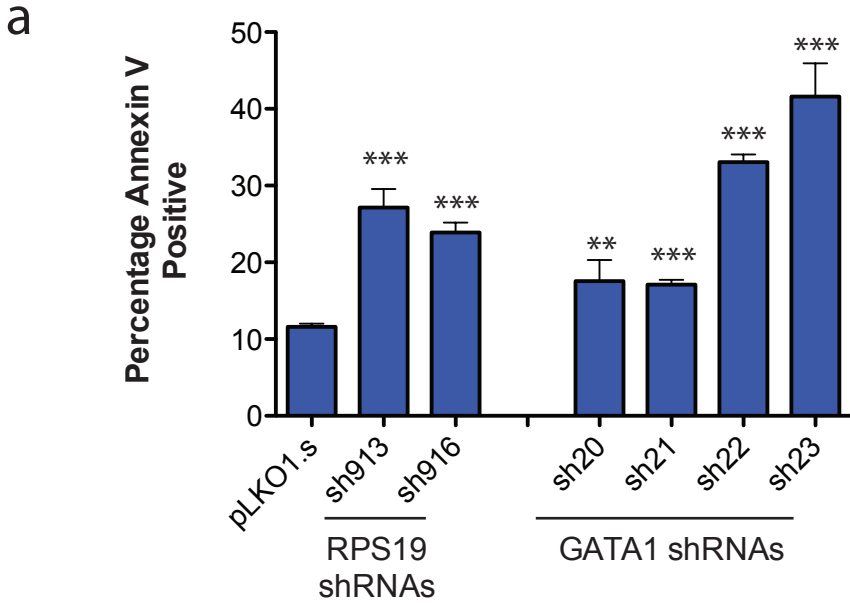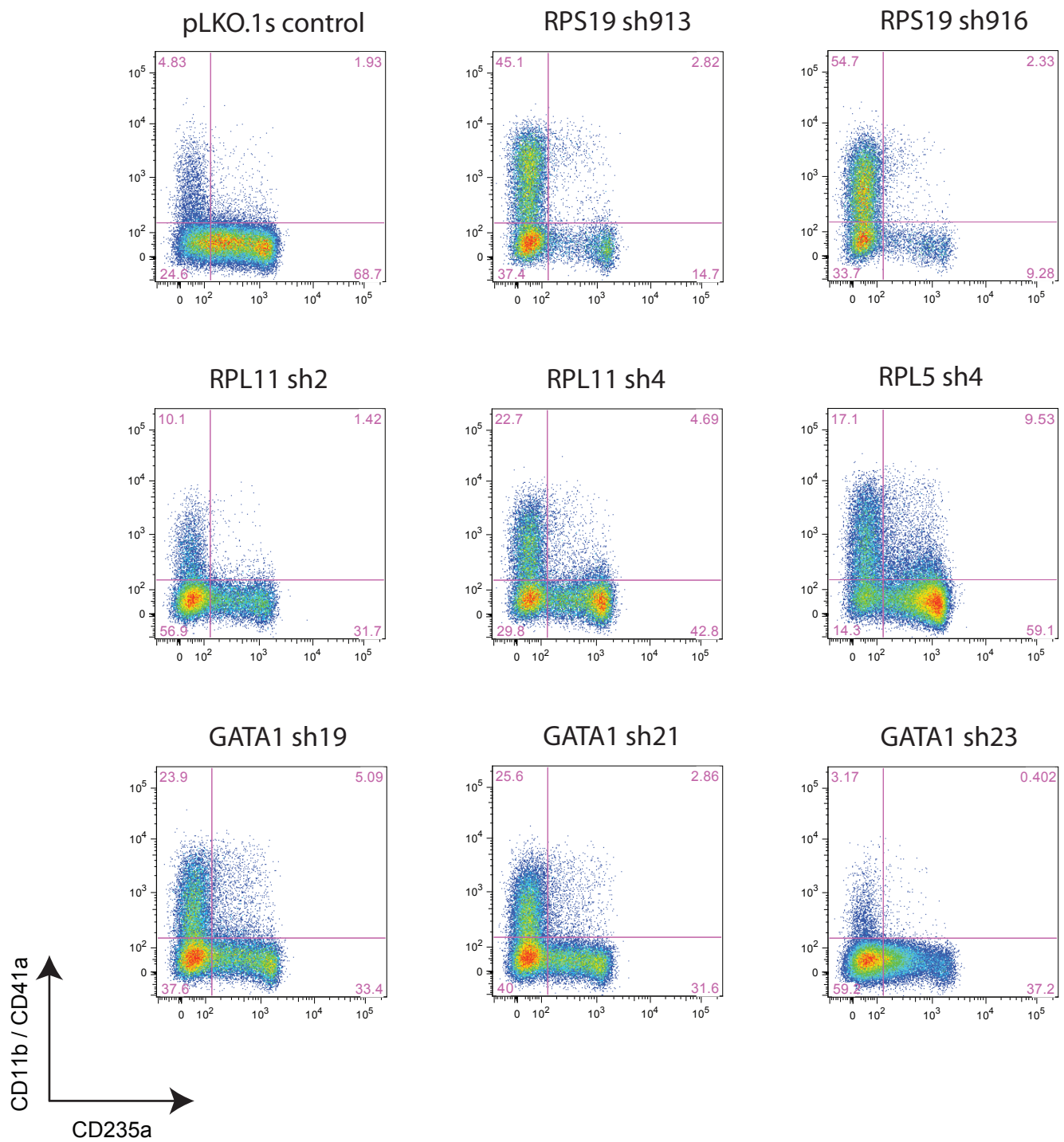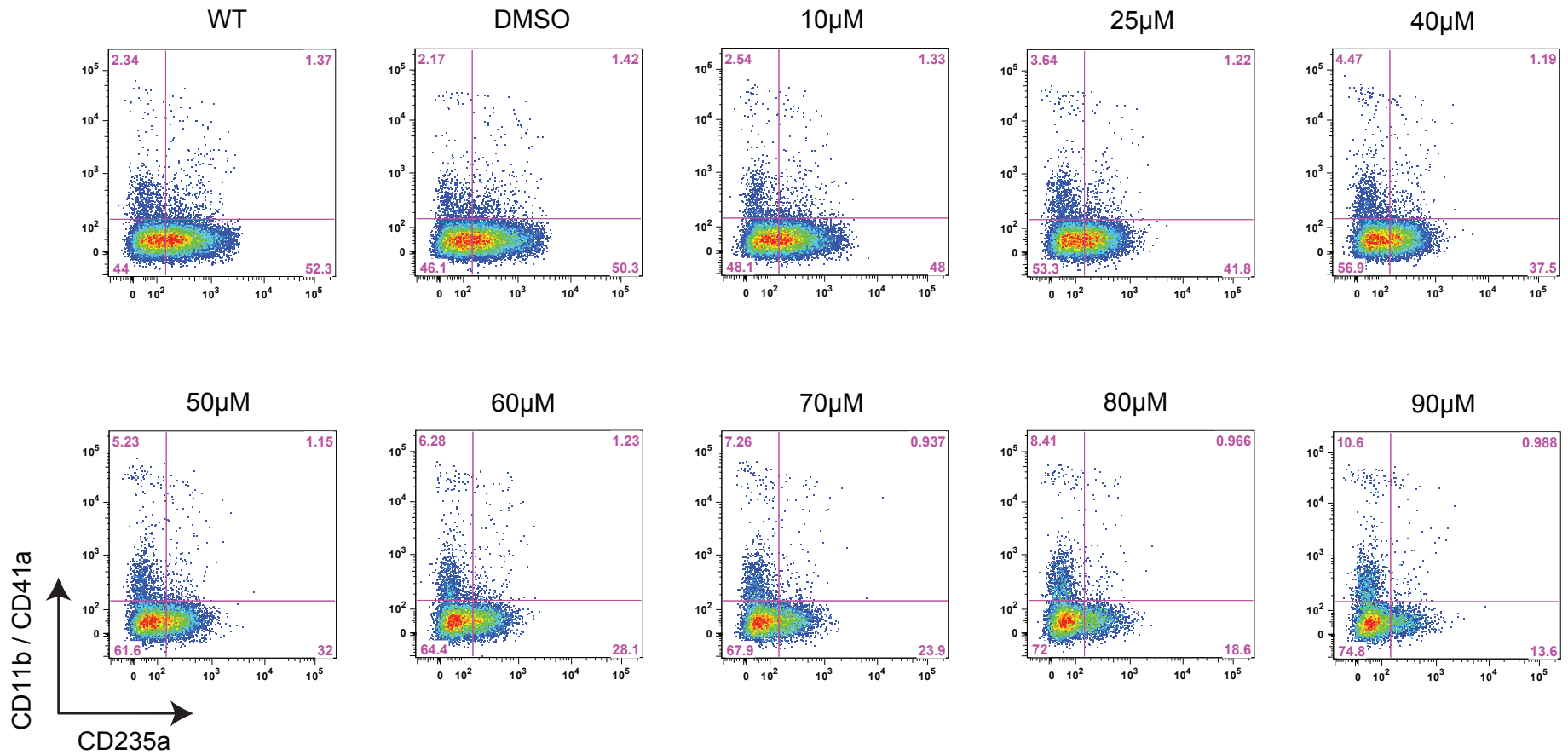Supplementary Figures 21-29 contain information about various aspects of GATA1 and DBA biology. Some of the data is preliminary and awaits more careful assessment but is relevant to the discussion and future work related to the project.

Experimental contribution

Aoi Wakabayashi and Jacob C. Ulirsch from the laboratory of Vijay Sankaran aided in obtaining and analyzing the data. I conducted most experiments with assistance from Aoi Wakabayashi. Aoi Wakabayashi sorted and processed cells for gene expression profiling. Jacob Ulirsch conducted bioinformatic analysis.

The following figures present my work:
Supplementary Figures 21a-b, 22, 25, 26, 27, 28b-c, 29.

2.1.1 GATA1 Activity in Diamond-Blackfan Anemia

To gain further insight into the role of GATA1 in DBA, primary human erythroid progenitor cells were infected with an shRNA control targeting luciferase (shluc) and shRNAs targeting *GATA1*. Knockdown of *GATA1* results in impaired erythroid differentiation as assessed by a reduction in $CD235a^+$ cells and an increase in the frequency of non-erythroid $CD11b^+$ or $CD41^+$ cells as has been reported previously (Supplementary Figure 10, 21a) (Flygare et al. 2005; Ebert et al. 2008). Similar disturbances in differentiation profiles for CD71 and CD235a can be observed when culturing bone marrow derived primary cells from a patient with the diagnosis DBA in comparison with two healthy donors (Supplementary Figure 22). For transcriptional analysis, stage-matched cells were sorted based on their surface marker expression and forwarded to microarray analysis (Supplementary Figure 21b). Both sh20 and sh21 result in the robust knockdown of *GATA1* mRNA (Supplementary Figure 21c). Scatter plots comparing the mean gene expression after *GATA1* knockdown show a high degree of correlation to the shluc control (Supplementary Figure 21d), and gene set enrichment analysis (GSEA) of differentially expressed genes reveals a specific decrease in GATA1 target genes (Supplementary Figure 23a, b). Thus direct knockdown of *GATA1* and transcriptional analysis in primary human erythroid cells resembles results obtained from DBA patient and healthy control cells (Figure 4a, c, d). To gain insight into the nature of the differentially expressed genes, an analysis of super enhancer (SE) regulatory elements was

conducted (Whyte et al. 2013; Pott and Lieb 2015). For this purpose a GATA1 chromatin immunoprecipitation sequencing (ChIP-seq) dataset in erythroid progenitor cells was used to derive SE elements (Supplementary Figure 24a, b) (Ulirsch et al. 2014). Interestingly, genes regulated by GATA1 SEs are expressed at higher absolute expression levels than genes regulated by traditional enhancers (TE) (Supplementary Figure 24c) and appear more sensitive to perturbation of GATA1 levels (Supplementary Figure 24d). Direct comparison of transcriptional profiles from samples with shRNA mediated knockdown of *GATA1* and DBA patients and respective controls reveal that erythroid progenitors from DBA are substantially enriched for the most differential expressed subset of genes as determined by GSEA (Supplementary Figure 24e). Consistent with an important role of GATA1 in DBA, the genes which expression is reduced the most (differential $\log_2$ expression < -0.5), are bound by GATA1 SEs at a significantly higher proportion than expected (Supplementary Figure 24f). Together, these data further support the contention that dosage-dependent reduction in GATA1 activity has a critical role in the pathogenesis of DBA.

2.1.2 Intracellular Flow Cytometry for GATA1

To directly assess the levels of GATA1 using flow cytometry an intracellular staining protocol was established. For this purpose a monoclonal antibody recognizing the N-terminus of GATA1 FL was used and evaluated by western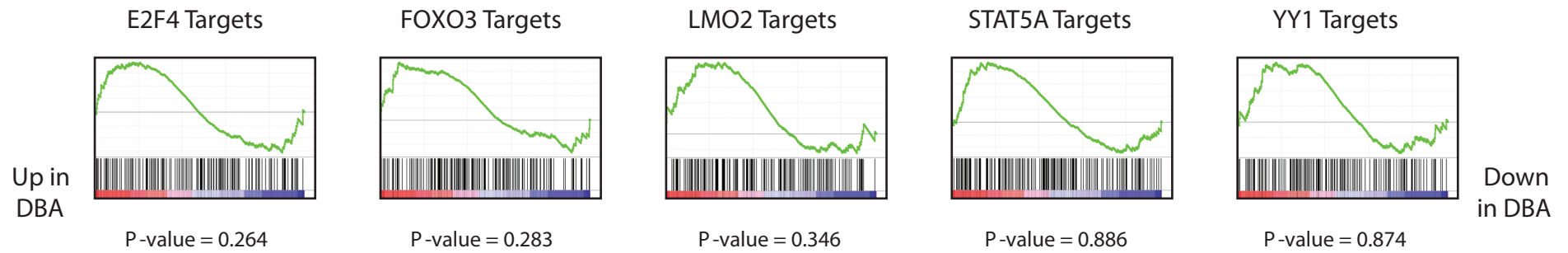 blot. The polyclonal antibody used throughout the study recognizes both GATA1 FL and GATA1s (Supplementary Figure 25a, b). As expected the monoclonal antibody only recognized GATA1 FL and not GATA1s, which lacks the N-terminal transactivation domain (TD) (Supplementary Figure 25c). The specificity of the antibody was further verified by direct depletion of GATA1 using five independent shRNAs (Supplementary Figure 25b, c). The simplified experimental workflow for intracellular flow cytometry is depicted in supplementary figure 26a. Staining of the human erythroid cell line K562 results in significant signal enrichment compared to the isotype control (Supplementary Figure 26b). Analysis of surface stained *in vitro* cultured primary human erythroid cells reveals an increase in signal intensity through successive stages of erythroid differentiation, consistent with the upregulation of GATA1 during erythropoiesis (Supplementary Figure 26c). Subsequent analysis and comparison of GATA1 expression in stage-matched *in vitro* cultured cells revealed differences in staining intensity across a number of healthy donors (Supplementary Figure 27). It remains unclear whether this reflects physiologic variation in GATA1 expression or technical inconsistencies. Due to the

scarce availability of DBA patient cells intracellular assessment of GATA1 was not conducted during the scope of this work, but will be essential for future studies.

2.1.3 Expression of GATA1 Isoforms in Human Hematopoiesis

GATA1 is expressed in a number of hematopoietic lineages (Shivdasani et al. 1997; Cantor et al. 2008; Nei et al. 2013), but the expression of GATA1 protein isoforms has not been thoroughly investigated in human hematopoiesis. The observation that DBA may be caused by mutations in the gene *GATA1*, which abrogate full length production, points to a specific role of GATA1 FL in erythropoiesis. After sorting different adult human bone marrow populations as outlined, GATA1 FL appears to be predominantly expressed in the erythroid $CD71^+CD235a^+$ compartment (Supplementary Figure 28). The erythroid cell line K562 was used as a positive control. $CD41^+$ megakaryocytes are known to express GATA1 and show a band (arrow) distinct from GATA1 FL and short (arrowheads). It remains unclear whether this reflects a posttranslational modification of GATA1 short. These results indicate high expression of GATA1 FL in erythroid cells and point to a specific function of this protein isoform in human erythropoiesis.

2.1.4 Modulation of mTOR signaling has no effect on GATA1 expression

Leucine has been implicated to alleviate the phenotype in DBA and 5q- syndrome and has been shown to act through the mTOR pathway (Pospisilova et al. 2007; Boultwood et al. 2013; Yip et al. 2013; Narla et al. 2014). To investigate a possible role of mTOR with respect to the expression of GATA1 K562 erythroid cells were treated with Torin1, a selective inhibitor of mTOR (Thoreen et al. 2009; Thoreen et al. 2012). As expected, Torin1 treatment significantly reduced cell proliferation (Thoreen et al. 2009) but in contrast to 4EGI-1 did not appear to affect GATA1 protein levels after 48h or 72h (Supplementary Figure 29, Figure 3d). Torin1 is known to cause severe defects in translation (Thoreen et al. 2012), but its consequences with respect to its regulated targets may be distinct from 4EGI-1 treatment and will require future studies to evaluate a role of mTOR in ribosomopathies with respect to GATA1 and DBA.

Supplementary Figure legends

Supplementary Figure 21. Knockdown of *GATA1* impairs erythropoiesis. **(a)** Representative flow cytometry plots showing impaired erythroid differentiation after shRNA-mediated knockdown of *GATA1* in primary human cells. Numbers in each quadrant represent the percentage of cells in each subpopulation. Staining for erythroid markers CD71, CD235a, and non-erythroid markers CD11b and CD41a, was performed as indicated to generate the plots shown. **(b)** Experimental setup showing the erythroid progenitor cells sorted (CD71^{high}/CD235a$^-$/CD11b$^-$/CD41a$^-$) for each sample used for RNA extraction and microarray analysis. **(c)** Relative *GATA1* mRNA levels shown for each sample and replicate. **(d)** Scatter plot of mean gene expression values in shluc control, sh20 and sh21 targeting *GATA1* of sorted primary erythroid progenitor samples (n = 3 for each sample). Pearson correlations are used to compare overall expression similarities between conditions.

Supplementary Figure 22. *In vitro* culture of bone marrow mononuclear derived cells from two healthy donors and one patient with Diamond-Blackfan anemia reveal defects in erythroid differentiation. Flow cytometry plots for CD71 and CD235a are shown at day 7 of differentiation.

Supplementary Figure 23. Differentially expressed genes upon knockdown of *GATA1*. **(a)** Heat map showing genes differentially up- or downregulated by $\log_2 > 1$ after shRNA treatment in shluc control vs. combined *GATA1* knockdown samples. **(b)** Enrichment profiles from gene set enrichment analysis (GSEA) comparing the relative expression of genes in shluc control versus *GATA1* knockdown samples and examining the distribution of both curated (top) or experimentally derived (bottom) GATA1 target genes in these datasets (shown as black bars). *P* values shown were calculated by examining the relative enrichment of the GATA1 gene sets compared to all genes in this data using a modified Kolmogorov-Smirnov test as implemented in GSEA. The most differentially expressed genes and their expression in each replicate are shown to the right for each gene set enrichment analysis. NES, normalized enrichment score.

Supplementary Figure 24. Genes regulated by GATA1 superenhancers are more sensitive to GATA1 downregulation. **(a)** Distribution of GATA1 chromatin-immunoprecipitation sequencing read occupancy intensity across regulatory elements is shown. Occupancy is not evenly distributed, with a subset of enhancers showing exceptionally high GATA1 binding. **(b)** Chromatin-immunoprecipitation sequencing profiles for two superenhancers in the *SLC4A1* and *ALAS2* locus is shown**. (c)** Box plots showing mean expression values for genes in the shluc control not regulated by GATA1 (no binding) and genes regulated by traditional (TE) and super- enhancers (SE) bound by GATA1. **(d)** Box plots showing changes in differential expression of genes in shluc control vs. *GATA1* knockdown samples for genes not regulated by GATA1 (no binding) and genes regulated by TE and SE bound by GATA1. Genes regulated by SE are more sensitive than other genes to downregulation of *GATA1* as indicated by larger changes in gene expression as indicated by the red arrow. **(e)** Enrichment profiles from GSEA comparing the relative expression of genes in shluc control versus *GATA1* knockdown samples amongst the top 100 up- or down- regulated genes in DBA patient vs. healthy control cells. *P* values shown were calculated by examining the relative enrichment of the GATA1 gene sets compared to all genes in this data using a modified Kolmogorov-Smirnov test as implemented in GSEA. NES, normalized enrichment score. **(f)** The proportion of genes bound by GATA1 superenhancers is shown for all genes and downregulated genes in DBA. A larger proportion of the downregulated genes in DBA are regulated by GATA1 superenhancers, suggesting a higher sensitivity to depletion of GATA1 protein.

Supplementary Figure 25. Evaluation of a GATA1 monoclonal antibody for intracellular flow cytometry. **(a)** Scheme illustrating *GATA1* mRNA and GATA1 FL and GATA1s protein. Numbers indicate exons. GATA1s lacks exon 2 encoding for the N-terminal transactivation domain (TD). **(b)** Western blot showing GATA1 expression in control and knockdown samples. RPS19 and GAPDH were used as loading controls. Multiple bands can be observed using this polyclonal antibody. GATA1 FL and GATA1s are indicated by arrowheads. **(c)** Western blot showing GATA1 expression in control and knockdown samples. RPS19 and ACTB were used as loading controls. Only GATA1 FL can be detected, as the monoclonal antibody only recognizes the N-terminus of GATA1. Samples shown in (b) and (c) are from independent experiments.

Supplementary Figure 26. Intracellular staining for GATA1 in erythroid cells. **(a)** Simplified experimental workflow for intracellular staining of GATA1 in erythroid cells. For permeabilization a saponin-based buffer was used from Invitrogen (C10418). ab125938 was used as an isotype control and a fluorescein isothiocyanate coupled goat anti-rabbit IgG was used (Jackson). **(b)** Histogram displaying GATA1 antibody and isotype control staining in erythroid K562 cells. **(c)** Flow cytometry plot showing distribution of CD71 and CD235a staining in primary erythroid cells (top). Gates I-IV indicate populations in which GATA1 expression levels based on intracellular staining were examined. Histogram plots below show expression of GATA1 within each gate and transition as indicated.

Supplementary Figure 27. Expression of GATA1 as measured by intracellular flow cytometry appears variable in indicated populations for a number of healthy donors.

Supplementary Figure 28. Expression of GATA1 protein isoforms in adult human bone marrow populations. **(a)** Scheme showing the sorted cell populations and their surface marker phenotype. **(b)** Flow cytometry plots showing sorted cell populations as indicated by gates and based on surface marker expression of CD41a, CD71 and CD235a. **(c)** Western blot showing GATA1 expression in sorted populations as indicated. Arrowheads point to GATA1 FL and GATA1s. The Arrow indicates a potentially modified form of GATA1s in CD41$^+$ megakaryocytes. RPS19 and Actin were used as loading controls. K562 cells were used to compare GATA1 expression levels. CD34$^+$ stem and progenitor cells were obtained from an independent healthy donor.

Supplementary Figure 29. Torin1 treatment has no effect on GATA1 expression levels. **(a)** Cell counts for K562 cells treated with Torin1 over 72h at indicated concentrations. **(b)** Expression of GATA1 after 48h to 72h of Torin1 treatment as examined by western blot. RPS19 and GAPDH were used as loading controls.

Supplementary Figure 21

a

b

Sort erythroid progenitor cells
shLuc
GATA1 sh20
GATA1 sh21

↓

RNA

↓

Microarray Analysis
(3 replicates for each sample)

c

d
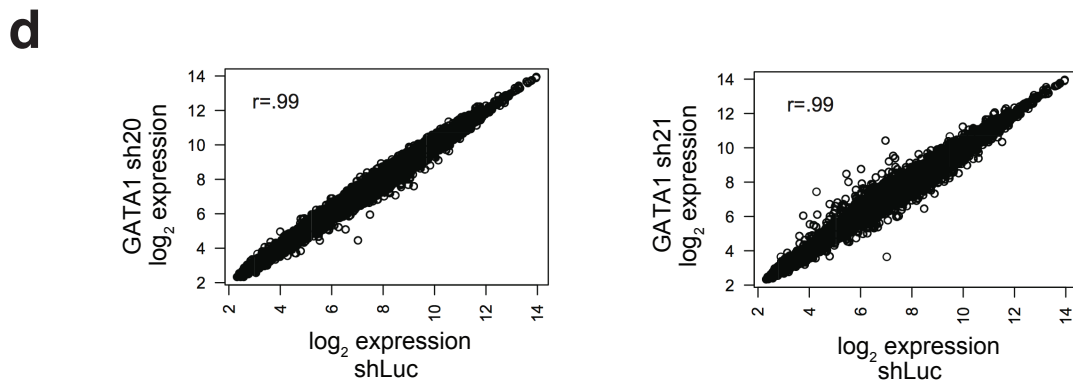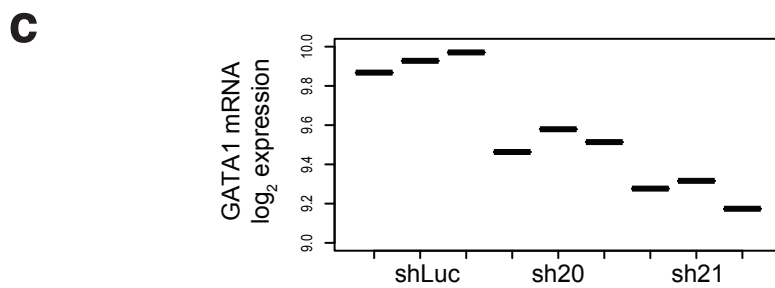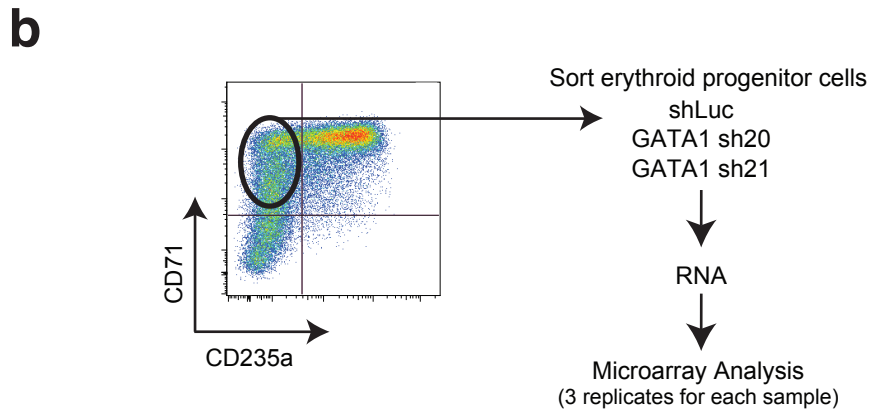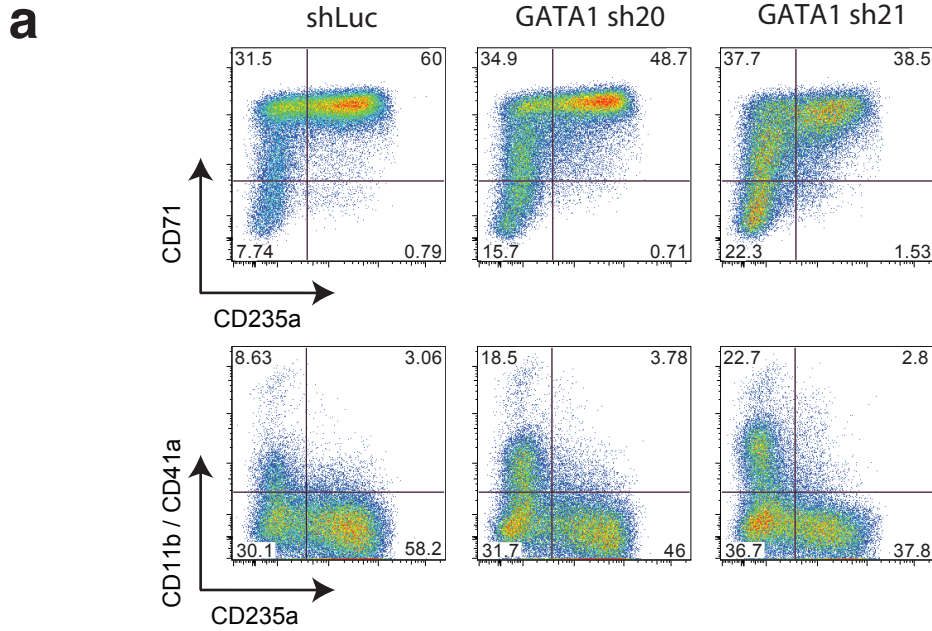
Supplementary Figure 22

Day 7 of differentiation
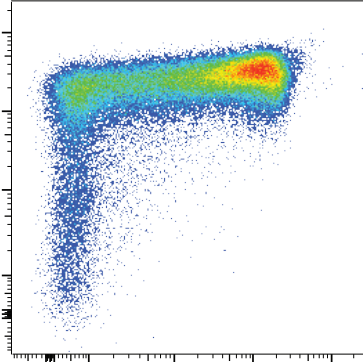
Healthy donor #1
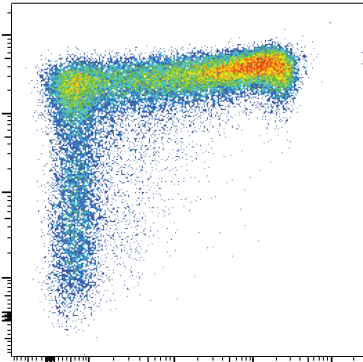
Healthy donor #2
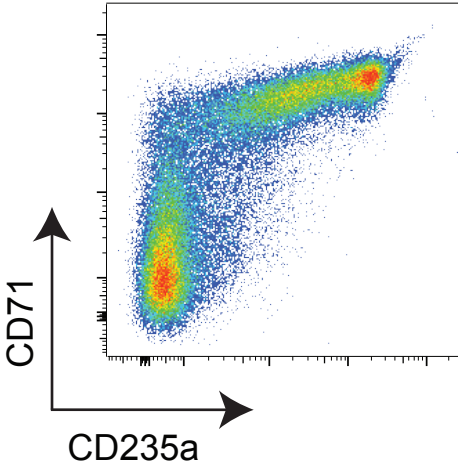
Diamond-Blackfan anemia

CD71

CD235a

Supplementary Figure 23

a

genes included if abs log$_2$ differential expression > 1

b

a

b

Super regulatory element (~29 kilobases)

SLC4A1

Super regulatory element (~17 kilobases)

ALAS2

c

d

e

sh20 top 100 down-regulated genes

NES = -1.68
p < 0.001

Up in DBA   Down in DBA

sh20 top 100 up-regulated genes

NES = 1.55
p = 0.01

Up in DBA   Down in DBA
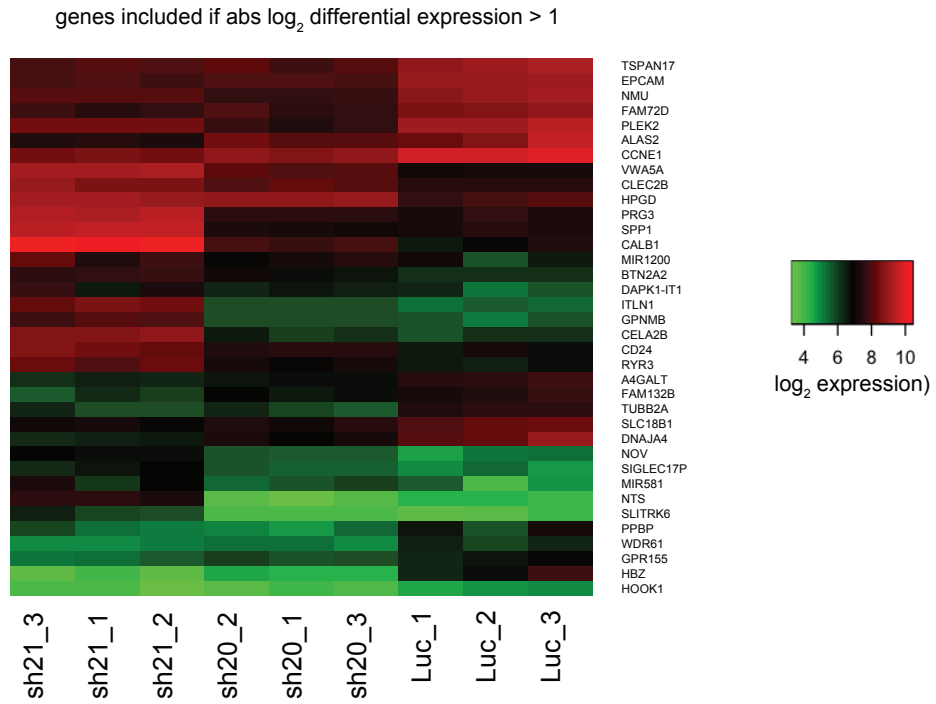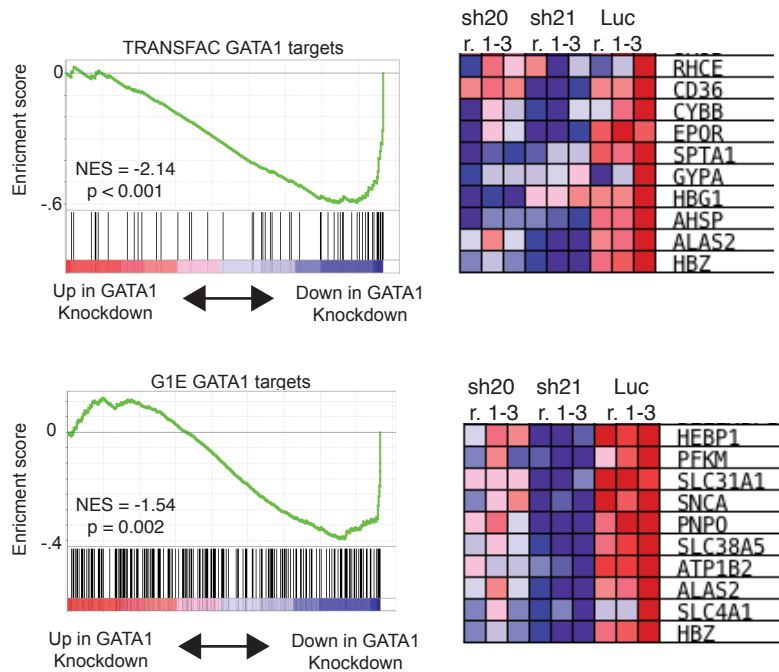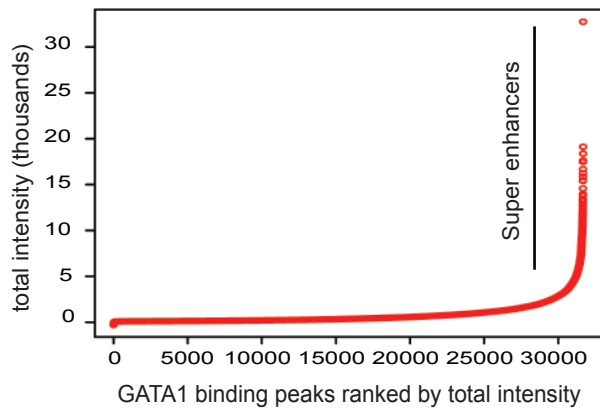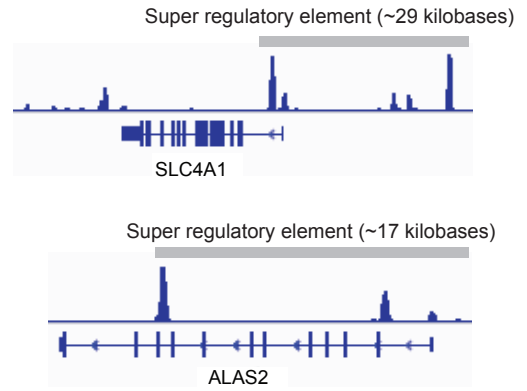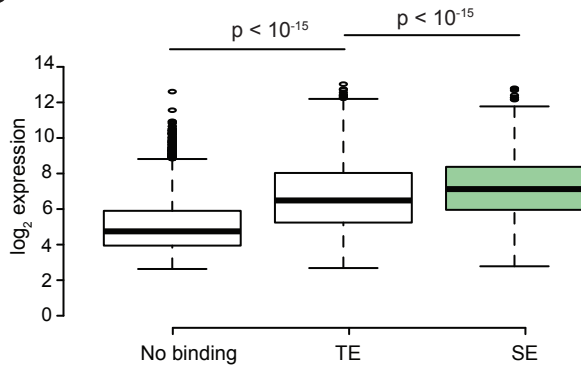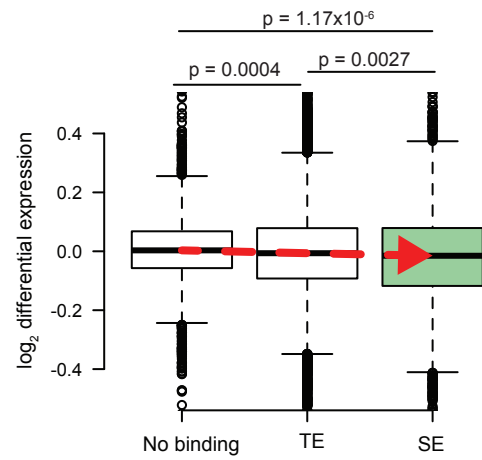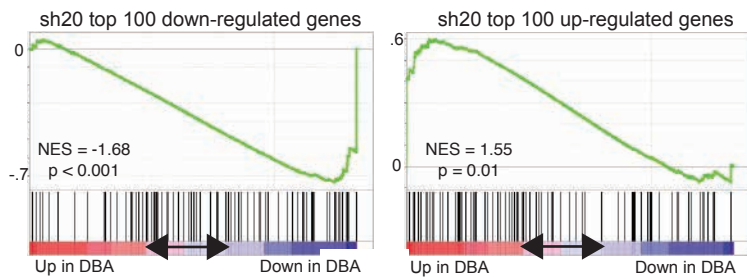
f
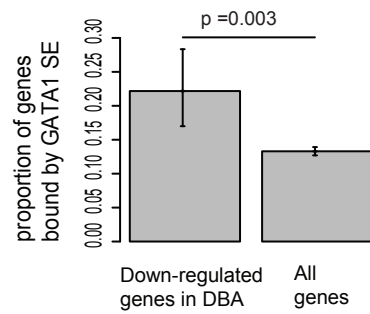
Supplementary Figure 25

a
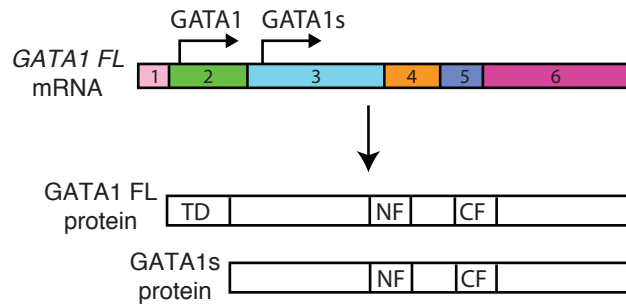
b

C-terminal polyclonal antibody recognizing GATA1 FL and GATA1 short
(M-20, sc-1234, Santa Cruz Biotechnology)
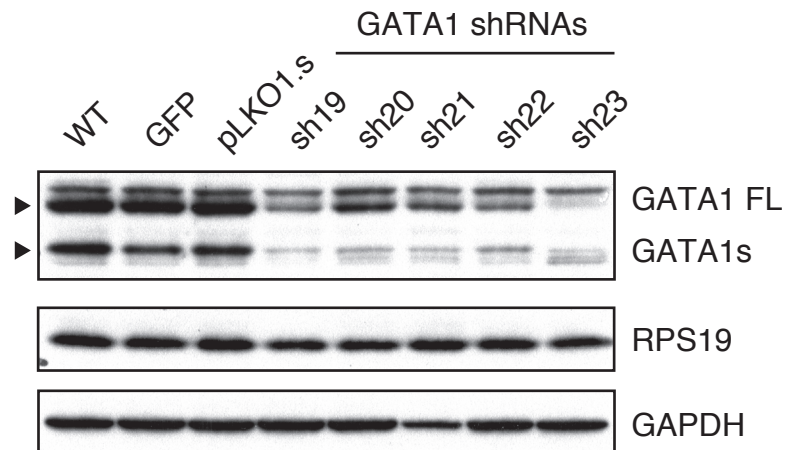
c

N-terminal monoclonal antibody recognizing GATA1 FL only
(EP2819Y, ab76121, Abcam)

Supplementary Figure 26

a

Surface staining

↓ Wash

Fixation 4% paraformaldehyde

↓ Wash

Saponin permeabilization

↓

Incubation anti-GATA1 or Isotype control

↓ Wash

Secondary antibody

↓ Wash

Flow cytometry analysis

b

K562 cells

c

Primary cells

Erythroid differentiation I ⟶ IV

GATA1 signal intensity

Supplementary Figure 27

Supplementary Figure 28

a

CD34+
Stem and Progenitor cells

CD71+ /CD235a+
Erythroblasts

CD41+
Megakaryocytes

CD71- /CD235a-/CD41-
Bone marrow cells

b

c

GATA1 FL

GATA1s

RPS19

β-Actin

Bone marrow

a

b

3. Manuscript II ― Cyclin D3 coordinates the cell cycle during differentiation to regulate erythrocyte size and number

Vijay G. Sankaran[1], Leif S. Ludwig[1], Ewa Sicinska[2], Jian Xu[2], Daniel E. Bauer[2], Jennifer C. Eng, Heide Christine Patterson, Ryan A. Metcalf, Yasodha Natkunam, Stuart H. Orkin, Piotr Sicinski, Eric S. Lander[3] and Harvey F. Lodish[3]

1 These authors contributed equally to this work.
2 These authors contributed equally to this work.
3 These authors contributed equally to this work.

Experimental contribution

Vijay Sankaran developed the initial idea and conducted genetic analysis. Vijay Sankaran and I developed, planned and conducted most experiments with assistance from Jennifer Eng. Ewa Sicinska and Piotr Sicinski conducted *in vivo* analysis of cyclin D knockout mice. Jian Xu and Daniel Bauer performed chromatin conformation capture and chromatin immunoprecipitation from donors heterozygous for rs9349205, respectively. Heide Patterson assessed histological specimens. Ryan Metcalf and Yasodha Natkunam provided and assessed cyclin D3 stained human bone marrow specimens. Stuart Orkin, Piotr Sicinski, Eric Lander and Harvey Lodish provided intellectual contributions to the project. Vijay Sankaran wrote the manuscript with assistance from myself, Eric Lander, Harvey Lodish and with input from all authors.

The following figures present my work:
Figures 1C-D, 2G-H, 3A-B, 4A-D.
Supplementary Figures 7, 8, 9.

The original article including supplemental information is included on the following pages and available online:
http://dx.doi.org/10.1101/gad.197020.112
http://genesdev.cshlp.org/content/26/18/2075.full

Cyclin D3 coordinates the cell cycle during differentiation to regulate erythrocyte size and number

Vijay G. Sankaran,[1,2,3,4,17] Leif S. Ludwig,[2,5,6,17] Ewa Sicinska,[7,18] Jian Xu,[8,9,18] Daniel E. Bauer,[8,9,18] Jennifer C. Eng,[1,2] Heide Christine Patterson,[2,10,11] Ryan A. Metcalf,[12] Yasodha Natkunam,[12] Stuart H. Orkin,[1,4,8,9] Piotr Sicinski,[13,14] Eric S. Lander,[1,15,16,19,20] and Harvey F. Lodish[1,2,16,19,20]

[1]Broad Institute, Cambridge, Massachusetts 02142, USA; [2]Whitehead Institute for Biomedical Research, Cambridge, Massachusetts 02142, USA; [3]Department of Medicine, Children's Hospital Boston, Boston, Massachusetts 02115, USA; [4]Department of Pediatrics, Harvard Medical School, Boston, Massachusetts 02115, USA; [5]Institute for Chemistry and Biochemistry, Freie Universität Berlin, 14195 Berlin, Germany; [6]Charité-Universitätsmedizin Berlin, Berlin 10117, Germany; [7]Center for Molecular Oncologic Pathology, Dana-Farber Cancer Institute, Boston, Massachusetts 02215, USA; [8]Division of Hematology/Oncology, Children's Hospital Boston, Boston, Massachusetts 02115, USA; [9]Department of Pediatric Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts, USA; [10]Institut für Klinische Chemie und Pathobiochemie, Klinikum rechts der Isar, Technische Universität München, 81675 Munich, Germany; [11]Department of Pathology, Brigham and Women's Hospital, Boston, Massachusetts 02115, USA; [12]Department of Pathology, Stanford University School of Medicine, Stanford, California 94304, USA; [13]Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, Massachusetts 02215, USA. [14]Department of Genetics, [15]Department of Systems Biology, Harvard Medical School, Boston, Massachusetts 02115, USA. [16]Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

Genome-wide association studies (GWASs) have identified a genetic variant of moderate effect size at 6p21.1 associated with erythrocyte traits in humans. We show that this variant affects an erythroid-specific enhancer of *CCND3*. A *Ccnd3* knockout mouse phenocopies these erythroid phenotypes, with a dramatic increase in erythrocyte size and a concomitant decrease in erythrocyte number. By examining human and mouse primary erythroid cells, we demonstrate that the *CCND3* gene product cyclin D3 regulates the number of cell divisions that erythroid precursors undergo during terminal differentiation, thereby controlling erythrocyte size and number. We illustrate how cell type-specific specialization can occur for general cell cycle components—a finding resulting from the biological follow-up of unbiased human genetic studies.

Genome-wide association studies (GWASs) have identified hundreds of loci associated with numerous phenotypes and diseases in humans (Manolio et al. 2008; Lander 2011). The description of these loci holds promise for the identification of important biological pathways and disease mechanisms. However, while there are notable exceptions (Sankaran et al. 2008a; Musunuru et al. 2010; Zhang et al. 2011), the majority of such molecular

pathways have yet to be elucidated (Lander 2011). Disorders affecting red blood cells (RBCs) or erythrocytes are among the most common genetic conditions in humans (Balarajan et al. 2011). This includes mutations affecting the globin genes and other erythrocyte components, such as the enzyme glucose-6-phosphate dehydrogenase. Anemia is estimated to affect nearly a quarter of the world's population, much of which is attributable to defective erythrocyte production (Balarajan et al. 2011). An increased understanding of the process through which erythrocytes are produced— erythropoiesis—is important in gaining insight into the pathophysiology of these diseases and may lead to avenues for rationally intervening in these conditions.

A number of recently published GWASs have uncovered variants that affect several erythrocyte traits (Ferreira

et al. 2009; Ganesh et al. 2009; Soranzo et al. 2009; Kamatani et al. 2010). Principal among these has been the number of RBCs (RBC count) and the size of these cells (mean corpuscular volume [MCV]), which show natural variation in humans and are both heritable traits. Moreover, variation in these traits is also characteristic of specific pathologic conditions affecting erythropoiesis. For example, the megaloblastic anemias and certain congenital anemias are characterized by a markedly elevated MCV. In addition, evolutionary biologists have noted for centuries that the two traits show extensive inverse variation across species (Boylan et al. 1991; Hawkey et al. 1991). The GWASs revealed a few loci that, in an analogous fashion, are inversely associated with RBC count and MCV in humans (Ganesh et al. 2009; Soranzo et al. 2009). One such locus is within the gene CCND3.

The CCND3 gene encodes cyclin D3, one of three well-characterized members of the D cyclin family that play critical roles as core members of the mammalian cell cycle machinery (Sicinska et al. 2003; Sherr and Roberts 2004; Malumbres and Barbacid 2009). Once induced, D cyclins bind to and activate the cyclin-dependent kinases (CDKs) 4 and 6, which can then phosphorylate the retinoblastoma tumor suppressor gene product (RB1) and RB1-related p107 and p130 proteins (Malumbres and Barbacid 2009). This phosphorylation in turn leads to release or derepression of E2F transcription factors that promote progression from the G1 to S phase of the cell cycle (Sherr and Roberts 2004; Malumbres and Barbacid 2009). Among the D cyclins, cyclin D3 is one of the least well studied. It has been suggested to have a role in certain cancers, particularly those of hematopoietic origin (Sicinska et al. 2003; Metcalf et al. 2010). Indeed, rearrangements involving the CCND3 gene have been seen in cases of multiple myeloma and diffuse large cell lymphoma. Mice with a knockout of the Ccnd3 gene have defects in T-lymphocyte development (Sicinska et al. 2003), B-lymphocyte maturation (Cooper et al. 2006), and neutrophil production (Sicinska et al. 2006). Although $Ccnd2^{-/-}$; $Ccnd3^{-/-}$ mice have been noted to be severely anemic and die in mid-gestation (Ciemerych et al. 2002), the role of cyclin D3 in erythropoiesis has not been previously explored.

Here we used insight from human genetics to study a variant near the CCND3 gene that shows an association of moderate effect size with both MCV and RBC count. We show that this variant affects an erythroid-specific enhancer element that regulates CCND3 expression; individuals with the weaker enhancer variant have larger but fewer red cells in their blood. We then used mouse genetics to show that the loss of Ccnd3 results in greatly enlarged erythrocyte size with a concomitant decrease in RBC count, consistent with the findings from the human genetic association studies. We explore the mechanism underlying this phenotype in mouse and human primary cells and show that cyclin D3 plays a critical role in regulating the number of cell divisions that erythroid precursors will undergo before production of a mature erythrocyte, thus controlling erythrocyte size and num-

ber. This unexpected link between the cell cycle, erythroid differentiation, and the regulation of erythrocyte production was only uncovered by pursuing the biology underlying the results from unbiased studies of human genetic variation.

Results

A variant near CCND3 affects an erythroid-specific enhancer element

In GWASs from the CHARGE and HaemGen consortia, the single-nucleotide polymorphism (SNP) most significantly associated with erythrocyte traits, including MCV and RBC count, was rs9349205 in the CCND3 locus at 6p21.1 ($P = 1.1 \times 10^{-31}$ for MCV) (Ganesh et al. 2009). Imputation of SNPs in a Japanese cohort study demonstrated that this same SNP was also the most significantly associated with MCV and RBC count in this population ($P = 6.2 \times 10^{-29}$ for MCV) (Kamatani et al. 2010). Given the significant associations seen with this SNP, we examined all SNPs in Caucasian (CEU) and Asian (CHB and JPT) populations from the 1000 Genomes Project Consortium (2010) that were in close linkage disequilibrium ($r^2 > 0.8$) with rs9349205 (Supplemental Fig. 1; Supplemental Tables 1, 2). We then intersected these two sets of SNPs and found that only four SNPs—rs9349205, rs9394841, rs1410492, and rs3218097—were in close linkage disequilibrium in both populations examined (Supplemental Table 3). Of these four SNPs, it appeared unlikely that rs3218097 or rs1410492 were the causal variants at this locus because rs3218097 is less significantly associated with the erythroid traits than rs9349205 (Kamatani et al. 2010), and rs1410492 is in complete linkage disequilibrium with rs3218097 in both the CEU and CHB/JPT populations. This left the SNPs rs9349205 and rs9394841 as potential causal variants. There may be untyped common variants that were missed in our analysis, but given the coverage of common variants in the 1000 Genomes Project Consortium (2010), we believe that we were relatively comprehensive in evaluating the majority of common genetic variation at this locus (our analysis should provide complete coverage of all variants present in at least 10 of the 120 samples from the CEU population).

To explore potential functional roles of these SNPs, we studied chromatin immunoprecipitation (ChIP) data from human adult primary erythroid cells. We examined the presence of histone 3 Lys 4 monomethylation (H3K4me1), a well-characterized marker of enhancer elements; histone 3 Lys 4 trimethylation (H3K4me3), a marker of active promoter regions; and histone 3 Lys 9 acetylation (H3K9ac), which marks active chromatin (Heintzman et al. 2007); along with the erythroid transcription factors GATA1, TAL1 (also known as SCL), and KLF1 that, in combination, delineate active chromatin and enhancer states (Cheng et al. 2009; Tallack et al. 2010; Wu et al. 2011). In the region surrounding rs9349205, we noticed that there was significant enrichment of the H3K4me1 mark as well as the presence of a site of erythroid-specific
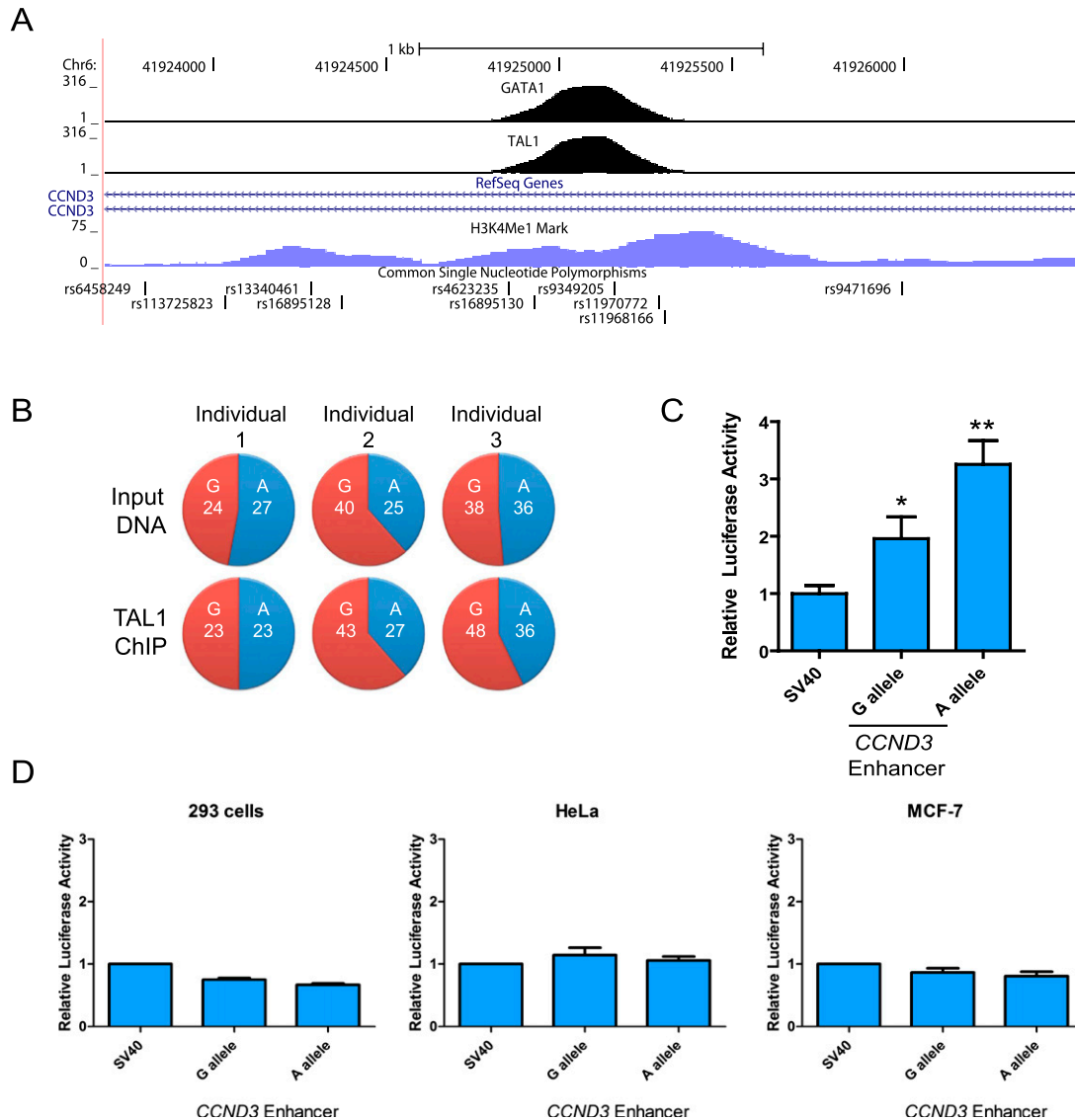
Figure 1. GWASs reveal a genetic variant affecting an erythroid-specific enhancer. (*A*) ChIP-seq (ChIP coupled with deep sequencing) data for GATA1, TAL1, and H3K4me1 are shown in the region surrounding rs9349205, along with other common SNPs in this region. Details of this data are described in the Materials and Methods section. (*B*) Allele-specific TAL1 ChIP shows no significant differences ($P > 0.1$ in all cases) for the number of clones with the G versus A allele at rs9349205. ChIP-PCR products were cloned and then sequenced. The number of clones from input DNA samples and TAL1 ChIP samples are shown in the *top* and *bottom* rows of the panel from three separate heterozygous individuals. (*C*) Results of a luciferase assay using the pGL3-SV40 promoter vector in the absence (indicated by the first bar) or presence of the erythroid-specific 280-bp enhancer region surrounding rs9349205. (Last two bars) The two alleles at rs9349205 were assayed in this experiment. The results are shown as the mean ± the standard error ($n \geq 3$ per group) and are normalized to *Renilla* luciferase activity, and the activity of the SV40 promoter construct alone was normalized to a value of 1. (*) $P < 0.05$; (**) $P < 0.01$. (*D*) The erythroid-specific enhancer region surrounding rs9349205 shows no enhancer activity in human nonerythroid 293, HeLa, and MCF-7 cells. Results of a luciferase assay using the pGL3-SV40 promoter vector in the absence (indicated by the first bar) or presence of the erythroid-specific 280-bp enhancer region surrounding rs9349205. (Last two bars) The two alleles at rs9349205 were assayed in this experiment. The results are shown as the mean ± the standard error ($n = 3$ per group) and are normalized to *Renilla* luciferase activity, and the activity of the SV40 promoter construct alone was normalized to a value of 1.

DNase I hypersensitivity, suggesting that this region was a canonical enhancer element (Fig. 1A; Supplemental Figs. 2A, 3A). This enhancer region was found ~15 kb upstream of the promoter of *CCND3*, where a variety of regulatory elements of this gene lie. In addition, the erythroid transcription factors TAL1, GATA1, and KLF1

all occupied chromatin in a region surrounding variant rs9349205 in independent ChIP data sets (Fig. 1A; Supplemental Fig. 3A). We noted that rs9349205 was located directly in the middle of the peak of TAL1 and GATA1 chromatin occupancy and was flanked by KLF1-binding peaks. We noted that no such binding sites or signatures

Sankaran et al.

of regulatory elements were found in the region surrounding variant rs9394841. When we examined the chromatin state in this region from nine human cell lines that were studied as part of the ENCODE project (Ernst et al. 2011), only the erythroid cell line K562 showed signs of having a strong enhancer within this region, whereas the other cell lines did not (Supplemental Fig. 2C). This suggests that this region is an erythroid-specific enhancer element that only shows the appropriate modifications in erythroid lineage cells but not in other cell types.

We explored whether the SNP rs9349205 might directly affect binding by one of the erythroid transcription factors examined above (Fig. 1A; Supplemental Fig. 3A). We focused on GATA1 and TAL1, since the ChIP data suggested that they were bound with their peak centered on rs9349205. We found no consensus transcription factor-binding motifs within the sequence immediately around this SNP (Supplemental Fig. 2B), although a potential noncanonical binding site for TAL1 may exist in the region and would support the finding of a ChIP peak centered at this variant (Fig. 1A; Supplemental Fig. 3A). As a result, we assayed TAL1 binding to chromatin in primary erythroid progenitors cultured from separate individuals heterozygous for the variant at rs9349205 but found no evidence of allele-specific binding in the region containing this SNP (Fig. 1B).

We next sought to address whether this element could functionally act as an enhancer in erythroid cells. A fragment of this region was cloned upstream of a ubiquitous mammalian promoter (SV40) and luciferase construct. When we introduced the 280-base-pair (bp) core region surrounding variant rs9349205, we were able to demonstrate enhancer activity in the K562 erythroid cell line (Fig. 1C). Moreover, when the minor A allele at rs9349205 was introduced into this construct, further increases in activity were noted compared with the G allele (Fig. 1C). This finding supported the hypothesis that the major allele at rs9349205 (G allele), associated with increased erythrocyte size and reduced RBC count, resulted in decreased enhancer activity upstream of the CCND3 gene. Using identical enhancer element assays in several other human nonerythroid cell types, we did not observe any signs of enhancer activity or of any allele-specific differences; these results suggest that rs9349205 affects an erythroid-specific enhancer (Fig. 1D). As noted above, the effect does not appear to be due to differential binding of the TAL1 transcription factor to the sequence surrounding the SNP; the exact mechanisms by which such differential enhancer activity occurs remain to be uncovered. This finding is reminiscent of a well-characterized human mutation in a GATA1-binding site that affects transcriptional activity without affecting GATA1 binding (Martin et al. 1989). We obtained additional evidence supporting the notion that this enhancer element acts specifically on the CCND3 gene; the chromatin conformation capture (3C) assay demonstrated an increased frequency of interaction of this enhancer element—in comparison with other surrounding regions—with the proximal promoter of CCND3 in human adult erythroid progenitors (Supplemental Fig. 3B). Together, these findings support a role of the rs9349205 SNP—most significantly associated with erythroid traits in the 6p21 region—in altering the erythroid-specific activity of an enhancer element upstream of CCND3.

Given the findings of the erythroid-specific enhancer element of CCND3, we examined expression of cyclin D3 in samples from in vitro differentiating human adult erythroid progenitors and sections of human bone marrow (Supplemental Fig. 4). We noticed that cyclin D3 was well expressed in early erythroid progenitors, with maximal expression in cells at the proerythroblast stage of differentiation, and underwent a decrease in expression as the cells underwent subsequent terminal differentiation, consistent with findings from prior studies (Dai et al. 2000; Metcalf et al. 2010). The decrease in cyclin D3 expression occurred concomitantly with the period when cell cycle exit occurs during terminal erythropoiesis, suggesting a potential role for this protein in the regulation of this process.

Erythropoiesis in Ccnd3 knockout mice

To study the in vivo role of Ccnd3 in erythropoiesis, we examined a germline knockout mouse of this gene (Sicinska et al. 2003, 2006; Cooper et al. 2006). A mild anemia was present in the adult homozygous knockout mice, with an average 13% reduction in the hematocrit compared with wild-type animals (Fig. 2A,B). Surprisingly, we noticed a 38% reduction in RBC count in the mutant animals (Fig. 2C). Along with this decrease in RBC count, these mice had a dramatic 40% increase in erythrocyte size (MCV) (Fig. 2D,G,H). To the best of our knowledge, these mice have the largest adult mouse erythrocytes that have been described to date (Fig. 2G,H). In fact, this RBC size approaches that seen in normal humans. The erythrocytes in these mice were relatively well hemoglobinized, albeit with a slight decrease in hemoglobin concentration (MCHC [mean corpuscular hemoglobin concentration]) compared to wild-type mice (Fig. 2E). The larger cells have more hemoglobin, and thus fewer RBCs can provide a sufficient total amount of hemoglobin so that little change in the oxygen-carrying capacity of these animals would occur. The mice additionally did not show signs of having defective erythropoiesis, as the reticulocyte count as a percent was not significantly different between the groups of mice (Fig. 2F), and when corrected for the RBC count, essentially no difference was observed (an average of 493×10^3 reticulocytes per microliter in controls vs. 488×10^3 reticulocytes per microliter in knockout animals). Other RBC parameters varied concordant with these parameters, and the platelet count showed no change between wild-type and mutant mice (Supplemental Fig. 5A). Of note, the human genetic variant associated with decreased enhancer function (the G allele of rs9349205) is associated with a larger MCV and reduced RBC count, entirely consistent with the phenotype observed in the Ccnd3 knockout animal.

Since erythropoietin (Epo) is a major mitogen and survival factor in erythroid progenitors and D cyclins are often induced by such pathways, we stimulated adult mice with pharmacological doses of Epo. The $Ccnd3^{-/-}$
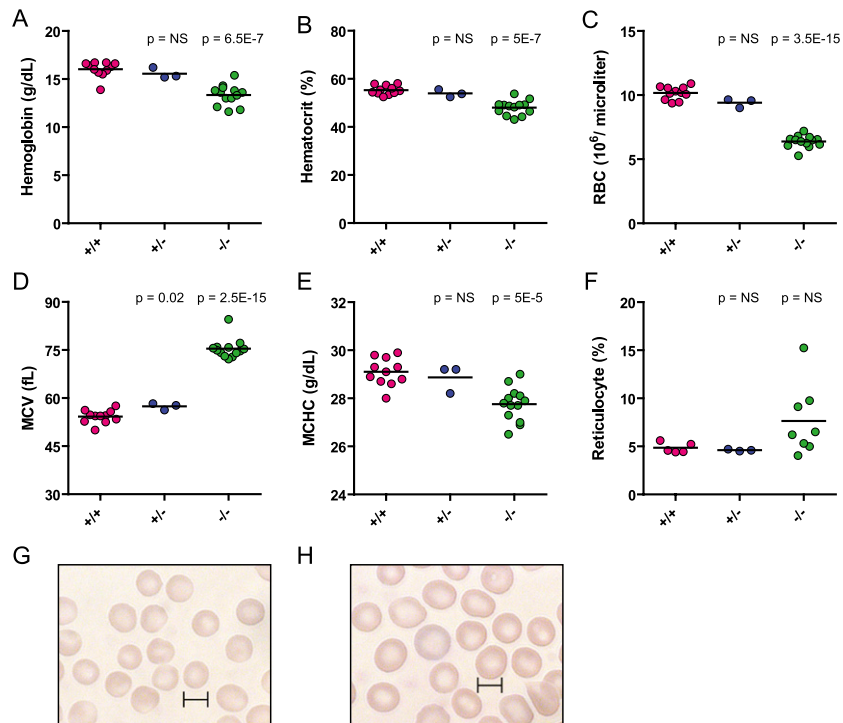
Figure 2. Hematological analysis of *Ccnd3* knockout mice reveals a dramatic increase in erythrocyte size with a concomitant decrease in RBC number. (*A–F*) The hematological values from complete blood counts on mice of various genotypes are shown. The genotype +/+ indicates wild-type littermate controls, +/− indicates heterozygous animals, and −/− indicates *Ccnd3* knockout animals. The data are shown as the mean, with the distribution of values for 11 +/+, three +/−, and 13 −/− animals, except for the reticulocyte percentage, which was done with five +/+, three +/−, and eight −/− animals. *P*-values are shown *above* the corresponding data and are based on a comparison with +/+ littermates. (*G,H*) Blood smears shown with identical (100×) magnification demonstrating the enlarged size and heterogeneity of erythrocytes from 8-wk-old *Ccnd3*⁻/⁻ mice (*H*) in contrast to age-matched littermate +/+ controls (*G*). Bars: *G,H*, 5 μm.

mice continued to respond appropriately to this stimulation despite the absence of cyclin D3, suggesting that cyclin D3 activity was independent of Epo (Supplemental Fig. 6A). Mice with a knockout of the two other D cyclins, *Ccnd1* and *Ccnd2*, were additionally noted to have ostensibly normal erythropoiesis, suggesting that only cyclin D3 had a critical role in terminal erythropoiesis (Supplemental Fig. 5B). This finding is consistent with the patterns of expression seen for these different D cyclins, with cyclin D3 being the predominant protein in terminal erythroid precursor cells (Ciemerych et al. 2002; Metcalf et al. 2010). When the bone marrow in adult mice was examined before and after stimulation with Epo, no major difference among progenitors at various stages of differentiation was noted (Supplemental Fig. 6B). Collectively, these results suggest that, overall, erythropoiesis proceeds in a relatively normal manner, although there is some perturbation in the differentiation process that causes fewer erythrocytes to be produced that are of a much larger size.

Effects of cyclin D3 perturbation on erythroid cells

To better understand the cellular mechanisms underlying the phenotype seen on loss of *Ccnd3* expression, we used in vitro cell culture of primary mouse fetal liver (FL) erythroid cells to examine the mechanistic basis for this alteration in erythrocyte size and number upon knockout or knockdown of *Ccnd3* (Zhang et al. 2003). We initially used FL cells derived from mice with a *Ccnd3* knockout allele, along with heterozygous and wild-type littermate controls. The cells were labeled with the plasma mem-

brane-labeling dye PKH26 in order to study the number of divisions that the cells undergo during differentiation. Normal differentiation proceeds over a period of 2–3 d in such a culture, beginning with undifferentiated erythroid progenitors (Ter119-negative cells from the FL), and generally involves an average of four to five cell divisions before terminal post-mitotic erythroid cells are produced (Zhang et al. 2003). We noticed that there was an average of 0.7 fewer cell divisions at 48 h in the knockout animals compared with wild-type controls ($P < 0.001$) (Fig. 3B). The cells from the knockout animals maintained brighter expression of PKH26 as a result of fewer cell divisions occurring between 24 and 48 h of culture, while maintaining a larger cell size at the 48-h time point, as assessed by forward scatter (Fig. 3A). No major difference was seen in phenotypic markers of erythroid cell differentiation among the knockout or control cultures (Supplemental Fig. 7).

To study the effect of acute knockdown of *Ccnd3* and avoid the potential compensatory responses seen in germline knockout cells, we cloned two shRNAs targeting *Ccnd3*, sh50 and sh79, in mouse retroviral vectors; these shRNAs reduced gene expression to ~15% and ~40% of control levels, respectively (Fig. 4A). We found that cells carrying sh50 or sh79 showed no significant difference in the number of divisions after 24 h, but showed significant differences ($P < 0.001$) after 48 h (Fig. 4B,C). There was an average of 0.9–1.0 fewer cell divisions with knockdown by the more robust sh50, and 0.2–0.3 fewer cell divisions with knockdown by sh79 (Fig. 4B). Concomitant with this reduction in the number of cell divisions, the cells were larger at 48 h by measurement of forward scatter (Fig. 4C).

Sankaran et al.

A

B
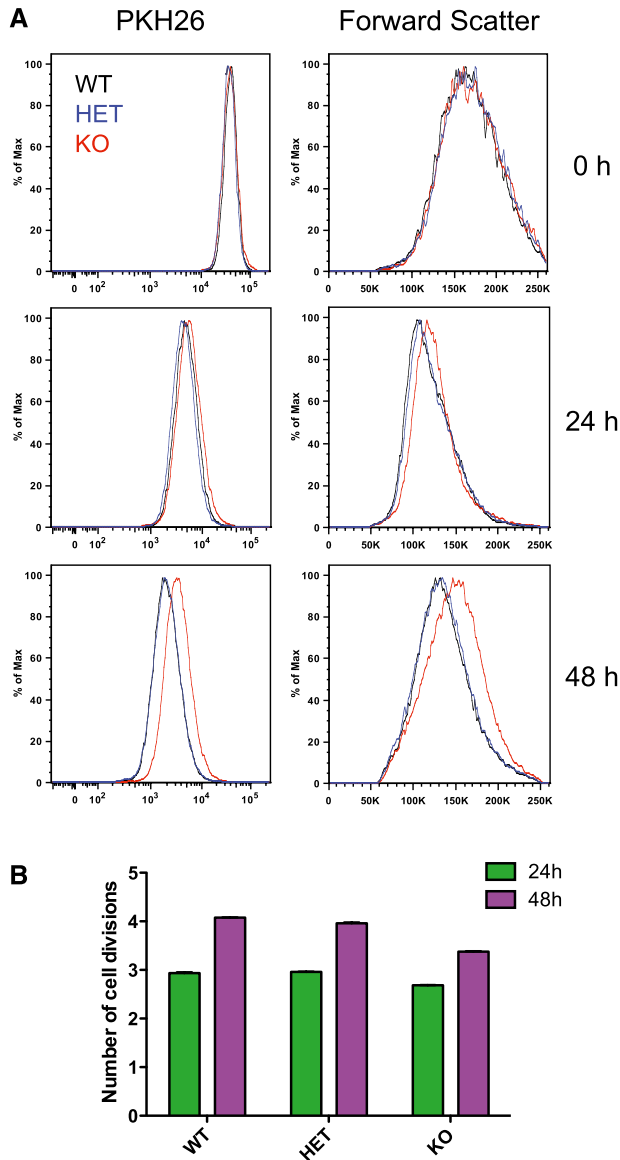
Figure 3. Cyclin D3 regulates the number of cell divisions and cell size during erythropoiesis. (A) PKH26 labeling and forward scatter are shown at 0, 24, and 48 h for knockout (KO), heterozygous (HET), or wild-type (WT) *Ccnd3* FL cells. (B) The average number of cell divisions calculated as discussed in the Materials and Methods from the mean fluorescence intensity measurements for PKH26-labeled FL cells of the various genotypes at 24 and 48 h.

A similar extent of increased size was seen in the subset of enucleated cells examined at 48 h, suggesting that this increased size is maintained in the most mature erythrocytes. Therefore, a reduction in the number of cell divisions occurs upon knockdown of cyclin D3 that results in the production of fewer but larger erythrocytes. The reduction in the number of cell divisions may be attributable to cells that are cycling slower and therefore undergo fewer divisions or may be due to cells that undergo premature cell cycle exit during differentiation.

The finding that only the terminal divisions were affected suggested that the former scenario was less likely to be the case. To directly test this, we labeled the cells at day 1 and day 2 of culture with bromodeoxyuridine (BrdU) and examined cell cycle progression in the cells (Fig. 4D). There was no major difference in the rate of cell cycle progression among the control and knockdown cells (Fig. 4D). Similar to the knockout analysis (Supplemental Fig. 7), we could not detect major differences in the expression of phenotypic markers of erythropoiesis with this knockdown, and enucleation continued to occur, suggesting that differentiation was proceeding in a relatively unperturbed manner. These results strongly suggest that cyclin D3 plays a critical role in regulating the number of cell divisions that occur during terminal erythropoiesis. To ascertain whether this erythroid phenotype was due to interaction between cyclin D3 and CDK4/6, we used the small molecule inhibitor of these kinases, PD0332991 (Toogood et al. 2005). Using a range of concentrations of this inhibitor, known to specifically act on CDK4 and CDK6 (Toogood et al. 2005), we observed identical phenotypes in terms of a reduction in the number of terminal cell divisions and increases in cell size at 48 h in the same FL cell culture system used in the previous experiments (Supplemental Fig. 8). We additionally studied erythroid cultures from adult bone marrow progenitors and obtained results similar to those discussed above. There was an average reduction of 0.6 cell divisions in cultures from the *Ccnd3* knockout mice compared with the controls, although this may be an underestimate of the actual effect on erythropoiesis given the heterogeneity of the progenitor cells cultured from this population (Supplemental Fig. 9). No major differences in cell differentiation, judged by induction of Ter119, or the rate of enucleation were seen when comparing cultures of bone marrow progenitors from *Ccnd3* knockout and control mice (Supplemental Fig. 9).

We used knockdown of cyclin D3 in adult human erythroid progenitor cell cultures to further study this phenomenon and discern whether the phenomena observed in the mice were also seen in humans (Sankaran et al. 2008a, 2011). We were able to achieve robust knockdown of cyclin D3 using a set of lentivirally expressed shRNAs (Fig. 5A; Moffat et al. 2006). Consistent with the results from the mouse primary cell culture, we noted that knockdown of *CCND3* resulted in a reduction in the number of cell divisions during terminal erythropoiesis, as assayed by PKH26 labeling (Fig. 5B). The distribution of PKH26 labeling was significantly different between the samples ($P < 0.001$ for all comparisons between knockdown and control samples). Consistent with the results in the mouse primary cell culture system, no differences in cell size or morphology were notable at the early stages of differentiation. However, we did note significant differences in the size of cells during the late stages of differentiation (Fig. 5C; Supplemental Fig. 10). These findings in the human cells—consistent with the data obtained from cultured mouse erythroid cells—reinforce the model that cyclin D3 levels dictate the number of cell divisions that will occur during terminal erythropoiesis, thus regulating the number and size of erythrocyte progeny (Fig. 6).
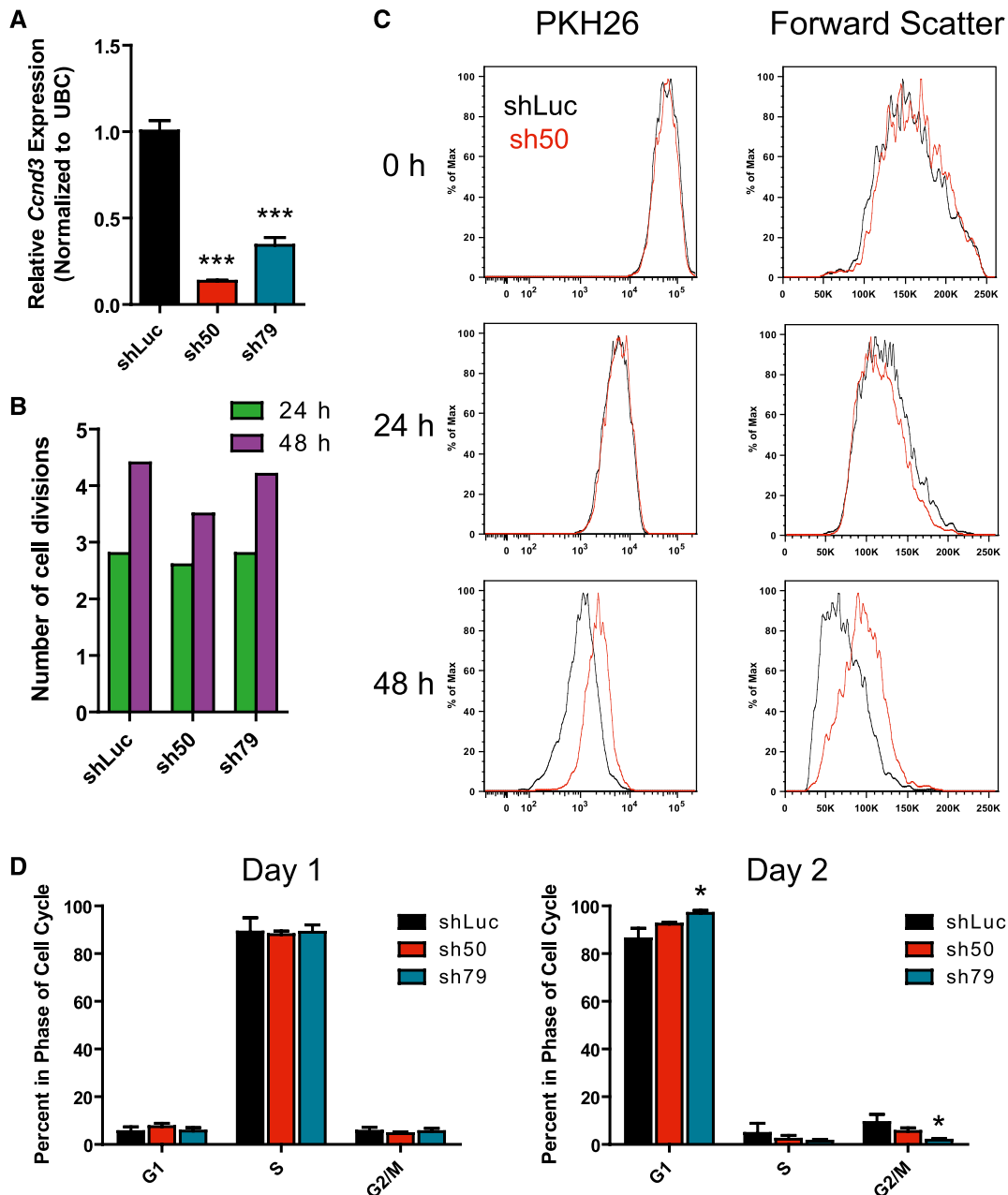
Figure 4. Cyclin D3 regulates the number of cell divisions during terminal mouse erythropoiesis. (*A*) Relative expression levels of *Ccnd3* are shown normalized to the *Ubc* control and were measured by quantitative RT–PCR from samples obtained on day 2 of culture. (*B*) The average number of cell divisions calculated as discussed in the Materials and Methods from the mean fluorescence intensity measurements for PKH26-labeled FL cells (transduced with the luciferase shRNA, shLuc, control, or sh50 and sh79 that target *Ccnd3*) at 24 and 48 h. (*C*) An example of the PKH26 labeling, along with forward scatter measurements at 0, 24, and 48 h, is shown for sh50 compared with shLuc. (*D*) The distribution of cells in various phases of the cell cycle as determined by BrdU labeling for 30 min on days 1 or 2 of culture is shown.

Discussion

Genetic studies of the metazoan cell cycle machinery have revealed specialized cell type-specific functions for many of the ubiquitously expressed proteins involved, revealing a previously unappreciated extent of complexity for the cell cycle in specialized cell types (Sherr and Roberts 2004; Malumbres and Barbacid 2009). Our studies on the role of cyclin D3 in erythropoiesis provide an important example of this. In humans, erythroid progenitors need to produce >2 million erythrocytes every second by ensuring that maturation and cell cycle exit on terminal differentiation are appropriately coordinated. By following up on clues from human genetic studies, we

Sankaran et al.
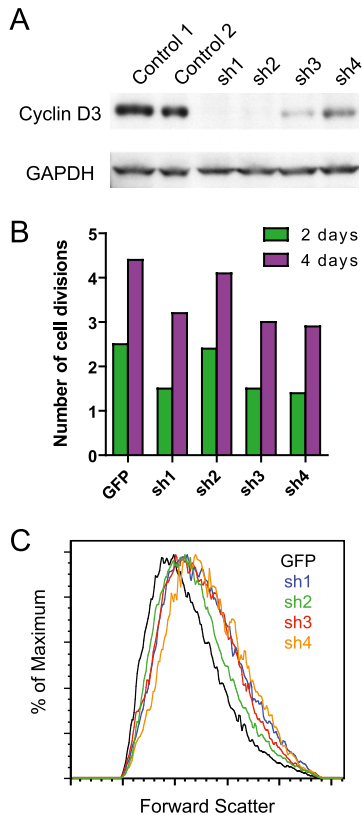
Figure 5. Knockdown of cyclin D3 in human erythropoiesis reduces the number of terminal divisions and results in increased cell size. (*A*) Knockdown of cyclin D3, as evaluated by Western blotting, in K562 cells transduced with the shRNAs targeting *CCND3* (sh1–4) or the pLKO.1 control vector (Controls 1 and 2). (*B*) The average number of terminal divisions is shown for primary adult erythroid cells transduced with sh1–4 or a GFP control lentivirus, which was calculated as discussed in the Materials and Methods from PKH26-labeling data. The cells were labeled on day 2 of differentiation and measured at 2-d intervals. (*C*) Forward scatter plots are shown for the GFP control or sh1–4 transduced erythroid cells at day 8 of differentiation, at a point when the cells are near the endpoint of terminal maturation.

demonstrated that variation in the expression of cyclin D3 regulates erythrocyte size and number through the coordination of the number of cell divisions that differentiating erythroid progenitors undergo (Fig. 6). Specifically, we showed that the variant rs9349205, which is most significantly associated with MCV and RBC count in humans, is located within an erythroid-specific enhancer element of the nearby *CCND3* gene. Moreover, the G variant at rs9349205, associated with larger MCV and reduced RBC count, is associated with decreased activity of this enhancer element, and these individuals therefore would have reduced cyclin D3 levels in developing erythroid progenitors. While we would like to measure the variation in *CCND3* mRNA expression in samples from humans with different variants at rs9349205, such experiments are confounded by the variation in both

cyclin D3 levels during erythropoiesis and the erythroid differentiation states in different human samples.

Additionally, the effect of variation in erythroid traits by the different alleles at rs9349205 is extremely subtle (0.6% of the variation in MCV is due to the allele at rs9349205) (Ganesh et al. 2009), and thus GWASs have required tens of thousands of individuals to delineate such an association with sufficient statistical power. Similarly, one would have to examine the cell cycle in erythroid cells from thousands of individuals with different alleles to reliably detect differences in cell cycle progression specifically due to rs9349205. Given the limitations that exist in following up in humans the subtle phenotypes detected by GWASs, we took the alternative approach of more dramatic perturbations of the candidate genes involved. Indeed, the phenotypes observed in the *Ccnd3* knockout animals and in knockdowns in cultured mouse and human erythroid progenitors are entirely consistent with the human genetic observations; reduced levels of cyclin D3 correlate with fewer terminal erythroid cell divisions, and fewer but larger terminally differentiated cells are produced. Thus, cyclin D3 appears to be a critical regulator of terminal erythroid proliferation through its interaction with CDK4 and CDK6. During terminal erythropoiesis, progenitors undergo progressive reductions in cell size, and premature exit from the cell cycle, triggered by reduced levels of cyclin D3, will therefore produce not only fewer cells, but also cells that are larger in size (Fig. 6).

The molecular mechanisms controlling erythrocyte size and number have been enigmatic despite extensive studies on this topic, particularly regarding the variation in these traits that has occurred in the course of evolution (Boylan et al. 1991; Hawkey et al. 1991; Hoffman 2001). In this study, we demonstrated one mechanism regulating both of these erythrocyte traits. While previous studies suggested that the expression of cell cycle components
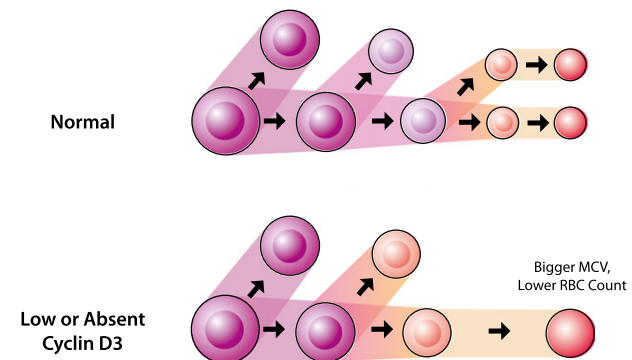
Figure 6. A model of how reduced or absent expression of cyclin D3 can modulate erythropoiesis and cause increased erythrocyte size and reduced RBC counts. The terminal erythroid cells (beginning at the proerythroblast stage of differentiation) undergo a reduced average number of cell divisions during terminal erythropoiesis, and therefore larger erythrocytes are produced with a reduction in the total number of RBCs. This model is supported by the human and mouse genetic data as well as the cellular studies presented here.

may correlate with changes in size and cell divisions during erythropoiesis (Dolznig et al. 1995), no studies have previously been able to directly examine this functional connection. Over 40 years ago, the number of cell divisions in erythroid precursors was postulated to regulate erythrocyte size, but no examples demonstrating such a connection have previously been described (Stohlman et al. 1968). Our findings show that deregulation of the number of cell divisions in erythroid precursors can have a dramatic impact on both erythrocyte size and number. Further work will be necessary to examine whether variation in the control of cell divisions during erythropoiesis may have given rise to the variations in erythrocyte size and number that have occurred in different species over the course of evolution (Hawkey et al. 1991). The relative preservation of hemoglobin levels seen among different species with inversely varying erythrocyte size and number (Boylan et al. 1991; Hawkey et al. 1991) suggests that such a mechanism coordinately impacting both of these parameters is likely to be involved.

It is interesting to note the agreement between the effects observed with reduction of *Ccnd3* levels in vitro as compared with an in vivo germline knockout of *Ccnd3*. We estimate that an average of 0.7 fewer cell divisions occur in the *Ccnd3* knockout mice based on the average reduction of 1.6-fold for RBC count in these mice [$\log_2(1.6) = 0.67$, assuming no other alterations in the kinetics of erythropoiesis]. In agreement with this, FL cultures from the *Ccnd3* knockout show a reduction of 0.7 fewer cell divisions, and an acutely mediated ~85% knockdown of *Ccnd3* in cultured cells reduced the number of divisions by ~0.9; these results collectively suggest that if compensatory responses occur in the knockout animals, these are likely to be extremely mild. However, we did note that the extent of knockdown with the lentiviral shRNAs in the K562 cell line did not consistently correlate with the extent of the effect on cell divisions in human primary cells. The exact reasons for this are not currently clear, although this may reflect cell type- or stage-specific variation in activity among various shRNAs. Nonetheless, it is remarkable that we have such overall congruent results from a variety of different experimental and genetic approaches.

Our study demonstrates the power of genetic association studies to provide insight into fundamental biological mechanisms. In this study, we not only took advantage of the results of GWASs to focus on relevant variants that affect human erythropoiesis, we also made use of genomic data to better understand the mechanisms underlying a noncoding genetic variant. Our study reveals the value of such approaches and suggests that even genetic variants with small or moderate effect sizes may provide important insight into basic biological mechanisms of great importance (Musunuru et al. 2010; Lander 2011). The approach that we used to study such subtle human genetic variation will likely be broadly applicable; more dramatic perturbation of genes in loci implicated from GWASs will be needed to study the biology underlying connections uncovered from such genetic studies. Given the large number of loci revealed from GWASs, it is

likely that other similarly important biological mechanisms have yet to be unveiled.

Materials and methods

Cell culture

293T, HeLa, and MCF-7 cells were maintained in DMEM with 10% fetal bovine serum (FBS), 2 mM L-glutamine, and 1% penicillin/streptomycin (P/S). For MCF-7 cells, the culture medium was additionally supplemented with 10 µg/mL human insulin (Sigma). For production of retro/lentiviruses, 293T cells were transfected with the appropriate viral packaging and genomic vectors using FuGene 6 reagent (Roche) according to the manufacturer's protocol. K562 cells were maintained at a density between 0.1×10^6 and 1×10^6 cells per milliliter in RPMI medium supplemented with 10% FBS, 2 mM L-glutamine, and 1% P/S. Culture of primary human and mouse cells is described below.

Luciferase reporter assay

K562, 293T, HeLa, and MCF-7 cells were seeded in a 24-well plate at an equal density of 50,000–100,000 cells per well. For transfection of K562 cells in each well, 0.5 µg of the luciferase reporter vectors pGL3-SV40 promoter, pGL3-SV40-Enhancer (G allele), or pGL3-SV40-Enhancer (A allele) were cotransfected with 50 ng per well of the pRL-SV40 vector using the Lipofectamine LTX and PLUS transfection reagent according to the manufacturer's protocol (Invitrogen). For transfection of 293T, HeLa, and MCF-7 cells in each well, 0.35 µg of the respective luciferase reporter vectors were cotransfected with 35 ng per well of the pRL-SV40 vector using the FuGene 6 reagent according to the manufacturer's protocol (Roche). Cells were incubated at 37°C with 5% CO_2 until analysis at 48 h. For measuring luciferase reporter activity, the Dual-Glo Luciferase assay system (Promega) was used according to the manufacturer's protocol. Briefly, cells were resuspended in Dual-Glo Luciferase assay reagent and incubated at room temperature for 10–30 min until measurement of firefly luminescence on a Safire 2 microplate reader (Tecan). Subsequently, Dual-Glo Stop and Glo reagent was added to the suspension and incubated for 10–30 min at room temperature until measurement of *Renilla* luciferase activity. For each sample, the ratio of Firefly:Renilla luminescence was calculated and normalized to the signal of the pGL3-SV40 promoter control vector.

Constructs

The shRNA constructs targeting mouse *Ccnd3* were sh50 (CCAACCTTCTCAGTTGCCAAAGTCGACTTTGGCAACTGAGAAGGTTGG) and sh79 (CCTTTGCGATGTATCCTCCATGTCGACATGGAGGATACATCGCAAAGG). These sequences were cloned into the BbsI sites of the MSCV-pgkGFP-U3-U6P vector, which coexpresses GFP from a PGK promoter.

The shRNA constructs targeting human *CCND3* were obtained from the Mission shRNA collection (Sigma-Aldrich). The constructs were in the pLKO.1 lentiviral vector. The following were the shRNA sequences that were used: sh1 (CCGGCGCTGTGAGGAGGAAGTCTTCCTCGAGGAAGACTTCCTCCTCACAGCGTTTTT), sh2 (CCGGGCACATGATTTCCTGGCCTTCCTCGAGGAAGGCCAGGAAATCATGTGCTTTTT), sh3 (CCGGCAGACCAGCACTCCTACAGATCTCGAGATCTGTAGGAGTGCTGGTCTGTTTTT), and sh4 (CCGGCCAGCACTCCTACAGATGTCACTCGAGTGACATCTGTAGGAGTGCTGGTTTTT).

Sankaran et al.

The 280-bp core region surrounding the variant rs9349205 was amplified from human genomic DNA using the Phusion High-Fidelity Polymerase (New England Biolabs) and cloned into the NheI and XhoI site of the pGL3-SV40 promoter vector (Promega) to yield pGL3-SV40-Enhancer (G allele). The amplified region was verified by DNA sequencing and was identical to the latest available human genome reference sequence (hg19). The major G allele was mutated to the minor A allele using the QuikChange II XL site-directed mutagenesis kit (Agilent Technologies), according to the manufacturer's protocol, to yield pGL3-SV40-Enhancer (A allele). All constructs were verified by DNA sequencing. The following primers were used for the cloning of the enhancer fragment and mutagenesis: Enhancer-FW (5′-GATCGCTA GCCATTGATCCTTCTGGCCCTA-3′), Enhancer-RV (5′-GATC CTCGAGCTTCAGTCCAAGGACCTGCT-3′), Enhancer-GtoA-FW (5′-GCCTTATCACCCACCACTGGCATCCTGCTTC-3′), and Enhancer-GtoA-RV (5′-GAAGCAGGATGCCAGTGGTGGGT GATAAGGC-3′).

Mouse FL and bone marrow erythroid progenitor purification, retrovirus infection, in vitro culture, and PD0332991 treatment

Embryonic day 14.5–15.5 (E14.5–E15.5) FL cells were homogenized in PBS supplemented with 2% FBS and 100 μM EDTA. Six-week-old to 12-wk-old mouse bone marrow was harvested in PBS supplemented with 2% FBS and 100 μM EDTA and homogenized, as was done for the FL cells. Mature erythrocytes were lysed by the addition of ammonium chloride solution (Stem Cell Technologies, Inc.) at a 1:4 ratio and incubation for 10 min on ice. After washing, the remaining cells were incubated with a cocktail of biotin-conjugated antibodies, including Lineage Cocktail (BD Pharmingen, 559971), Ter119 (eBioscience, 13-5921-85), CD16/32 (Abcam, 25249), Sca-1 (BD Pharmingen, 553334), CD34 (MCA,1825B), and CD41 (MCA, 2245B). After magnetic depletion with streptavidin beads (BD Pharmingen, 557812), a pure FL Ter119-negative erythroid progenitor population was obtained (Flygare et al. 2011).

For retroviral infection, 293T cells were transfected with the retroviral construct described above along with the pCL-eco packaging vector. Medium was changed the day after transfection. After 24 h, this medium was collected and filtered at 0.45 μm immediately prior to infection of purified erythroid progenitor cells. The cells were mixed with viral supernatant, and polybrene (filtered 4 mg/mL stock) was added to the mixture at 0.4 μL/mL medium in a 24-well plate at a density of 100,000 cells per well. The cells were spun at 2000 rpm for 90 min at 32°C.

Subsequently, for differentiation, cells were resuspended in IMDM containing 15% FBS and 0.5 U/mL Epo (Amgen) for up to 48 h at 37°C with 5% CO_2. Where indicated, cultures were supplemented with 0–2000 nM PD0332991 or DMSO (PD0332991 was kindly supplied by the laboratory of P. Sicinski).

Human CD34+ cell-derived erythroid cultures and lentiviral transduction

Culture of human adult peripheral blood-mobilized CD34+ progenitors was performed using a two-stage culture method, as has been described previously (Sankaran et al. 2008a, 2011). CD34+ cells were obtained from magnetically sorted mononuclear samples of G-CSF-mobilized peripheral blood from donors and were frozen after isolation. Cells were obtained from the Yale Center of Excellence in Molecular Hematology (YCEMH). Cells were thawed and washed into RPMI-1640 with 10% FCS, pelleted, and then seeded in StemSpan SFEM medium (Stem Cell Technologies, Inc.) with 1× CC100 cytokine mix (Stem Cell

Technologies, Inc.) and 2% P/S. Cells were maintained in this expansion medium at a density of 0.1×10^6–1×10^6 cells per milliliter, with medium changes every other day as necessary. Cells were lentivirally transduced at day 3 of expansion and kept in expansion medium for a total of 6 d. After this expansion phase, the cells were reseeded into StemSpan SFEM medium with 2% P/S, 20 ng/mL SCF (PeproTech, Inc.), 1 U/mL Epo (Amgen), 5 ng/mL IL-3 (PeproTech, Inc.), 2 μM dexamethasone (Sigma-Aldrich), and 1 μM β-estradiol (Sigma-Aldrich). Cells were maintained in differentiation medium, with medium changes every other or every third day as needed. Cells were maintained at a density of 0.1×10^6–1×10^6 cells per milliliter. By day 3 of differentiation, homogeneous larger blasts were present in the culture. By day 5, the majority of cells had proerythroblast morphology, and on day 7, the majority of the cells had basophilic erythroblast morphology. By day 12 of differentiation, the majority of cells demonstrated ortho-chromatophilic and polychromatophilic erythroblast morphology. This morphological classification has been confirmed previously using phenotypic markers of erythropoiesis, including CD235, CD71, CD45, and CD36.

May-Grünwald–Giemsa staining

Approximately 50,000–200,000 cells were placed in 200 μL of PBS with 1% FBS and centrifuged onto poly-L-lysine-coated slides by spinning in the cytocentrifugation apparatus for 4 min at 300 rpm. After drying, the slides were stained in May-Grünwald solution (Harleco, Inc.) for 5 min. Slides were washed with water three times (30 sec per wash) and stained in Giemsa (Sigma-Aldrich) at a 1:20 dilution for 15 min. Slides were subsequently washed five times (30 sec per wash) with water and dried. Then, slides were mounted with coverslips and examined. All images shown from cytospin sections were taken at 100× magnification.

Immunohistochemistry

Immunohistochemistry was performed on sections of normal bone marrow biopsies as has been described previously (Metcalf et al. 2010). Secondary staining with glycophorin A was performed in some samples to detect colocalization of cyclin D3 with erythroid cells and has been shown in a previously published report (Metcalf et al. 2010). Representative sections are shown in the figures.

Analysis of mouse hematopoiesis

$Ccnd3^{-/-}$, $Ccnd2^{-/-}$, and $Ccnd1^{-/-}$ mice have been described previously (Ciemerych et al. 2002; Sicinska et al. 2003, 2006; Kozar et al. 2004; Cooper et al. 2006). Genotyping of these mice was performed as described in the prior publications. The mouse hematopoietic analysis involved performing routine complete blood counts, morphological analysis of peripheral blood and bone marrow, and other methods using standard approaches that have been described in detail before (Sankaran et al. 2008b). Epo-mediated stimulation of erythropoiesis was carried out in a manner similar to what has been described previously (Singbrant et al. 2011), with daily injections of the Epo over the course of 8 d. All mice were maintained on a C57Bl/6 background.

FLs were derived from E14.5–E15.5 embryos using timed matings of C57Bl/6 wild-type and $Ccnd3^{+/-}$ mice, as discussed above in the section regarding the isolation of progenitors from mouse FL.

All animal work was performed in accordance with guidelines set forth by the appropriate committees on animal care at the

Massachusetts Institute of Technology and the Dana-Farber Cancer Institute.

Cell cycle analyses

In vitro cultured erythroid cells were pulsed with 10 μM BrdU for 30 min, and BrdU incorporation was detected using a BrdU flow kit (BD Pharmingen, 552598) at indicated time points. Briefly, pulsed cells were fixed and permeabilized, treated with DNase to expose incorporated BrdU epitopes, and stained by APC-conjugated anti-BrdU antibodies. 7-aminoactinomycin (7-AAD) was added to stain for DNA content. Stained cells were analyzed at a rate of no more than 100 events per second on a BD Pharmingen LSR II flow cytometer. The 7-AAD signal data was acquired on a linear scale.

To allow tracking of the number of cell divisions, FL Ter119-negative erythroid progenitor cells from $Ccnd3^{+/+}$, $Ccnd3^{+/-}$, and $Ccnd3^{-/-}$ or C57Bl/6 shRNA-infected cells were labeled with the PKH26 red fluorescent cell linker kit (Sigma-Aldrich, PKH26GL-1KT). Alternatively, lentivirally-transduced human erythroid progenitors were used for this labeling procedure, typically at day 2 of differentiation. Briefly, an equal number of cells was washed with PBS and resuspended in Diluent C. The PKH26 dye was prepared immediately prior to staining, and an equal volume of a 4 μM solution in Diluent C was added to the cell suspension and mixed immediately by pipetting. The cells were incubated for 5 min at room temperature. The staining reaction was stopped by addition of an equal volume of FBS and incubation for 1 min. Subsequently, the labeled cells were washed three times with IMDM with 10% FBS, resuspended in Epo-containing medium described above, and cultured at 37°C with 5% CO_2 until analysis at the indicated time points. An aliquot of the labeled cells was used to measure the mean fluorescence intensity (MFI) of PKH26 immediately after labeling (0 h).

Assuming an equal distribution of the fluorophore dye PKH26 on the daughter cells after each cell division, we would anticipate a reduction of the PKH26 MFI by the factor 2 with each division. Accordingly, we would anticipate a reduction of 2^x after x divisions (Eq. 1) after y hours in culture.

$$MFI(y \text{ h}) = \frac{MFI(0 \text{ h})}{2^x} \qquad (1)$$

Therefore, measuring the MFI after y hours of culture would allow one to approximate the number of cell divisions, x, as follows (Eqs. 2–5):

$$2^x = \frac{MFI(0 \text{ h})}{MFI(y \text{ h})} \qquad (2)$$

$$\log[2^x] = \log\left[\frac{MFI(0 \text{ h})}{MFI(y \text{ h})}\right] \qquad (3)$$

$$X \cdot \log[2] = \log\left[\frac{MFI(0 \text{ h})}{MFI(y \text{ h})}\right] \qquad (4)$$

$$X = \frac{\log\left[\frac{MFI(0 \text{ h})}{MFI(y \text{ h})}\right]}{\log[2]} \qquad (5)$$

Flow cytometry analysis and sorting

For flow cytometry analysis, in vitro cultured erythroid cells were washed in PBS and stained with 7-AAD or 1 μg/mL propidium iodide (PI), 1:100 APC-conjugated Ter119 (eBioscience, 17-5921-83), 1:300 PE-conjugated CD71 (eBioscience, 12-0711-83), and 1 μg/mL Hoechst, followed by FACS analysis (BD Bioscience LSR II flow cytometer) (Ji et al. 2008). Data were analyzed using FlowJo 8.6.9 (TreeStar).

For flow cytometry sorting of GFP+ cells, in vitro cultured erythroid cells were washed with PBS and stained with 1 μg/mL PI. GFP-positive/PI-negative cells were sorted on a BD FACSAria cell sorter.

Quantitative RT–PCR

Isolation of RNA was performed using the miRNeasy minikit (Qiagen). An on-column DNase (Qiagen) digestion was performed according to the 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: *Ccnd3* forward (5′-CGAGCCTCCTACTTCCAGTG-3′) and reverse (5′-GGACAG GTAGCGATCCAGGT-3′), and *Ubc* forward (5′-GAGTTCC GTCTGCTGTGTGA-3′) and reverse (5′-CCTCCAGGGTGA TGGTCTTA-3′).

Western blotting

Approximately 2.5×10^6 cells were harvested at indicated time points, washed twice in PBS, resuspended in RIPA lysis buffer (50 mM Tris-HCl at pH 7.4, 150 mM NaCl, 0.1% SDS, 1% NP-40, 0.25% sodium deoxycholate), and incubated for 10 min on ice. After centrifugation at 14,000 rpm for 5 min at 4°C to remove cellular debris, the remaining supernatant was transferred to a new tube, supplemented with sample buffer, and incubated for 10 min at 60°C. Proteins were separated by SDS gel electrophoresis using the NuPAGE Bis-Tris gel system (Invitrogen) and MOPS running buffer under reducing conditions. Subsequently, proteins were transferred onto a nitrocellulose membrane using the NuPAGE transfer buffer (Invitrogen). Membranes were blocked with 5% milk-PBST for at least 1 h and probed with cyclin D3 mouse monoclonal antibody (D7; sc-6283, Santa Cruz Biotechnology) at a 1:500 dilution or GAPDH rabbit polyclonal antibody (FL-335; sc-25778, Santa Cruz Biotechnology) at a 1:1000 dilution in 2.5% milk-PBST for 1 h at room temperature or overnight at 4°C. Membranes were washed four times with PBST, incubated with sheep anti-mouse or donkey anti-rabbit peroxidase-coupled secondary antibodies (NA931 and NA934, GE Healthcare) at a 1:2000 dilution in 2.5% milk-PBST for 1 h at room temperature, washed three times with PBST, and incubated for 2 min with Western Lightning Plus-ECL substrate (Perkin Elmer). Proteins were visualized by exposure to scientific imaging film (Kodak).

ChIP and 3C

ChIP was performed as has been described previously (Xu et al. 2010). Briefly, 0.5×10^7–1×10^7 cells per immunoprecipitation were cross-linked with 1% formaldehyde for 5 min at room temperature. Chromatin was sonicated in RIPA buffer with 0.3 M NaCl to ~500 bp. Dynabeads Protein A or G (Invitrogen) was used for collection of chromatin. Antibodies were incubated with beads for 3 h before incubating with sonicated chromatin overnight. Sequencing of ChIP material took place on the HeliScope single-molecule sequencer, with 5 million–10 million

reads per sample used in the analysis. The following antibodies were used in the ChIP coupled with deep sequencing (ChIP-seq) analyses: H3K4me1 (Abcam, ab8895), H3K4me3 (Millipore, 04-745), H3K9ac (Millipore, 06-942), GATA1 (Abcam, ab47490), TAL1 (Santa Cruz Biotechnology, sc-12984), and KLF1 (Abcam, ab49158). 3C was performed on in vitro derived human adult proerythroblasts (day 5 of differentiation), as has been described previously. A bacterial artificial chromosome (clone RP11-720D9 from the Children's Hospital Oakland Research Institute [CHORI]) that covered the *CCND3* gene and the upstream enhancer region was used as a control for the 3C reactions.

For allele-specific ChIP assays, ChIP was performed with TAL1 antibody (Santa Cruz Biotechnology, sc-12984), as discussed above, on CD34-derived erythroid cell samples prospectively genotyped as being heterozygous at rs9349205. PCR of ChIP DNA was performed with AmpliTaq Gold DNA Polymerase (Applied Biosystems) and *CCND3* forward and reverse primers encompassing rs9349205 (5′-CTTCAGTCCAAGGACCTGCT-3′ and 5′-AAGGGAAGAGGCTGGATTTG-3′). The 191-bp amplicon was cloned into pCR2.1 (Invitrogen TA cloning kit) according to the manufacturer's directions. Sanger sequencing of individual clones was performed with the forward primer (5′-CTTCAGTCCAAGGACCTGCT-3′) to determine genotype at rs9329205.

Statistical analysis of experimental data

Pairwise comparisons of experimental results were performed using either unpaired two-sided Student's *t*-tests for experimental data with appropriate biological replicates (comparing control samples with experimental samples) or the Fisher's exact tests for comparison of allele frequencies for allele-specific ChIP assays that were done by comparing the input DNA control with the corresponding ChIP sample from a single individual.

Human genetic analysis

The data from GWASs of erythrocyte traits were analyzed, and the most significantly associated SNPs from each GWAS at the 6p21.1 locus were assessed. We used the SNAP tool (http://www.broadinstitute.org/mpg/snap) (Johnson et al. 2008) or ad hoc scripts to analyze all common genetic variants from the appropriate 1000 Genomes Project Consortium (2010) populations that are in close linkage disequilibrium with the most significantly associated variants. Given the strong genetic evidence from multiple GWASs that support rs9349205 as being the most significantly associated variant at the 6p21.1 locus, we limited our analysis to variants in close linkage disequilibrium with this variant ($r^2 \geq 0.8$). The University of California at Santa Cruz (UCSC) Genome Browser (http://genome.ucsc.edu) was loaded with custom annotation tracks using data derived from ChIP-seq analysis of various histone modifications and transcription factors in human erythroid cells to analyze the presence of such modifications at the site of the genetic variants. Analysis of chromatin state segmentation (Ernst et al. 2011) was assessed using the appropriate tracks from the ENCODE annotations on the UCSC Genome Browser. The chromatin states were defined as described previously (Ernst et al. 2011).

References

1000 Genomes Project Consortium. 2010. A map of human genome variation from population-scale sequencing. *Nature* **467:** 1061–1073.

Balarajan Y, Ramakrishnan U, Ozaltin E, Shankar AH, Subramanian SV. 2011. Anaemia in low-income and middle-income countries. *Lancet* **378:** 2123–2135.

Boylan JW, Van Liew JB, Feig PU. 1991. Inverse changes in erythroid cell volume and number regulate the hematocrit in newborn genetically hypertensive rats. *Proc Natl Acad Sci* **88:** 9848–9852.

Cheng Y, Wu W, Kumar SA, Yu D, Deng W, Tripic T, King DC, Chen KB, Zhang Y, Drautz D, et al. 2009. Erythroid GATA1 function revealed by genome-wide analysis of transcription factor occupancy, histone modifications, and mRNA expression. *Genome Res* **19:** 2172–2184.

Ciemerych MA, Kenney AM, Sicinska E, Kalaszczynska I, Bronson RT, Rowitch DH, Gardner H, Sicinski P. 2002. Development of mice expressing a single D-type cyclin. *Genes Dev* **16:** 3277–3289.

Cooper AB, Sawai CM, Sicinska E, Powers SE, Sicinski P, Clark MR, Aifantis I. 2006. A unique function for cyclin D3 in early B cell development. *Nat Immunol* **7:** 489–497.

Dai MS, Mantel CR, Xia ZB, Broxmeyer HE, Lu L. 2000. An expansion phase precedes terminal erythroid differentiation of hematopoietic progenitor cells from cord blood in vitro and is associated with up-regulation of cyclin E and cyclin-dependent kinase 2. *Blood* **96:** 3985–3987.

Dolznig H, Bartunek P, Nasmyth K, Mullner EW, Beug H. 1995. Terminal differentiation of normal chicken erythroid progenitors: Shortening of G1 correlates with loss of D-cyclin/cdk4 expression and altered cell size control. *Cell Growth Differ* **6:** 1341–1352.

Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M, et al. 2011. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473:** 43–49.

Ferreira MA, Hottenga JJ, Warrington NM, Medland SE, Willemsen G, Lawrence RW, Gordon S, de Geus EJ, Henders AK, Smit JH, et al. 2009. Sequence variants in three loci influence monocyte counts and erythrocyte volume. *Am J Hum Genet* **85:** 745–749.

Flygare J, Rayon Estrada V, Shin C, Gupta S, Lodish HF. 2011. HIF1α synergizes with glucocorticoids to promote BFU-E progenitor self-renewal. *Blood* **117:** 3435–3444.

Ganesh SK, Zakai NA, van Rooij FJ, Soranzo N, Smith AV, Nalls MA, Chen MH, Kottgen A, Glazer NL, Dehghan A, et al. 2009. Multiple loci influence erythrocyte phenotypes in the CHARGE Consortium. *Nat Genet* **41:** 1191–1198.

Hawkey CM, Bennett PM, Gascoyne SC, Hart MG, Kirkwood JK. 1991. Erythrocyte size, number and haemoglobin content in vertebrates. *Br J Haematol* **77:** 392–397.

Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, Barrera LO, Van Calcar S, Qu C, Ching KA, et al. 2007. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* **39:** 311–318.

Hoffman JF. 2001. Questions for red blood cell physiologists to ponder in this millenium. *Blood Cells Mol Dis* **27:** 57–61.

Ji P, Jayapal SR, Lodish HF. 2008. Enucleation of cultured mouse fetal erythroblasts requires Rac GTPases and mDia2. *Nat Cell Biol* **10:** 314–321.

Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'Donnell CJ, de Bakker PI. 2008. SNAP: A Web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics* **24:** 2938–2939.

Kamatani Y, Matsuda K, Okada Y, Kubo M, Hosono N, Daigo Y, Nakamura Y, Kamatani N. 2010. Genome-wide association study of hematological and biochemical traits in a Japanese population. *Nat Genet* **42:** 210–215.

Kozar K, Ciemerych MA, Rebel VI, Shigematsu H, Zagozdzon A, Sicinska E, Geng Y, Yu Q, Bhattacharya S, Bronson RT, et al. 2004. Mouse development and cell proliferation in the absence of D-cyclins. *Cell* **118:** 477–491.

Lander ES. 2011. Initial impact of the sequencing of the human genome. *Nature* **470:** 187–197.

Malumbres M, Barbacid M. 2009. Cell cycle, CDKs and cancer: A changing paradigm. *Natl Rev* **9:** 153–166.

Manolio TA, Brooks LD, Collins FS. 2008. A HapMap harvest of insights into the genetics of common disease. *J Clin Invest* **118:** 1590–1605.

Martin DI, Tsai SF, Orkin SH. 1989. Increased γ-globin expression in a nondeletion HPFH mediated by an erythroid-specific DNA-binding factor. *Nature* **338:** 435–438.

Metcalf RA, Zhao S, Anderson MW, Lu ZS, Galperin I, Marinelli RJ, Cherry AM, Lossos IS, Natkunam Y. 2010. Characterization of D-cyclin proteins in hematolymphoid neoplasms: Lack of specificity of cyclin-D2 and D3 expression in lymphoma subtypes. *Mod Pathol* **23:** 420–433.

Moffat J, Grueneberg DA, Yang X, Kim SY, Kloepfer AM, Hinkle G, Piqani B, Eisenhaure TM, Luo B, Grenier JK, et al. 2006. A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. *Cell* **124:** 1283–1298.

Musunuru K, Strong A, Frank-Kamenetsky M, Lee NE, Ahfeldt T, Sachs KV, Li X, Li H, Kuperwasser N, Ruda VM, et al. 2010. From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. *Nature* **466:** 714–719.

Sankaran VG, Menne TF, Xu J, Akie TE, Lettre G, Van Handel B, Mikkola HK, Hirschhorn JN, Cantor AB, Orkin SH. 2008a. Human fetal hemoglobin expression is regulated by the developmental stage-specific repressor BCL11A. *Science* **322:** 1839–1842.

Sankaran VG, Orkin SH, Walkley CR. 2008b. Rb intrinsically promotes erythropoiesis by coupling cell cycle exit with mitochondrial biogenesis. *Genes Dev* **22:** 463–475.

Sankaran VG, Menne TF, Scepanovic D, Vergilio JA, Ji P, Kim J, Thiru P, Orkin SH, Lander ES, Lodish HF. 2011. MicroRNA-15a and -16-1 act via MYB to elevate fetal hemoglobin expression in human trisomy 13. *Proc Natl Acad Sci* **108:** 1519–1524.

Sherr CJ, Roberts JM. 2004. Living with or without cyclins and cyclin-dependent kinases. *Genes Dev* **18:** 2699–2711.

Sicinska E, Aifantis I, Le Cam L, Swat W, Borowski C, Yu Q, Ferrando AA, Levin SD, Geng Y, von Boehmer H, et al. 2003. Requirement for cyclin D3 in lymphocyte development and T cell leukemias. *Cancer Cell* **4:** 451–461.

Sicinska E, Lee YM, Gits J, Shigematsu H, Yu Q, Rebel VI, Geng Y, Marshall CJ, Akashi K, Dorfman DM, et al. 2006. Essential role for cyclin D3 in granulocyte colony-stimulating factor-driven expansion of neutrophil granulocytes. *Mol Cell Biol* **26:** 8052–8060.

Singbrant S, Russell MR, Jovic T, Liddicoat B, Izon DJ, Purton LE, Sims NA, Martin TJ, Sankaran VG, Walkley CR. 2011. Erythropoietin couples erythropoiesis, B-lymphopoiesis, and bone homeostasis within the bone marrow microenvironment. *Blood* **117:** 5631–5642.

Soranzo N, Spector TD, Mangino M, Kuhnel B, Rendon A, Teumer A, Willenborg C, Wright B, Chen L, Li M, et al. 2009. A genome-wide meta-analysis identifies 22 loci associated with eight hematological parameters in the HaemGen consortium. *Nat Genet* **41:** 1182–1190.

Stohlman F Jr, Ebbe S, Morse B, Howard D, Donovan J. 1968. Regulation of erythropoiesis. XX. Kinetics of red cell production. *Ann N Y Acad Sci* **149:** 156–172.

Tallack MR, Whitington T, Yuen WS, Wainwright EN, Keys JR, Gardiner BB, Nourbakhsh E, Cloonan N, Grimmond SM, Bailey TL, et al. 2010. A global role for KLF1 in erythropoiesis revealed by ChIP-seq in primary erythroid cells. *Genome Res* **20:** 1052–1063.

Toogood PL, Harvey PJ, Repine JT, Sheehan DJ, VanderWel SN, Zhou H, Keller PR, McNamara DJ, Sherry D, Zhu T, et al. 2005. Discovery of a potent and selective inhibitor of cyclin-dependent kinase 4/6. *J Med Chem* **48:** 2388–2406.

Wu W, Cheng Y, Keller CA, Ernst J, Kumar SA, Mishra T, Morrissey C, Dorman CM, Chen KB, Drautz D, et al. 2011. Dynamics of the epigenetic landscape during erythroid differentiation after GATA1 restoration. *Genome Res* **21:** 1659–1671.

Xu J, Sankaran VG, Ni M, Menne TF, Puram RV, Kim W, Orkin SH. 2010. Transcriptional silencing of γ-globin by BCL11A involves long-range interactions and cooperation with SOX6. *Genes Dev* **24:** 783–798.

Zhang J, Socolovsky M, Gross AW, Lodish HF. 2003. Role of Ras signaling in erythroid differentiation of mouse fetal liver cells: Functional analysis by a flow cytometry-based novel culture system. *Blood* **102:** 3938–3946.

Zhang J, Zahir N, Jiang Q, Miliotis H, Heyraud S, Meng X, Dong B, Xie G, Qiu F, Hao Z, et al. 2011. The autoimmune disease-associated PTPN22 variant promotes calpain-mediated Lyp/Pep degradation associated with lymphocyte and dendritic cell hyperresponsiveness. *Nat Genet* **43:** 902–907.

Cyclin D3 coordinates the cell cycle during differentiation to regulate erythrocyte size and number

Vijay G. Sankaran, Leif S. Ludwig, Ewa Sicinska, Jian Xu, Daniel E. Bauer, Jennifer C. Eng, Heide Christine Patterson, Ryan A. Metcalf, Yasodha Natkunam, Stuart H. Orkin, Piotr Sicinski, Eric S. Lander, Harvey F. Lodish

Supplementary Figure 1. Linkage disequilibrium (LD) pattern for SNPs from the 1000 Genomes project populations with rs9349205. (**A**) Pattern of pairwise LD for SNPs from the CEU population with rs9349205. SNPs in close LD ($r^2 \geq 0.8$) with rs9349205 are shown above the dotted line. The intensity of the orange color for each SNP is shown proportional to its r^2 value. (**B**) Pattern of pairwise LD for SNPs from the CHB/JPT populations with rs9349205. SNPs in close LD ($r^2 \geq 0.8$) with rs9349205 are shown above the dotted line. The intensity of the purple color for each SNP is shown proportional to its r^2 value.

Supplementary Figure 2. Chromatin and epigenetic states in the region surrounding rs9349205. (**A**) DNAse I hypersensitivity data is shown from erythroid K562 cells and primary human CD34+ derived cells in the region of rs9349205. This demonstrates that rs9349205 is within a DNAse I hypersensitive region in erythroid cells. (**B**) The DNA sequence around rs9349205 is shown. The only canonical transcription factor binding motif found in this region is the WGATAR (GATA1 binding) motif that is underlined (**C**) The chromatin state segmentations are shown for each of the 9 cell lines below the region around rs9349205. The diagrams and the chromatin states are based on the references discussed in the Methods.

Supplementary Figure 3. Chromatin modifications and looping at the *CCND3* locus on 6p21 associated with erythroid traits. (**A**) ChIP-seq data for histone modifications and erythroid transcription factors KLF1, GATA1, and TAL1 are shown in the region upstream of *CCND3* on 6p21 from primary human adult erythroid progenitors. The top panel is a larger scale view of this region. A box in this upper panel highlights the region shown at higher resolution in the lower panel. The variant rs9349205 that is most highly associated with MCV and RBC count is shown at the bottom of these two plots. (**B**) 3C assay showing the relative enrichment for

various regions surrounding the *CCND3* gene in human adult erythroid progenitors. The samples were normalized to a bacterial artificial chromosome containing this genomic region and the relative enrichment at the rs9349205 locus was set to a value 1. The results are shown as the mean ± the standard deviation (n=3). The anchor region is shown in the diagram, along with the position of the *CCND3* gene (hg19 coordinates).

Supplementary Figure 4. Pattern of cyclin D3 expression during human erythropoiesis. (**A**) Western blot showing pattern of expression of cyclin D3 during *in vitro* differentiation of human adult CD34+ cells. The cells represent early erythroid progenitors (day 1), CFU-Es (day 3), proerythroblasts (day 5), basophilic erythroblasts (day 7), and polychromatophilic/ orthochromatophilic erythroblasts (day 9). (**B, C**) Immunohistochemical staining of human bone marrow sections demonstrates robust nuclear cyclin D3 staining in early precursors of the erythroid lineage. This pattern of expression could be confirmed by double staining sections with an anti-glycophorin A antibody.

Supplementary Figure 5. Peripheral blood hematological parameters for adult *Ccnd1*, *Ccnd2* and *Ccnd3* knockout (-/-) mice. (**A**) This diagram shows the mean corpuscular hemoglobin (MCH), the red cell distribution width, and the number of platelets in *Ccnd3* knockout mice. The p-values compared with wild-type littermate controls are shown above the appropriate sample. The data are shown as the mean with the distribution of values for 11 +/+, 3 +/-, and 13 -/- animals. (**B**) Ostensibly normal erythropoiesis in *Ccnd1* and *Ccnd2* knockout mice. The erythrocyte indices hemoglobin, MCV, and RBC count) are shown for *Ccnd1* -/- (n=5), *Ccnd2* -/- (n=6), and control wild-type littermates (n=11).

Supplementary Figure 6. *Ccnd3* -/- mice continue to respond to erythropoietin. (**A**) The plots show peripheral blood indices (including hemoglobin, RBC count, MCV, and platelet count)

before and after 8 days of stimulation with erythropoietin. The data are shown as the mean with the distribution of values (n=3 per group). (**B**) Normal erythropoiesis in the bone marrow of *Ccnd3* -/- animals before and after stimulation with erythropoietin. The distribution of various morphologically distinguishable erythroid progenitors is shown from bone marrow smears evaluated by hematopathologists with > 100 cells counted per animal (n=3 animals per group).

Supplementary Figure 7. CD71, Ter119 and Hoechst dye staining of FL cells from *Ccnd3* -/- and control wildtype and heterozygous littermates at 24 and 48 hours. The frequency of enucleated cells is indicated. Representative samples are shown from one of several biological replicates.

Supplementary Figure 8. PKH26 staining and forward scatter profile of FL cells treated with PD0332991 at indicated concentrations for 48 hours. Representative samples are shown from one of several biological replicates.

Supplementary Figure 9. (**A**) CD71, Ter119 and Hoechst dye staining of cultures of bone marrow progenitor cells from *Ccnd3* -/- and control wildtype littermates at 24 and 48 hours. The frequency of enucleated cells is indicated. Representative samples are shown from one of several biological replicates. (**B**) The average number of cell divisions calculated as discussed in the Methods from the mean fluorescence intensity measurements for PKH26 labeled bone marrow cells from *Ccnd3* -/- and control wildtype littermates at 48 hours.

Supplementary Figure 10. Cytospin images (at 100X magnification) are shown of GFP control or sh1-4 cells. This image demonstrates the larger size of cyclin D3 knockdown cells compared to controls.

Supplementary Table 1: SNPs in close LD with rs9349205 in the 1000 Genomes CEU Population

SNP	Distance from rs9349205	r^2	D'	Chromosome	Coordinate hg18
rs9349205	0	1	1	chr6	42033137
rs9394841	1776	0.917	1	chr6	42034913
rs1410492	17304	0.883	1	chr6	42015833
rs3218097	19884	0.883	1	chr6	42013253
rs9349204	10781	0.848	1	chr6	42022356

Supplementary Table 2: SNPs in close LD with rs9349205 in the 1000 Genomes CHB/JPT Population

SNP	Distance from rs9349205	r^2	D'	Chromosome	Coordinate hg18
rs9349205	0	1	1	chr6	42033137
rs9394841	1776	0.954	1	chr6	42034913
rs3218108	22504	0.871	1	chr6	42010633
rs1410492	17304	0.834	1	chr6	42015833
rs3218097	19884	0.834	1	chr6	42013253
rs9357371	39840	0.834	1	chr6	41993297
rs3806113	48824	0.834	1	chr6	41984313
rs9381097	77359	0.834	1	chr6	41955778
rs9369313	78094	0.834	1	chr6	41955043
rs9462746	108880	0.834	1	chr6	41924257
rs9471676	109967	0.834	1	chr6	41923170
rs9394834	118366	0.834	1	chr6	41914771
rs9381095	118451	0.834	1	chr6	41914686
rs9349202	119578	0.834	1	chr6	41913559
rs2395795	122805	0.834	1	chr6	41910332
rs9381094	124530	0.834	1	chr6	41908607
rs6899876	128189	0.834	1	chr6	41904948
rs13193235	132500	0.834	1	chr6	41900637
rs9357366	138183	0.834	1	chr6	41894954
rs71558769	143052	0.834	1	chr6	41890085
rs3747749	154555	0.834	1	chr6	41878582
rs9367112	155239	0.834	1	chr6	41877898
rs1891453	159894	0.834	1	chr6	41873243
rs3747750	164787	0.834	1	chr6	41868350
rs3827633	165313	0.834	1	chr6	41867824
rs9471653	169749	0.834	1	chr6	41863388

Supplementary Table 3: Overlap of SNPs in close LD with rs9349205 in the 1000 Genomes CEU and CHB/JPT Populations

SNP	Distance from rs9349205	r^2	D'	Chromosome	Coordinate hg18
rs9349205	0	1	1	chr6	42033137
rs9394841	1776	0.917	1	chr6	42034913
rs1410492	17304	0.883	1	chr6	42015833
rs3218097	19884	0.883	1	chr6	42013253

Supplementary Figure 1

A

rs9349205 (CEU)

B

rs9349205 (CHBJPT)

Supplementary Figure 2

A

Scale
chr6:
K562 DS
0.1 -
0
10
CD34+ Mob Sg 1
1

41,924,500   500 bases⊢   hg19
                   41,925,000   41,925,500

rs13340461
rs114789046
rs16895128
rs4623235
rs16895130
rs76552804
rs9349205
rs11970772
rs11968166
rs114074977

B
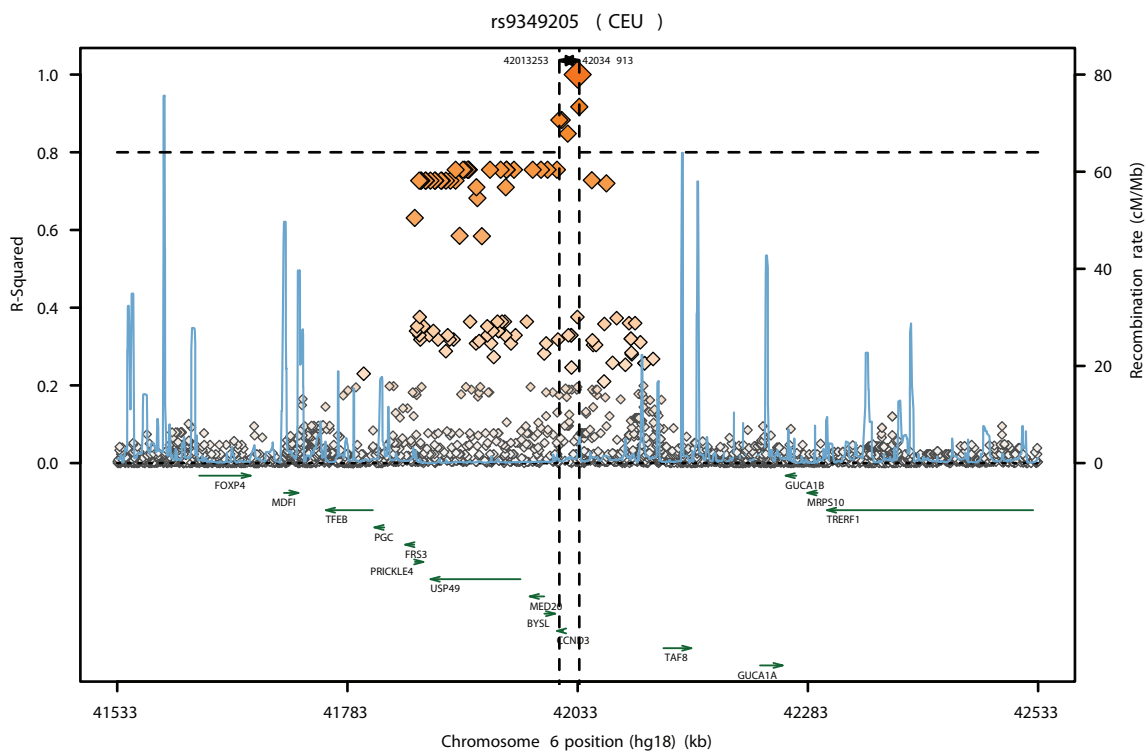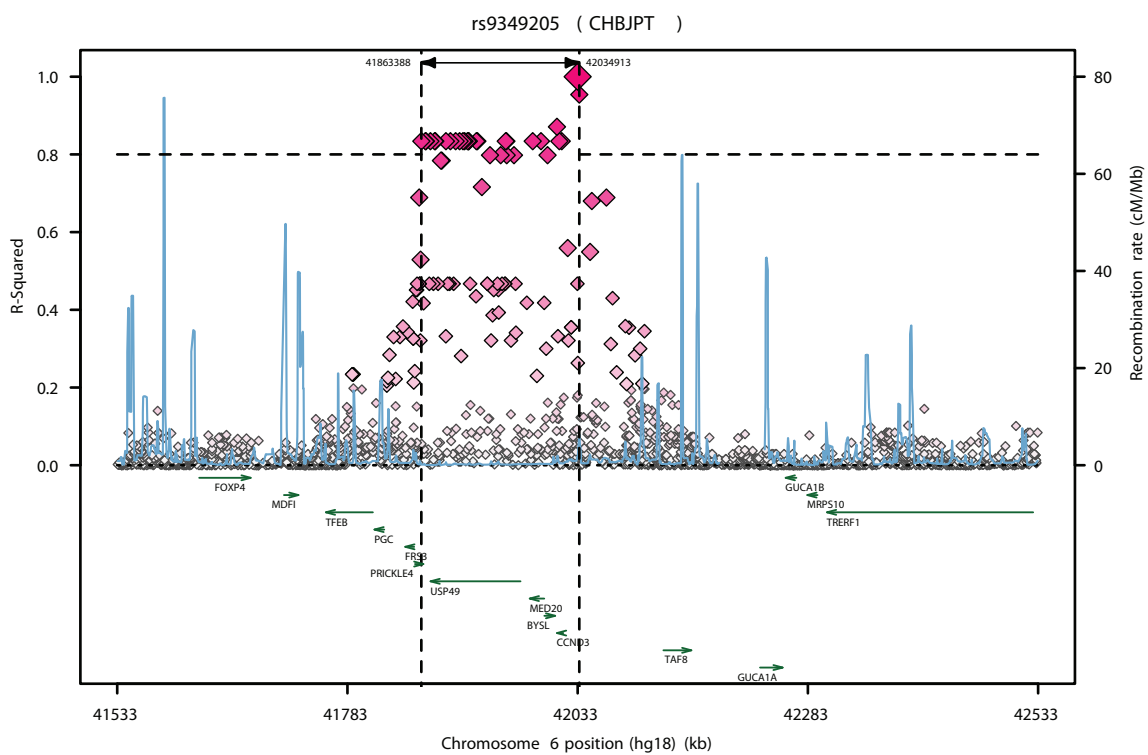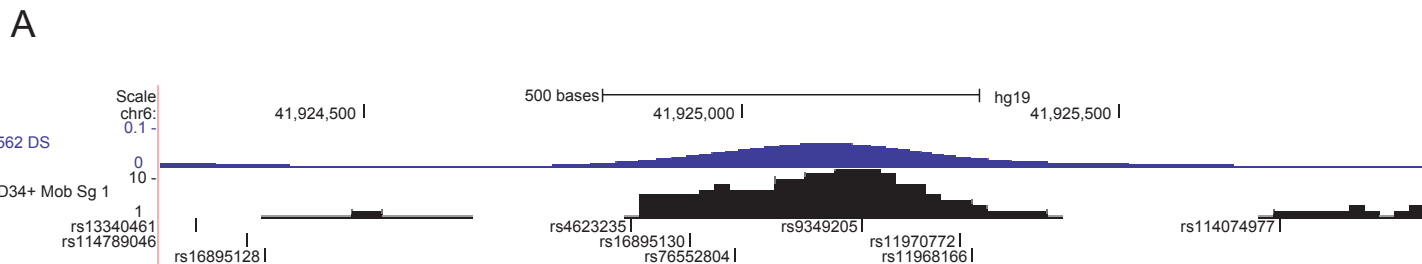
AGCCCACCTCTGGGCCTTATCACCCACCGCTGGCATCCTGCTTCTTTCTCC

GATA1 motif   rs9349205

C

Scale
chr6:   41924000   41924500   1 kb⊢   41925000   41925500   41926000   41926500

STS Markers on Genetic (blue) and Radiation Hybrid (black) Maps
NHGRI Catalog of Published Genome-Wide Association Studies

rs9349205
rs9349205
rs11970772
rs11968166

UCSC Genes Based on RefSeq, UniProt, GenBank, CCDS and Comparative Genomics
CCND3
CCND3
CCND3
RefSeq Genes
CCND3
CCND3

GM12878 Chromatin State Segmentation by HMM from ENCODE/Broad
13_Heterochrom/lo
7_Weak_Enhancer

H1-hESC Chromatin State Segmentation by HMM from ENCODE/Broad
13_Heterochrom/lo

K562 Chromatin State Segmentation by HMM from ENCODE/Broad
7_Weak_Enhancer   4_Strong_Enhancer   5_Strong_Enhancer
5_Strong_Enhancer   7_Weak_Enhancer

HepG2 Chromatin State Segmentation by HMM from ENCODE/Broad
13_Heterochrom/lo   11_Weak_Txn
6_Weak_Enhancer
7_Weak_Enhancer

HMEC Chromatin State Segmentation by HMM from ENCODE/Broad
13_Heterochrom/lo

HSMM Chromatin State Segmentation by HMM from ENCODE/Broad
13_Heterochrom/lo

HUVEC Chromatin State Segmentation by HMM from ENCODE/Broad
13_Heterochrom/lo

NHEK Chromatin State Segmentation by HMM from ENCODE/Broad
11_Weak_Txn   11_Weak_Txn   7_Weak_Enhancer
7_Weak_Enhancer   11_Weak_Txn

NHLF Chromatin State Segmentation by HMM from ENCODE/Broad
13_Heterochrom/lo

75 _   K562 H3K4me1 Histone Modifications by ChIP-seq Signal from ENCODE/Broad
K562 H3K4me1

1
75 _   K562 H3K4me3 Histone Modifications by ChIP-seq Signal from ENCODE/Broad
K562 H3K4me3

1

Supplementary Figure 3

Supplementary Figure 4

A

Day of Differentiation

1   3   5   7   9

Cyclin D3

GAPDH

B

60X Normal Bone Marrow

C

60X Normal Bone Marrow

Supplementary Figure 5

Supplementary Figure 6

A

B

Supplementary Figure 7

Supplementary Figure 8

PKH26

Forward
Scatter

Supplementary Figure 9

A

B
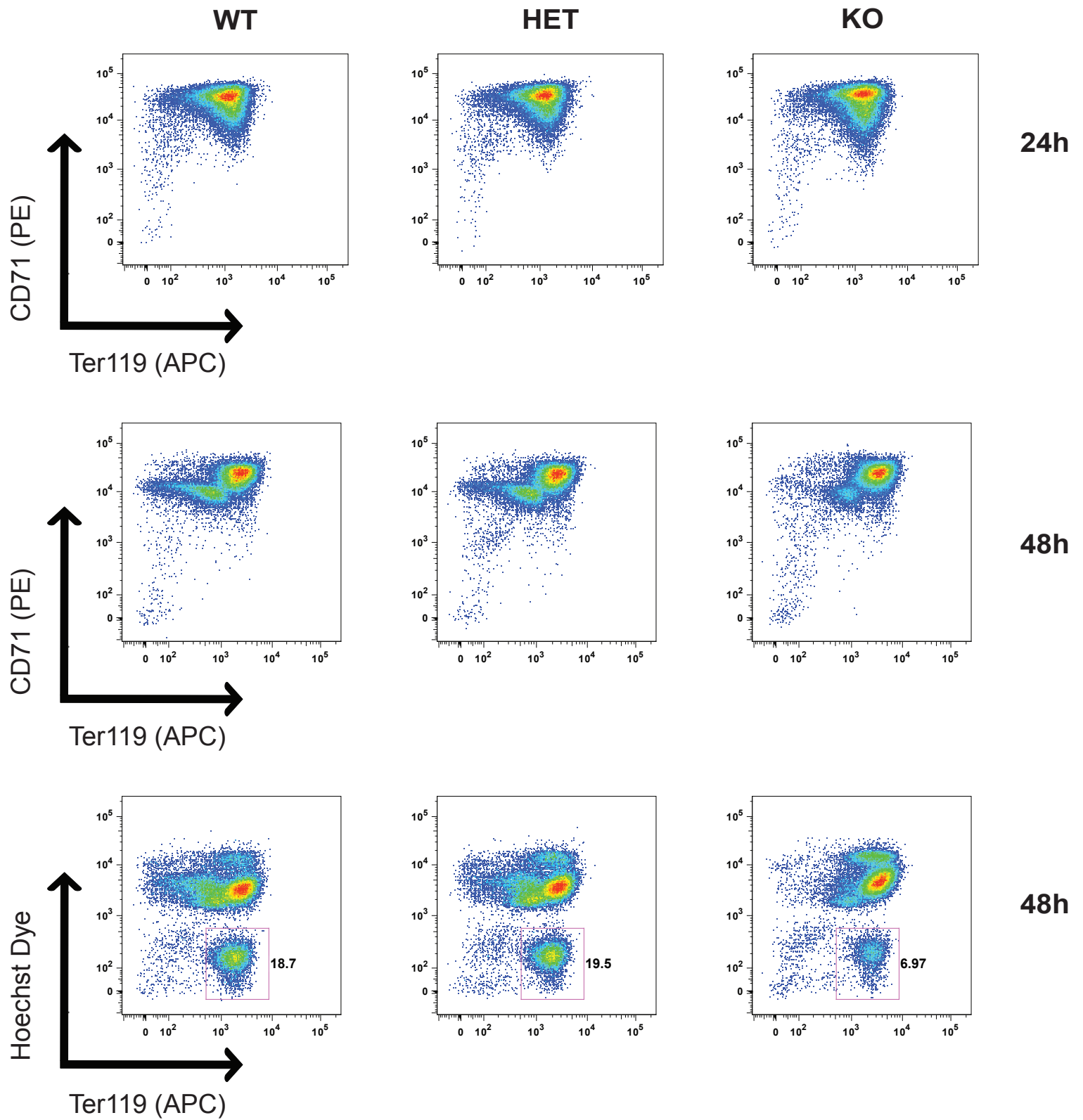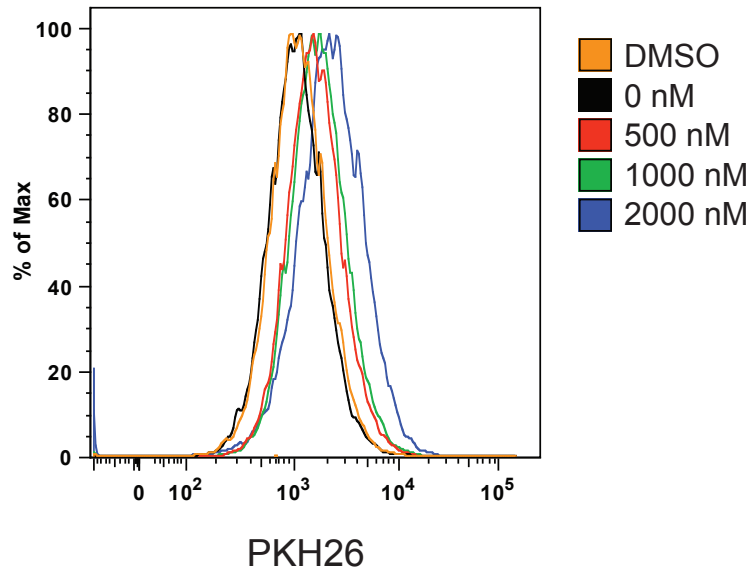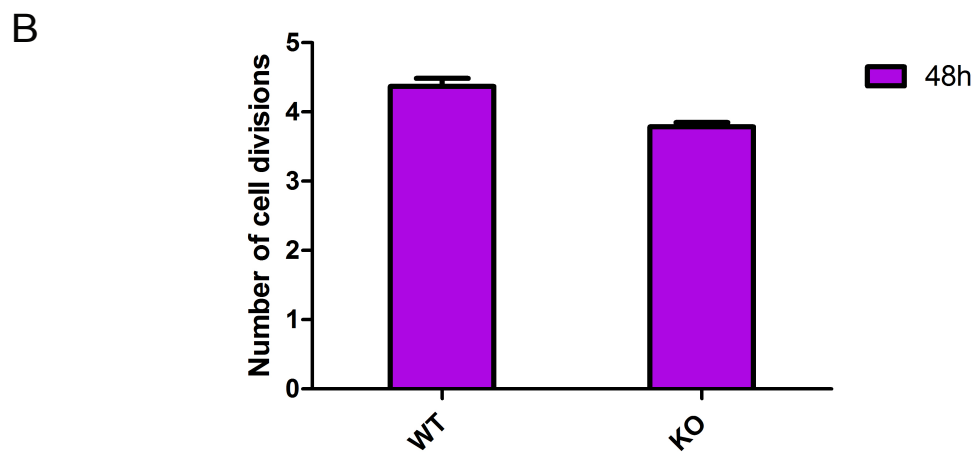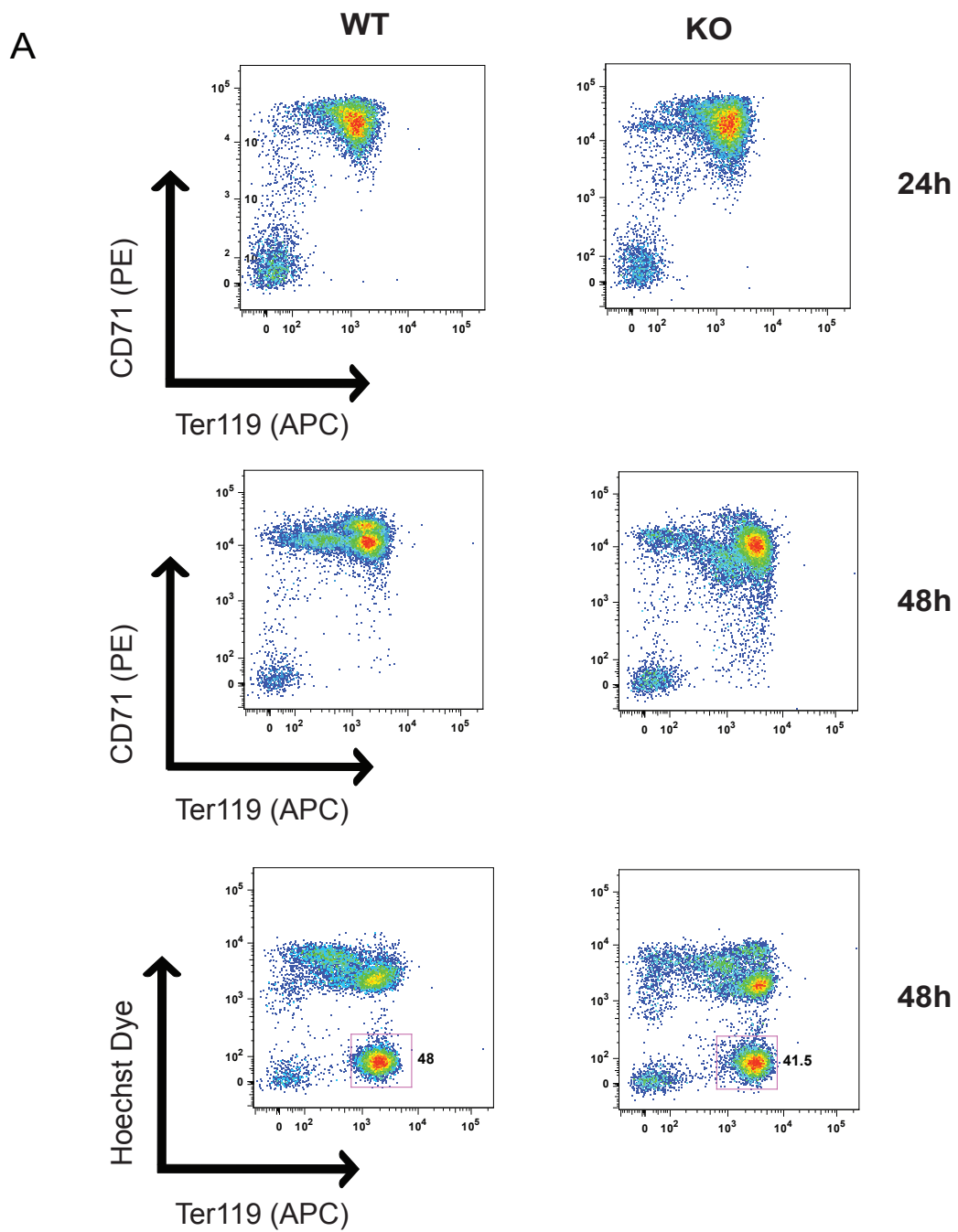
Supplementary Figure 10
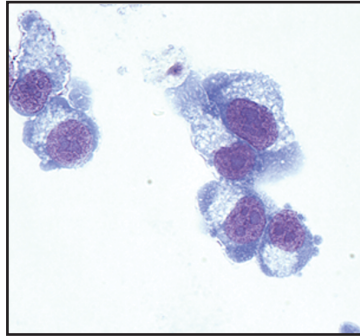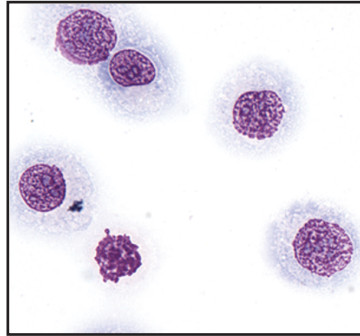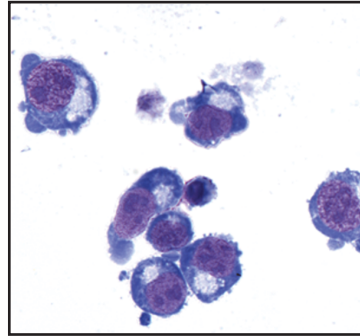
| GFP | sh1 | sh2 | sh3 | sh4 |
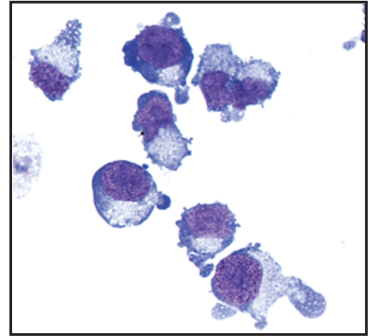
3.2 Additional information not published in the original manuscript

Supplementary Figures 11-13 contain information obtained from *in vivo* analysis of different developmental stages of erythropoiesis in *Ccnd3* knockout mice and controls.

Experimental contribution

Heide Patterson aided in immunohistochemical staining and analysis. Ference Reinhardt assisted in the timed mating of knockout mice. I performed and analyzed all other experiments with assistance from Jennifer Eng.

The following figures present my work:
Supplementary Figures 11, 12 A-B, D, 13.

3.2.1 Analysis of *in vivo* Erythropoiesis in adult *Ccnd3*$^{-/-}$ mice

Examination of 6-8 week old adult mice reveals reduced bone marrow cellularity in *Ccnd3*$^{-/-}$ mice, which is consistent with prior reports (Supplementary Figure 11A) (Sicinska et al. 2006). Staining of bone marrow cells with murine erythroid surface markers CD71 and Ter119 show comparable distribution of erythroid cells but altered frequencies (Supplementary Figure 11B), which is challenging to interpret given the defects in other hematopoietic lineages in these mice (Sicinska et al. 2003; Cooper et al. 2006; Sicinska et al. 2006). However, consistent with the reduction of peripheral red blood cells counts in the adult mice, the number of Ter119^{+} erythroid cells is reduced in bone marrow (Supplementary Figure 11C). Similar to peripheral blood erythrocytes, the size of bone marrow derived enucleated erythroid cells (Ter119^{+}Hoechst^{-}) is increased as determined by forward scatter (Supplementary Figure 11B, bottom). *Ccnd3*$^{-/-}$ mice show no signs of compensatory extramedullary erythropoiesis as assessed by spleen weight and size (Supplementary Figure 11D, E) and frequencies of differentiating CD71^{+}Ter119^{+} erythroid cells in the spleen (Supplementary Figure 11F, G) (Socolovsky et al. 2001). Thus, *Ccnd3*$^{-/-}$ adult mice present with larger, but reduced erythroid cells in the bone marrow with no major indication of defective erythropoiesis.

3.2.2 Analysis of *in vivo* Erythropoiesis in fetal development *of Ccnd3*[-/-] mice

Total cellularity and number of erythroid Ter119[+] cells in the mouse fetal liver as assessed by flow cytometry and immunohistochemistry is reduced throughout embryonic days E13.5 to E16.5 in *Ccnd3*[-/-] embryos (Supplementary Figure 12A, B, C). Staining of fetal liver cells for erythroid markers also indicate a decreased frequency of cells in the more mature CD71[low]Ter119[+] compartment (Supplementary Figure 12D). Consistent with a reduced cell count in the fetal liver, the peripheral blood counts in embryos is reduced (Supplementary Figure 13A). However, similar to adult mice the fetal liver derived enucleated (Ter119[+]Hoechst[-]) *Ccnd3*[-/-] erythroid cells present with a higher hemoglobin content (MCH) and increased cell size as assessed by microscopy and forward scatter (Supplementary Figure 13B, C, D bottom row). Interestingly, the enucleated knockout cells present with higher erythroid surface marker expression of CD71 and Ter119 (Supplementary Figure 13D top row). Presumably, this may be attributed to their larger cell surface, as erythroid progenitor cells deficient for cyclin D3 undergo fewer divisions in terminal erythropoiesis and retain a larger volume and thus surface area. Proper interpretation of altered CD71 and Ter119 differentiation profiles is thus impeded (Supplementary Figure 12D).

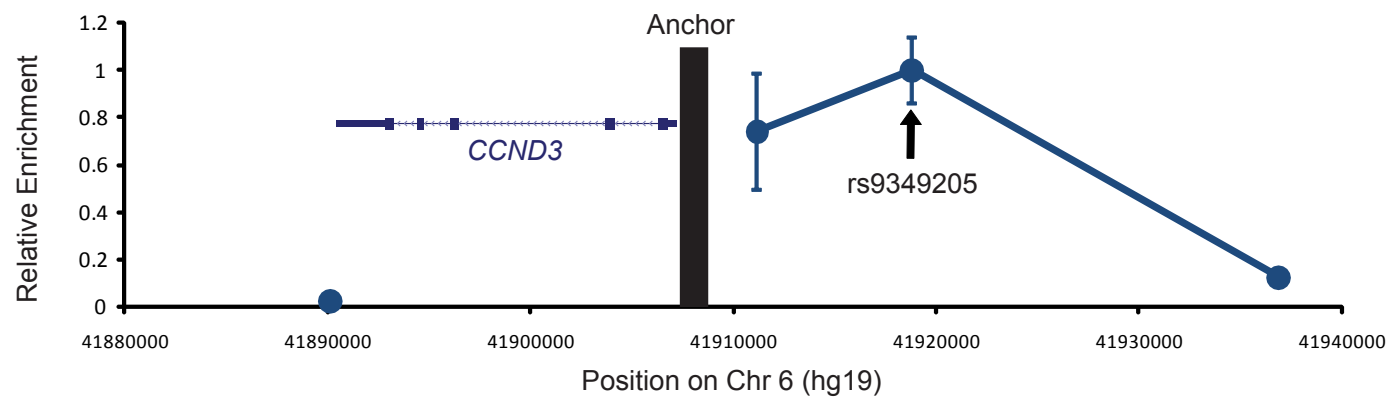Primitive erythroid cells derive from the yolk sac and still retain their nuclei at E12.5 (Ter119[+]Hoechst[+]) (McGrath and Palis 2008; Palis 2014) and also present with an increase in cell size as measured by forward and side scatter (Supplementary Figure 13E). This suggests a conserved role for cyclin D3 throughout all developmental stages of murine erythropoiesis. Of note, cyclin D3 deficient mice present with reduced body weight at 6-8 weeks of age: Male WT 24.3±2.1 g vs. KO 17.6±2.1 g (mean value ± standard deviation, WT n=9, KO n=6); Female WT 20.0±0.9 g vs. KO 14.1±0.3 g (WT n=9, KO n=5). It remains difficult to assess to what extent the reduced red blood cell count observed in fetal and adult mice may contribute to this observation in the context of a complete knockout, as nearly one third of gene knockouts in mice show alterations in body size, mass and growth (Smemo et al. 2014).
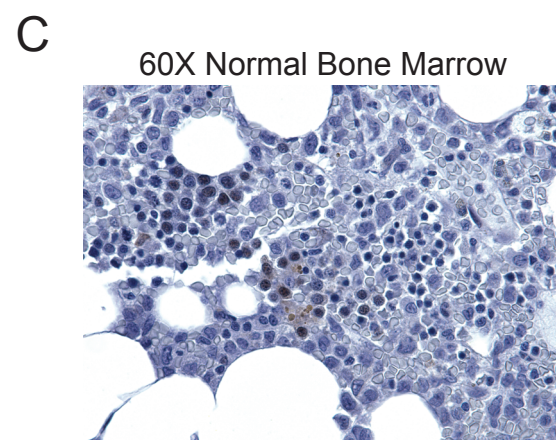
Supplementary Figure legends

Supplementary Figure 11. Analysis of *in vivo* erythropoiesis in 6-8 week old adult *Ccnd3*[-/-] mice. **(A)** Total bone marrow cellularity per femur (WT n=17, KO n=9). **(B)** Flow cytometry plots showing stained total bone marrow cells. Two representative plots are shown for CD71 and Ter119 for each genotype (Top = WT, Bottom = KO). A Hoechst and Ter119 plot is shown below with a gate indicating enucleated (Hoechst⁻) cells. Forward scatter (FSC) histogram of enucleated cells for each genotype is shown. **(C)** Total number of Ter119[+] cells in bone marrow per femur (WT n=17, KO n=9). **(D)** Spleen weight of 6-8 week old mice (WT n=9, KO n=5). **(E)** Representative images of WT and KO spleens. **(F)** Flow cytometry plots showing CD71 and Ter119 stained cells of whole spleen. **(G)** Frequency of CD71[+]Ter119[+] as gated in (F) (WT n=9, KO n=5).

Supplementary Figure 12. Erythroid cells are reduced in the fetal liver of *Ccnd3*[-/-] embryos. **(A)** Total fetal liver cellularity at indicated time-points. Note different scales. **(B)** Total number of Ter119[+] cells in fetal liver at indicated time-points. For (A) and (B): E13.5: WT n=2, KO n=3; E14.5: WT n=10, KO n=6; E16.5: WT n=4, KO n=6. **(C)** Ter119 immunostained fetal liver from E14.5 embryos is shown for each genotype at E14.5 at 20x magnification (Top = WT, Bottom = KO). Thymus tissue was used as a negative control (4x). **(D)** Representative flow cytometry plots showing CD71 and Ter119 stained cells of whole fetal liver at indicated time-points for each genotype (Top = WT, Bottom = KO).

Supplementary Figure 13. Erythroid cells are reduced in the peripheral blood of *Ccnd3*[-/-] embryos. **(A)** Total peripheral blood counts at indicated time-points (E13.5: WT n=2, KO n=3; E14.5: WT n=9, KO n=7; E16.5: WT n=4, KO n=6). Note different scales. **(B)** Mean corpuscular hemoglobin (MCH) from E14.5 peripheral blood cells (WT n=4, KO n=7). **(C)** Representative cytospin images from MayGrundwald-Giemsa stained E14.5 peripheral blood shown at 63x magnification. Scalebar = 5μm. **(D)** Hoechst and Ter119 plot of total fetal liver stained cells is shown with a gate indicating enucleated (Hoechst negative) cells. Ter119, CD71 and FSC histograms are shown for enucleated cells at indicated time-points. **(E)** Hoechst and Ter119 plot is shown with a gate indicating nucleated (Hoechst[+]) cells from E12.5 peripheral blood. FSC and side scatter (SSC) histograms are shown for nucleated cells from three independent samples for each genotype.
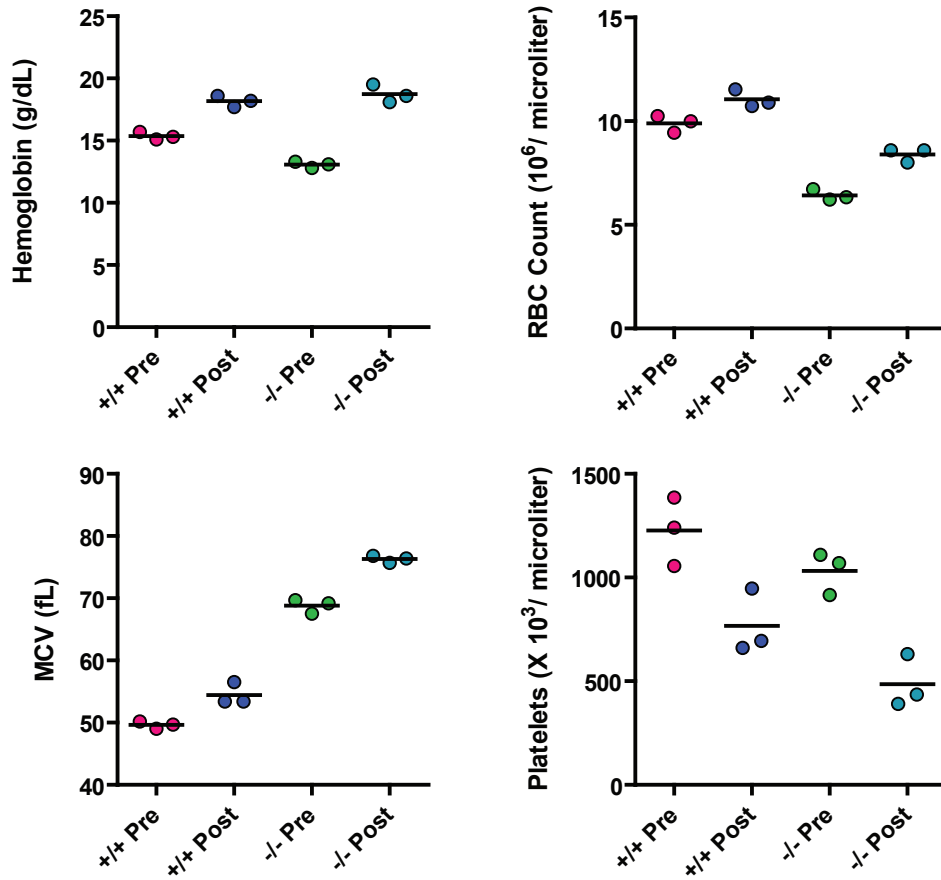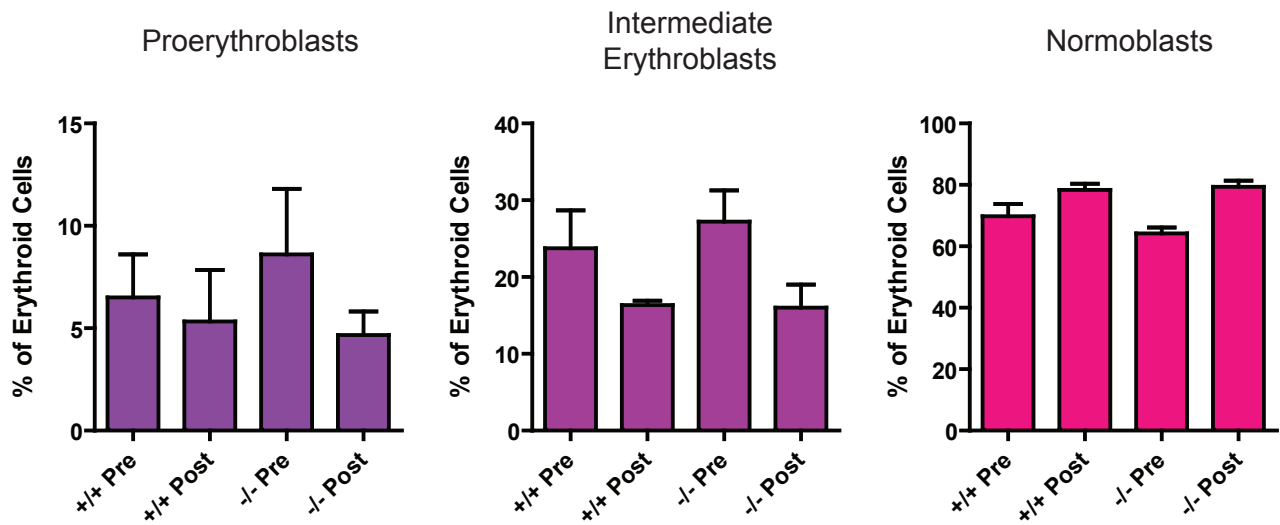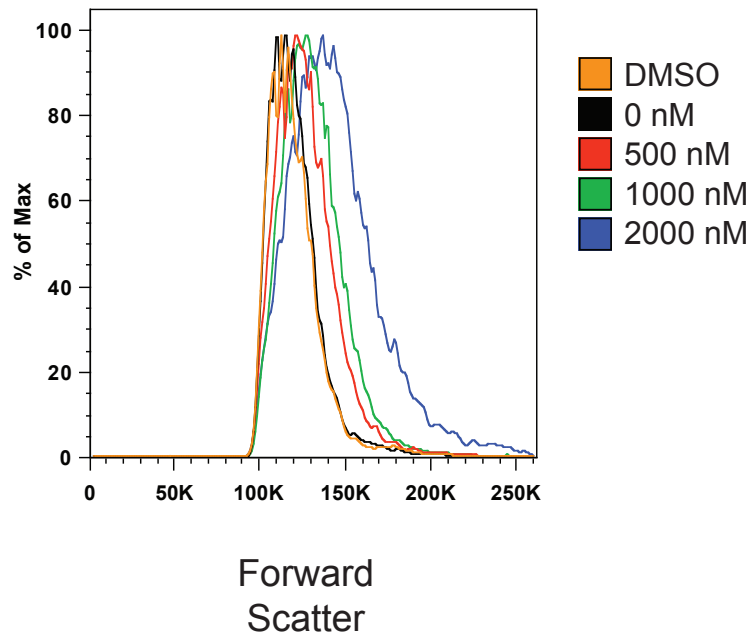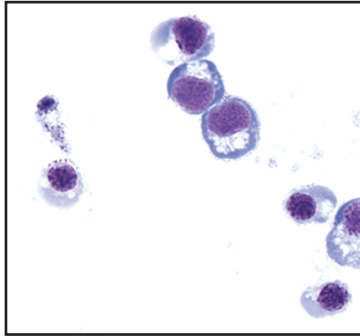
Supplementary Figure 11

Supplementary Figure 12

A

E13.5

Fetal liver cellularity x10⁷

WT — KO

E14.5

x10⁷

WT — KO

E16.5

x10⁷

WT — KO

B

E13.5

Ter119+ cells x10⁷

WT — KO

E14.5

x10⁷

WT — KO

E16.5

x10⁷

WT — KO

C

E14.5 Fetal liver

WT

KO

Thymus (negative control)

D

E13.5

E14.5

E16.5

WT

KO

CD71

Ter119
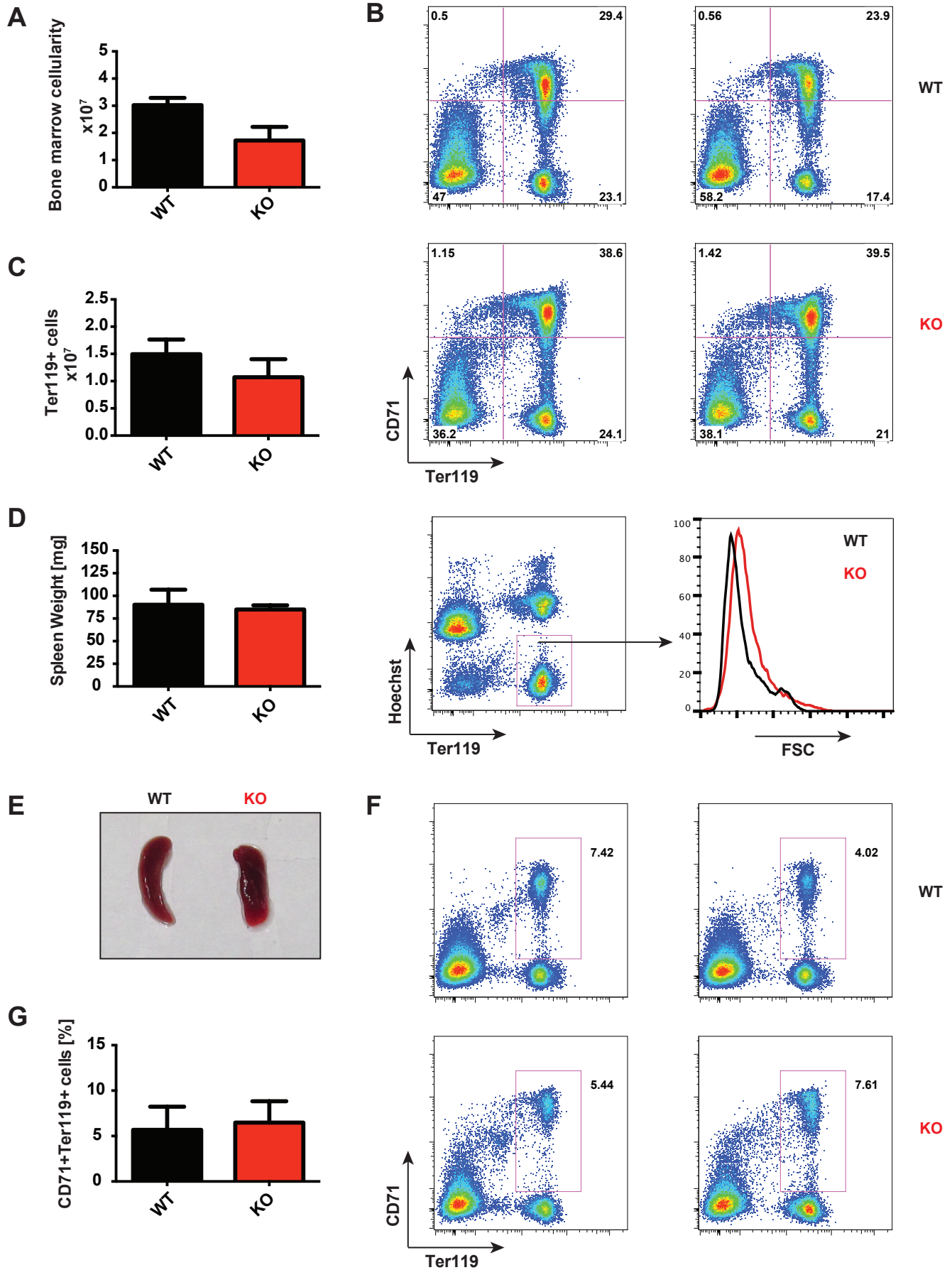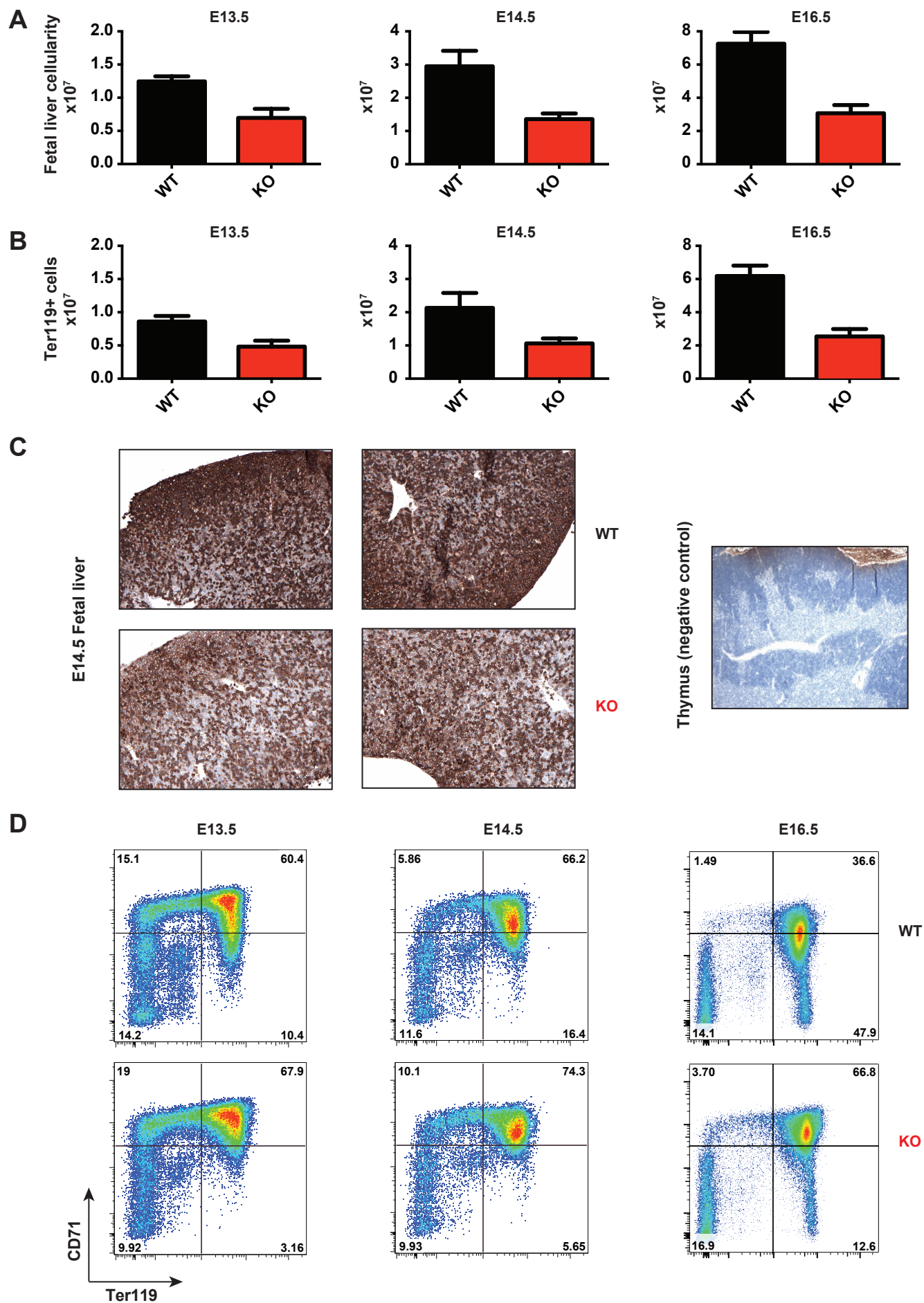
Supplementary Figure 13
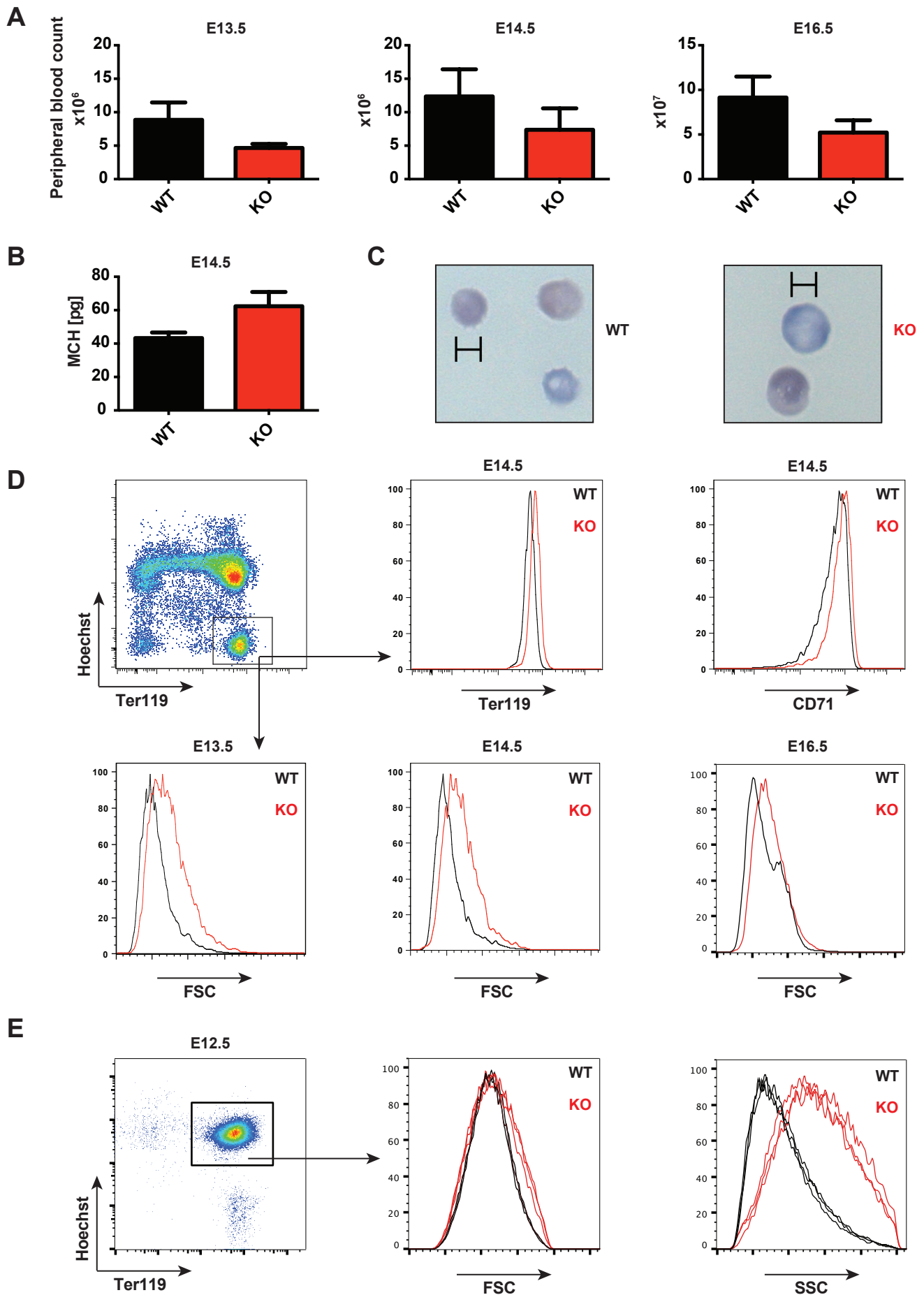
4. Discussion

Figure numbers refer to figures in the introduction or discussion, unless labeled M1 and M2 for the individual manuscripts. M1: Altered translation of GATA1 in Diamond-Blackfan anemia. M2: Cyclin D3 coordinates the cell cycle during differentiation to regulate erythrocyte size and number.

4.1 The Role of GATA1 in Diamond-Blackfan Anemia

GATA1 is an essential erythroid transcription factor, and numerous human mutations have been associated with anemia and / or thrombocytopenia (Pevny et al. 1995; Campbell et al. 2013; Crispino and Weiss 2014). Whole-exome sequencing approaches have identified *GATA1* mutations in patients with the clinical diagnosis of Diamond-Blackfan anemia, which previously has only been associated with genetic lesions in ribosomal proteins and is thus considered a classical ribosomopathy (Sankaran et al. 2012a). However, with no known molecular link of GATA1 to the ribosome it remained unclear, whether these mutations are in fact truly linked at the molecular level to the majority of DBA cases (Weiss et al. 2012). The altered translation of *GATA1* mRNA due to ribosomal haploinsufficiency now provides a common pathophysiology for the two distinct sets of genetic lesions identified in DBA (Figure 6) (Boultwood and Pellagatti 2014). Additionally, the reduced levels of erythroid-essential GATA1 protein provide an intriguing explanation for the specificity of the observed erythroid defect in DBA. GATA1 transcriptional targets are globally and specifically reduced in DBA cells with ribosomal protein mutations and reflect the decrease of GATA1 levels, which ultimately impair commitment and progression through the erythroid lineage. In fact, no other transcription factor gene set showed significant reduction in patient cells, demonstrating the specificity of the effect observed for GATA1 activity in DBA (M1 Supplementary Figure 16a).

These findings are further underlined by independent studies, which have recognized *GATA1* mutations in patients with the diagnosis of DBA, identifying *GATA1* as a true DBA gene (Klar et al. 2014; Parrella et al. 2014). Hollanda and colleagues identified identical mutations favoring the production of *GATA1s* in a family also diagnosed with anemia, but who additionally presented with neutropenia and mild multilineage dysplasia (Hollanda et al. 2006; Weiss et al. 2012). The clinical variability observed within the different patients with similar mutations is not understood but may be attributed to variable expression of *GATA1* (Sankaran et al. 2012a). Such differences have been seen in studies of mice with variable

Gata1 levels presenting with different phenotypes with respect to erythroid differentiation and proliferation (McDevitt et al. 1997; Pan et al. 2005). Similarly, the clinical heterogeneity in DBA patients with ribosomal protein mutations is also elusive. Genotype and phenotype correlate poorly, and siblings with identical mutations may present with different degrees of anemia (Vlachos et al. 2008; Lipton and Ellis 2009). In one particular case, a DBA patient received a hematopoietic stem cell transplant from an asymptomatic sibling. The transplant failed to cure the anemia, and subsequent analysis revealed the identical genetic defect within the same ribosomal protein gene in both siblings (Marcin Wlodarski, Pediatric Hematology, University Freiburg, personal communication).
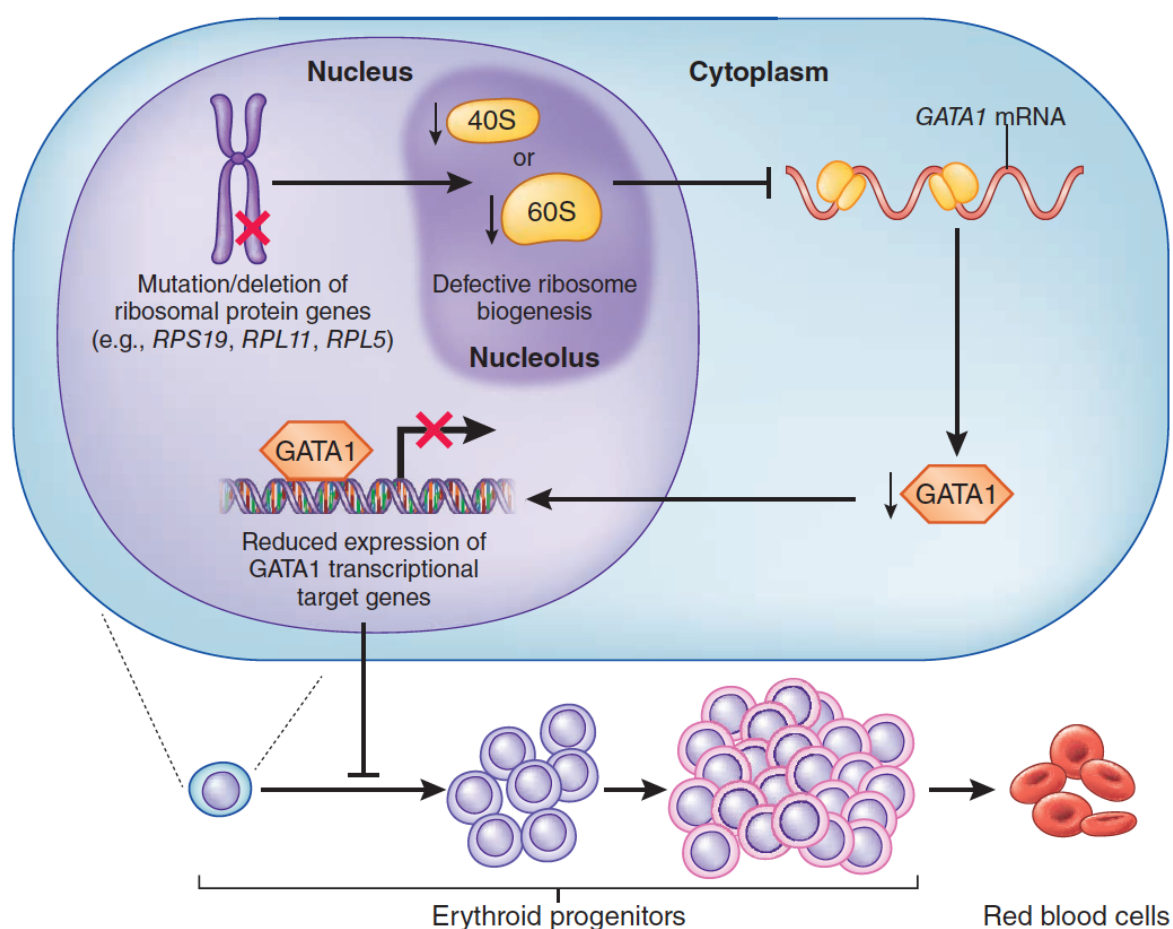
Figure 6 The role of GATA1 in DBA. Haploinsufficiency of ribosomal proteins (resulting from gene mutation or deletion or from experimental knockdown) affects the maturation of the small (40S) or large (60S) ribosomal subunits, leading to defective ribosome biogenesis. When the translation initiation potential is reduced, such as in the case of ribosomal protein haploinsufficiency, translation of the *GATA1* mRNA is impaired, as indicated by a decrease in *GATA1* mRNA abundance in polysomes, due to its highly structured 5'UTR, leading to a decrease in GATA1 protein levels. Global and specific reduction of GATA1 transcriptional target genes reflects the decreased GATA1 levels and plays a role in the erythroid differentiation defect in DBA. From Boultwood and Pellagatti 2014.

It is not known whether the variability in anemia and requirement for treatment may be attributed to varying GATA1 protein levels. This could be addressed by direct measurement of GATA1 by western blot or intracellular flow cytometry in erythroid progenitor cells from patients and healthy donors. Separation of defined progenitor cells by surface markers will minimize confounding factors and additionally allow comparing GATA1 levels across different stages of erythropoiesis. The intracellular method developed here showed variation with respect to GATA1 measurement in healthy donor cells (M1 Supplementary Figure 26, 27). It remains unclear whether this reflects physiologic donor-to-donor differences or may be attributed to the combination of surface markers used. Recent studies have shown different surface markers to be more suitable to reflect distinct stages of human erythropoiesis (Hu et al. 2013; Li et al. 2014a; Iskander et al. 2015). The combination of these markers with refined intracellular flow cytometry or potentially single-cell mass cytometry approaches (Bjornson et al. 2013; Lujan et al. 2015) may help to elucidate whether GATA1 levels are reduced and whether this can be correlated with the degree of anemia in DBA.

Alternatively, GATA1 expression may not be significantly different in erythroid progenitor cells from DBA patients and healthy controls. *In vitro* cultured erythroid cells from a DBA patient show differential surface marker profiles when compared to healthy donor cells (M1 Supplementary Figure 22). One set of cells remains negative for CD71 and erythroid marker CD235a, whereas a subset of CD71$^+$CD235$^+$ cells appears to differentiate with similar kinetics when compared to controls. Thus, a threshold amount of GATA1 may be required to initiate and progress through early stages of erythropoiesis and produce mature red blood cells (Pan et al. 2005). However, the number of progenitor cells sufficiently translating *GATA1* mRNA to commit and advance through the erythroid lineage may be significantly reduced in DBA and subsequently leads to anemia.

4.1.1 Knockdown of *GATA1* mimics Aspects of Diamond-Blackfan Anemia

Direct shRNA-mediated reduction of *GATA1* in primary human cells phenocopies a number of aspects of ribosome protein knockdown. Both result in reduced proliferation of erythroid cells, impaired differentiation towards the erythroid lineage and an increase in the frequency of non-erythroid cells (M1 Supplementary Figure 4, 9, 10). To more thoroughly investigate the transcriptional basis for these defects, stage matched human primary control

and *GATA1* knockdown cells were used for gene expression analysis (M1 Supplementary Figure 21-24). As expected, GSEA analysis revealed a specific reduction in the expression of GATA1 target genes as has been observed in DBA patient cells with ribosomal protein mutations (M1 Supplementary Figure 23b, M1 Figure 4c, d). When directly comparing transcriptional profiles of *GATA1* knockdown cells and DBA patient cells with ribosomal protein mutations, there was a significant overlap of differentially regulated genes (M1 Supplementary Figure 24e). Thus, GATA1 activity plays a critical role in the pathophysiology of DBA. Importantly, expression of GATA1 FL is able to rescue aspects of these defects more readily than the mutant, favoring production of GATA1s (M1 Supplementary Figure 17).

Two qualities of DBA with respect to the role of GATA1 FL in DBA may have to be distinguished. In case of ribosomal haploinsufficiency protein production of GATA1 FL and GATA1s is impaired, whereas the *GATA1* mutations observed in DBA present with significant reduction of predominantly GATA1 FL. Preliminary results investigating the expression of GATA1 in different human hematopoietic populations also reveal GATA1 FL to be significantly higher expressed in CD71$^+$CD235$^+$ erythroid cells than in other hematopoietic populations (M1 Supplementary Figure 28). These results point to a specific role of GATA1 FL in human erythropoiesis, which cannot be fully compensated for by the short form as patients with these mutations present with anemia (Sankaran et al. 2012a). Consistently, the *in vitro* culture of cells from a DBA patient only expressing GATA1s revealed it to be expressed at levels exceeding the combined level of both GATA1 isoforms in the controls, but erythroid differentiation remained severely compromised (O'Brien et al. 2014). However, the exact function of both GATA1 protein isoforms and the specific role of the N-terminus in erythropoiesis are not completely understood.

4.1.2 The Role of the N-terminus of GATA1

GATA1s lacks amino acids 1-83, encoding for the N-terminal transactivation domain named by its ability to activate reporter gene transcription in non-erythroid cells (Calligaris et al. 1995). Two recent studies utilized global transcriptome and chromatin occupancy analysis of GATA1 protein isoforms in murine cell lines and human induced pluripotent stem cells from patients with *GATA1* truncating mutations to define how the N-terminus regulates hematopoiesis (Byrska-Bishop et al. 2015; Chlon et al. 2015). Compared with GATA1 FL,

progenitor cells only expressing GATA1s appear impaired in both promotion of erythroid differentiation and maturation of committed erythroid cells. This is underlined by gene expression analysis showing a selective deficiency of the short form to activate erythroid specific gene expression. Consistently, chromatin occupancy analysis reveals impaired binding of GATA1s to regulatory regions of specific erythroid genes, but not at non-erythroid sites, including megakaryocytic and myeloid target genes. Equivalent occupancy of both GATA1 protein isoforms were observed at many sites, but regions bound more by GATA1 FL than GATA1s were highly enriched for genes with erythroid functions. Thus, the N-terminus appears to allow for selective DNA binding, even within the same lineage of cells. Together, these findings suggest that *GATA1s* mutations ultimately impair the expression of multiple erythroid important regulators and result in anemia (Byrska-Bishop et al. 2015).

How exactly the N-terminus allows for selective DNA binding to regulate chromatin occupancy of GATA1 remains an open question (Byrska-Bishop et al. 2015; Chlon et al. 2015). Although RUNX1 has been suggested to bind the N-terminus of GATA1, its loss primarily impairs megakaryocyte development (Elagib et al. 2003; Goldfarb 2009). The retinoblastoma protein also has been suggested to bind the same domain, but it primarily appears to affect terminal maturation and not early erythropoiesis (Walkley et al. 2008; Kadri et al. 2009). Previous studies of human *GATA1* missense mutations in the N-terminal zinc finger domain have shown to lead to diminished binding of GATA1 cofactors FOG1 and TAL1 (Campbell et al. 2013; Crispino and Weiss 2014). Likewise, the erythroid specific functions of the N-terminus of GATA1 are likely mediated or facilitated by the recruitment of other erythroid specific cofactors (Byrska-Bishop et al. 2015; Chlon et al. 2015). A number of proteins are known to interact with GATA1, and analysis of their overlap with chromatin occupancy sites shared with GATA1 FL may shed further light on how the red blood cell specific transcriptional program is regulated with respect to erythropoiesis and DBA. Importantly, fundamental differences in mice and human GATA1 biology and erythropoiesis will have to be considered (Pishesha et al. 2014; Ulirsch et al. 2014). The mutations favoring production of GATA1s in humans primarily result in anemia (Sankaran et al. 2012a; Ludwig et al. 2014), whereas a functionally identical mutation presents with thrombocytopenia in mice (Majewski et al. 2006).

4.1.3 GATA1 Protein Isoforms in Hematopoiesis

GATA1 is known to regulate hematopoiesis by activating specific genes of one cell type and repressing genes of other lineages (Orkin and Zon 2008). The MEP is the bipotential megakaryocyte and erythroid progenitor cell and gives rise to erythrocytes and megakaryocytes, which produce platelets / thrombocytes (megakaryopoiesis) (Orkin and Zon 2008). Interestingly, knockdown of ribosomal proteins in a culture of stem and progenitor cells leads to decreased GATA1 levels and an increased frequency of non-erythroid cells, including CD41[+] megakaryocytes (M1 Supplementary Figure 9, 10). In contrast, the ectopic expression of predominantly GATA1 FL in cells with knockdown of *RPS19* or from DBA patients with ribosomal protein mutations decreases the number of these non-erythroid cells (M1 Figure 4e, f, M1 Supplementary Figure 17). Although GATA1s appears deficient in sufficiently promoting erythroid commitment as well as progression through erythropoiesis, the activation of megakaryocytic genes by GATA1s appears unaltered and is consistent with predominant expression of a potentially modified short form in human CD41[+] megakaryocytes (M1 Supplementary Figure 28) (Byrska-Bishop et al. 2015). These observations indicate that GATA1 FL and GATA1s may play antagonistic roles in lineage commitment. Although both GATA1 protein isoforms are expressed in erythroid cells, the full-length version appears to be the predominant driver of erythropoiesis. It is not exactly known which protein isoforms of GATA1 are expressed in different hematopoietic cell types and how the expression/translation of individual or both protein isoforms is regulated.

Interestingly, GATA1s expression is also reduced upon ribosomal protein knockdown (M1 Supplementary Figure 9), and in a healthy donor total GATA1 expression levels in CD41[+] megakaryocytes also appear to be lower when compared to CD71[+]CD235[+] erythroid cells (M1 Supplementary Figure 28). These data indicate that high levels of GATA1 FL may be required for erythropoiesis, whereas lower levels of GATA1s may be sufficient to drive megakaryocyte and platelet development. Additionally, the failure to properly upregulate GATA FL for erythropoiesis in progenitor cells may promote alternative cell fates such as megakaryocyte development. Consistent with these observations, DBA patients often present with thrombocytosis at diagnosis (Willig et al. 1999), as do mice after induction of *Rps19* deficiency (Jaako et al. 2011). These findings are in agreement with the increased frequency of CD41[+] megakaryocytes after ribosomal protein knockdown in cell culture models of DBA (Flygare et al. 2005; Ebert et al. 2008; Ludwig et al. 2014).

4.1.4 *GATA1* Gene Therapy and small Molecules in the Treatment of DBA

Lentiviral hematopoietic stem cell (HSC) gene therapy and targeted genome editing of human HSCs provide a rationale for targeted therapies in patients with congenital hematopoietic defects (Aiuti et al. 2013; Genovese et al. 2014b). The finding of impaired GATA1 production as a final common pathway in DBA offers the possibility of such treatment approaches to any patient with DBA regardless of their genetic lesion. Allogeneic HSC transplantation has been conducted in DBA patients and is the only curative option for DBA (Roy et al. 2005; Mugishima et al. 2007; Fagioli et al. 2014). However, it is associated with significant morbidity and mortality and a matched donor is not readily available for the majority of patients. Therefore, the prospect of avoiding harsh conditioning and immunosuppressive approaches using gene therapy of autologous HSCs appears attractive in DBA. Development of clinically usable vectors would need to allow for regulated expression of *GATA1* in hematopoietic cells, while minimizing insertional mutagenesis. Importantly, *Gata1* regulatory elements in mice have been identified, which are capable of driving regulated expression of marker genes solely in the cell types where *Gata1* is normally expressed and are sufficient to allow appropriate rescue of knockout mice using *Gata1* cDNA (Shimizu et al. 2013; Takai et al. 2013). Thus, these mouse elements and their human homologues may be used to drive expression of human *GATA1* cDNA (or fluorescent reporters) in human HSCs using lentiviral gene delivery. As transgene expression can result in toxicity at the HSC stage (Gentner et al. 2010) and ectopic *GATA1* expression in HSCs results in the loss of self-renewal activity (Iwasaki et al. 2003), the functionality of these regulatory elements in human HSCs will require careful monitoring. MicroRNA-126 is expressed in HSCs and early progenitors, and the incorporation of microRNA-126 target sequences into gene expression vectors has successfully suppressed ectopic gene expression in HSCs, while maintaining expression in differentiated cells (Gentner et al. 2010). Similar approaches may also be exploited for *GATA1* gene therapy. *In vitro* culture of transduced HSCs supporting multi-lineage hematopoiesis will allow assessing the specificity of expression and measuring the effect of the vectors upon differentiation (Figure 7). This approach may also be applied to assess rescue of differentiation from cells derived from patients with DBA. Furthermore, long-term multi-lineage repopulation of manipulated human cells may be conducted in severe combined immune deficient mice to test safety and efficacy of the outlined approach (Genovese et al. 2014b). Overall, gene therapy approaches involving *GATA1* will require careful assessment and additional studies to translate to the clinic. Nevertheless, the prospect of using a single vector for all cases of DBA, potentially even

those with unknown genetic lesions, presents an attractive option motivating immediate research and investigation.
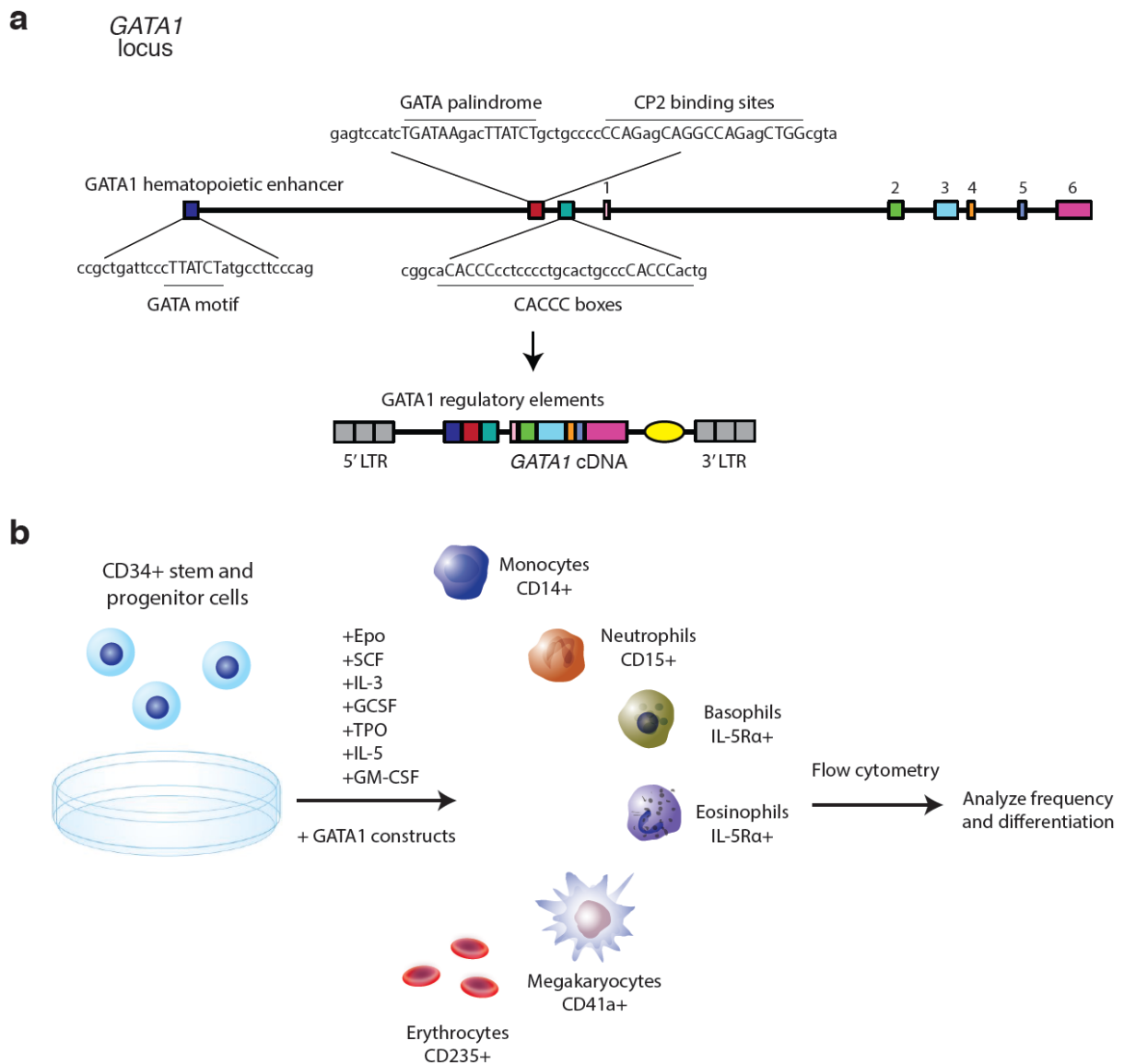
Figure 7 Gene therapy in Diamond-Blackfan anemia. **(a)** The endogenous *GATA1* locus with cis-regulatory elements is shown above. A simplified vector scheme with regulatory elements to allow regulated expression of *GATA1* is illustrated below. **(b)** Simplified scheme to illustrate how efficacy of different vectors would be monitored. Human CD34+ stem and progenitor cells would be transduced with control and *GATA1* constructs and differentiated *in vitro* under conditions favoring growth of diverse hematopoietic cell types. Frequency and differentiation would be assessed using flow cytometry using indicated markers to monitor potential skewing of differentiation with the vectors used.

Simultaneously, the search continues for small molecules and drugs, which may alleviate the anemia in DBA. A number of compounds have been clinically tested but so far only corticosteroids have been proven to be therapeutically beneficial in DBA (Vlachos et al. 2008; Sjogren and Flygare 2012). A recent study identified PPARα agonists to synergize with

corticosteroids to stimulate self-renewal of early erythroid progenitor cells in healthy and *RPS19* knockdown cells (Lee et al. 2015). These and other compounds including leucine await further clinical assessment in DBA (Sjogren and Flygare 2012; Dussiot et al. 2014; Suragani et al. 2014). With respect to direct modulation of the ribosome it will be interesting to follow the path of Ataluren for the treatment of patients with genetic disorders caused by premature stop codons due to nonsense mutations. Ataluren is a small molecule suggested to modulate the ribosome by selectively inducing ribosomal read-through of premature but not terminal termination codons (Welch et al. 2007), although the exact mechanism has been called into question (McElroy et al. 2013). A variety of genetic disorders may be caused by premature stop codons and thus Ataluren may be of value for a fraction of patients across a spectrum of diseases. Trials have been primarily conducted in patients with Duchenne muscular dystrophy and cystic fibrosis and have reported beneficial effects in certain instances (Sermet-Gaudelus et al. 2010; Wilschanski et al. 2011; Bushby et al. 2014; Kerem et al. 2014; Haas et al. 2015). It remains to be seen whether it will be possible to specifically modulate the translational defects in DBA by the use of small molecules.

4.1.5 The Role of p53 and Modulation by GATA1

A prevalent model of DBA suggests p53 activation downstream of the specific loss of ribosomal proteins to result in cell cycle arrest and apoptosis (Figure 3) (Weiss et al. 2012). However, the role of p53 through the proposed mechanism remains controversial, as knockdown of ribosomal proteins RPL5 or RPL11 do not lead to activation of p53 (Bursac et al. 2012; Teng et al. 2013). Both genes are frequently mutated in patients with DBA (Gazda et al. 2008), and it is unclear whether p53 activation is a core aspect of the pathophysiology in DBA. Importantly, 4EGI-1 treatment did not result in major perturbation of ribosomal protein levels, but phenocopied *in vitro* characteristics of DBA and impaired production of GATA1 (M1 Figure 3e-f, M1 Supplementary Figure 11). Furthermore, GATA1 has been shown to interact and inhibit p53 (Trainor et al. 2009). Consistently, the knockdown of GATA1 increases p53 levels in human primary erythroid cells (M1 Supplementary Figure 20). In contrast, a global upregulation of p53 transcriptional targets could not be observed in DBA patient cells with ribosomal protein mutations (M1 Supplementary Figure 16b). Nevertheless, transcriptional analysis reveals GATA1 to inhibit the expression of proapoptotic genes and activate antiapoptotic genes (Cheng et al. 2009; Weiss et al. 2012). Loss of antiapoptotic LRF, a direct target of GATA1, has been shown to lead to lethal anemia in embryos

independent of p53 (Maeda et al. 2009). Thus, the reduced translation of *GATA1* in erythroid progenitor cells of DBA may lead to apoptosis by a number of mechanisms, some of which may be dependent as well as independent of p53. Additional studies will be required to delineate the role of p53, its exact interplay with GATA1, and the extent to which loss of ribosomal proteins contribute to p53 activation and the erythroid failure in DBA.

4.2 Altered Translation in Diamond-Blackfan Anemia

Impaired translation of a number of transcripts has been implicated in DBA, including *GATA1*, *BAG1*, *CSDE1* and *BCAT1* (Horos et al. 2012; Ludwig et al. 2014; Pereboom et al. 2014). Beyond these, a variety of additional transcripts are likely to be affected in the context of ribosomal haploinsufficiency and may contribute to the anemia (Horos et al. 2012). However, only *GATA1* mutations have been identified in patients with DBA as sufficient to cause disease. As GATA1 FL is essential for erythroid commitment and maturation, altered translation of its mRNA may be more detrimental than for other transcripts. The exome aggregation consortium (ExAC) provides exome-sequencing data for over 60,000 unrelated individuals. In agreement with its essential role, *GATA1* mutations appear less tolerated as suggested by ExAC data reporting fewer missense or loss of functions mutations for *GATA1* than for *BAG1* or *BCAT1* (http://exac.broadinstitute.org/, May 2015).

The 5'UTR of *GATA1* is predicted to be highly structured and have a higher folding energy than UTRs of similar length and reporter expression analysis show it to restrict efficient translation (M1 Supplementary Figure 13). This is consistent with other studies reporting an increased initiation potential by the cell to translate transcripts with highly structured 5'UTRs (Babendure et al. 2006; Moerke et al. 2007; Shah et al. 2013). Translation initiation would be further compromised upon ribosomal haploinsufficiency, as the number of available ribosomes is a major determinant of initiation (Shah et al. 2013). Thus, in DBA transcripts with highly structured 5'UTRs are less likely to be translated and are out-competed away by mRNAs with short unstructured 5'UTRs, which show less stringent requirements for translation initiation. These observations provide a mechanistic rationale to explain a selective reduction in the translation of *GATA1* mRNA in the context of ribosomal haploinsufficiency. An unusually long 5'UTR has also been attributed to impair the translation of *BCAT1*, whereas IRES elements in the 5'UTR have been held responsible for *BAG1* and *CSDE1* (Horos et al. 2012; Pereboom et al. 2014). Both studies forwarded polysome fractions for

transcript identification by microarrays and found selected transcripts to be differentially translated but have not been able to elucidate general mechanistic rules for how ribosomal haploinsufficiency influences the translational profile of a cell.

Two approaches may aid in the identification of additional transcripts and their underlying characteristics, which are subject to impaired translation. Thorough comparative proteomic approaches of control and cells with knockdown of ribosomal genes may identify differentially expressed proteins (Schwanhausser et al. 2011; De Keersmaecker et al. 2015). Ribosome profiling represents an alternative technique by generating ribosome protected mRNA fragments to determine transcripts that are actively translated (Guo et al. 2010; Ingolia et al. 2011). Ribosome profiles of control and ribosomal protein knockdown cells may then be compared to identify transcripts that are depleted or increased on actively translating ribosomes. Subsequent analysis of their 5'UTRs and other mRNA characteristics may aid to generate more mechanistic insight with respect to altered translation of *GATA1* mRNA and other transcripts in DBA. An erythroid cell line may be suitable for this approach, unless stage-matched primary human cells are sorted, as ribosomal protein knockdown mediated changes in lineage and cell fate may confound the analysis. Generally, changes in gene transcription, including those due to reduced GATA1 transcriptional activity, may hinder proper analysis and justify performing experiments shortly after ribosomal protein knockdown. Individual depletion of proteins from the small or the large subunit would allow investigation into whether a common or a distinct set of transcripts is depleted from translating ribosomes. Similar profiles would strengthen the notion of a quantitative defect. Although dosage effects will have to be considered, distinct profiles may indicate specific functions for individual ribosomal proteins in translating a subset of transcripts.

4.2.1 Experimental Challenges

It is not clear whether inaccurate annotation of 5'UTRs prohibited identification of explicit rules of translational regulation in DBA in prior studies (Horos et al. 2012; Pereboom et al. 2014). Many genes are expressed in multiple tissues and different alternative transcriptional start sites result in the annotation of different 5'UTRs. Rapid amplification of the 5' cDNA ends (5'RACE) to determine the 5'UTR of *GATA1* transcripts also identified a previously unreported 5' end. Thus, proper annotation of 5' ends, possibly across different stages of erythropoiesis, may be required for proper interpretation and analysis and may be

achieved by cap analysis gene expression (CAGE) (Shiraki et al. 2003; Motakis et al. 2014). CAGE captures 5'ends of mRNA and generates fragments forwarded for sequencing to allow locating exact transcription start sites. Nonetheless, a complete knowledge of UTRs may not be immediately revealing, as even a thorough analysis of a library of short 10 bp UTRs followed by computational modeling only allows prediction of around 70 % of the observed expression variability (Dvir et al. 2013). However, most UTRs are much longer and deciphering the regulatory code of 5'UTRs may thus require additional measures, including going beyond structural predictions and experimentally verifying RNA structural motifs (Siegfried et al. 2014). It is also unclear how the rest of the mRNA molecule may contribute to translational efficiency. Additionally, RNA binding proteins may contribute to the structure and translation of mRNA transcripts and changes in RNA binding proteins levels due to ribosomal defects may feedback on translation. Thus, the elucidation of a definitive translational code as embedded within mRNA transcripts represents a significant challenge.

The sensitivity of GATA1 levels in DBA may also be modulated by a number of additional factors. GATA1 is known to autoregulate its own transcription (Kobayashi et al. 2001; Nishikawa et al. 2003), and the RNA binding protein Elavl1 stabilizes *gata1* mRNA in zebrafish (Li et al. 2014b). GATA1 has also been suggested to have a short half-life (Bibikova et al. 2014). Thus, reduced translation of *GATA1* mRNA, a particularly short protein half-life, and reduced transcription or mRNA levels may ultimately corrupt sufficient accumulation of GATA1 to allow proper initiation and progression through erythroid differentiation. However, *GATA1* mRNA levels appeared similar in early erythroid cell populations of DBA patients with ribosomal protein mutations and healthy controls (M1 Figures 2d, 4b). Thus, autoregulation of *GATA1* or the stability of its mRNA may not play a significant role until later in differentiation. Furthermore, it remains unclear whether GATA1 protein stability may be altered in DBA. HSP70 has been suggested to stabilize GATA1 in human erythropoiesis (Ribeil et al. 2007), and altered HSP70 levels may consequently negatively influence GATA1 stability and activity. Overall, a number of factors may potentially contribute to the sensitivity of GATA1 levels or in fact any protein and thorough analysis of ribosome profiling may ideally be coupled with proteomics approaches to draw definitive conclusions.

4.2.2 The Role of mTOR and the translational Response in DBA

Leucine has been shown to improve anemia in one patient and in cellular models of DBA by modulating mTOR signaling (Pospisilova et al. 2007; Jaako et al. 2012; Narla et al. 2014). The mTOR kinase is the catalytic subunit of two complexes, mTORC1 and 2 (Laplante and Sabatini 2012). A major function of the mTORC1 complex is the regulation of protein synthesis. However, the effects of mTORC1 on translation are pleiotropic and pharmaceutical modulation of this pathway has shown discordant results with respect to mechanisms of translational regulation (Hsieh et al. 2012; Thoreen et al. 2012). A primary mode of action appears to be the modulation of the cap-binding proteins, including eIF4E and the inhibitory eIF4E-binding proteins (EIF4BP), to modulate translation initiation complex formation. However, in preliminary experiments, Torin1-mediated inhibition of mTORC1 (Thoreen et al. 2009) impaired proliferation but did not show obvious reduction of GATA1 protein expression levels in the erythroid K562 cell line (M1 Supplementary Figure 29). In contrast, inhibition of the eIF4E/eIF4G complex by 4EGI-1 (Moerke et al. 2007) had profound effects on GATA1 protein expression and erythroid differentiation and phenocopied aspects of ribosomal protein knockdown and DBA (M1 Figure 3d-f, M1 Supplementary Figure 11). The discrepancy of these findings remains unclear. However, modulation of mTOR signaling may have more pleotropic effects and its role in DBA and whether leucine has an effect on translation of *GATA1* and/or also acts by generally modulating the translational machinery will be the subject of future studies.

Overall, it has not been carefully scrutinized whether and how a cell with ribosomal haploinsufficiency may respond to overcome the ribosomal defect. A recent study identified activation of eukaryotic translation elongation factor 2 kinase (eEF2K) with a consequent inhibition of translation elongation in response to knockdown of ribosomal proteins (Gismondi et al. 2014). A decrease in global protein synthesis was accompanied by increased recruitment of a group of transcripts with 5' terminal oligopyrimidine tract (TOP) mRNAs to the ribosome. A TOP motif at the 5' end can be identified in mRNAs of all ribosomal proteins and translation elongation factors in vertebrates and this response would thus help to restore a sufficient level of ribosomes in the cell (Gismondi et al. 2014). However, no global analysis of transcripts and their recruitment to the ribosome was conducted and it remains unknown whether this applies to all mRNAs with a TOP motif and is thus a general response. A number of proteins are known to regulate aspects of translation and are subject to posttranslational modification (Sonenberg and Hinnebusch 2009; Hsieh et al. 2012; Laplante

and Sabatini 2012; Thoreen et al. 2012). Their activity or changes thereof upon ribosomal protein knockdown have not been thoroughly investigated and may provide further clues towards elucidating general mechanisms in DBA.

4.2.3 Quantitative and qualitative Defects in Translation

Quantitative and qualitative defects have been discussed with respect to DBA (Xue and Barna 2012). Although multiple ribosomal proteins have been identified as mutated in DBA, a specific function of one or multiple of these proteins with respect to aiding in the translation of a subset of mRNAs has not yet been demonstrated. Furthermore, their distribution over the ribosomal surface and the small and large ribosomal subunit make a common function with respect to the recruitment of specific mRNAs unlikely (Ben-Shem et al. 2011; Sulima et al. 2014). Considering the ribosome biogenesis defects, altered polysome profiles, and the reduced protein synthesis capacity of DBA cells (Cmejlova et al. 2006; Garcon et al. 2013), the majority of the evidence points to a quantitative defect in translation in DBA. As pointed out before, a reduced number of available ribosomes would limit the translation initiation capacity, a rate limiting step in translation, and transcripts with a higher threshold for initiation including *GATA1* are less likely to be efficiently translated (Babendure et al. 2006; Moerke et al. 2007; Shah et al. 2013). However, a special role in translation for individual ribosomal proteins may not be excluded. *RPL5* and *RPL11* mutations are more frequently associated with cleft palate and abnormal thumbs (Gazda et al. 2008), which may implicate a unique role for these proteins in the respective cell types (Xue and Barna 2012; Pereboom et al. 2014).

A recent study identified a mutation in the X-linked gene *TSR2* in two cousins with DBA, who also presented with physical findings reminiscent of another ribosomopathy named Treacher Collins syndrome (TCS) (Gripp et al. 2014). TCS patients primarily present with craniofacial abnormalities that are also observed in DBA. Both syndromes show clinical heterogeneity and overlap and it is probable that a minority of patients with TCS have a DBA genotype (Lipton and Ellis 2009). Interestingly, in yeast Tsr2 interacts with Rps26, shields it from proteolysis, and coordinates its transfer to the assembling ribosome (Schutz et al. 2014). Depletion of Tsr2 reduces growth and alters polysome profiles, both features of cellular models of DBA (Garcon et al. 2013). *RPS26* is also commonly mutated in DBA (Doherty et al. 2010), and it seems plausible to hypothesize that *TSR2* mutations may impair RPS26

stability in human cells. Although this will require additional studies, the knockdown of TSR2 would be expected to ultimately result in ribosome biogenesis defects, reduced ribosome abundance and impaired translation of *GATA1*.

In addition, a number of qualitative defects have been implicated with respect to lesions affecting the ribosome and the altered translation of specific transcripts. Heterozygous loss of function mutations in *Rpl38* have been identified to result in tissue-specific skeletal patterning defects in mice (Kondrashov et al. 2011). Unlike in DBA, global protein synthesis is unchanged and polysome profiles appear unaltered. Interestingly, the translation of a set of Homebox (*Hox*) mRNAs was perturbed, and loss of function studies of specific *Hox* mRNAs phenocopied *Rpl38* knockdown. The analysis of a variety of mice with ribosomal protein deficiencies, which are accompanied by global changes in protein synthesis did not present with patterning defects (Kondrashov et al. 2011). Subsequent analysis identified RNA elements in the 5'UTRs of *Hox* mRNAs, which resemble viral IRES and are essential for translation of Hox proteins (Xue et al. 2015). Together, these observations indicate a special role of Rpl38 to facilitate *Hox* transcript recruitment to the ribosome.

Haploinsufficient *RPSA* mutations have been reported in isolated congenital asplenia, which is characterized by the absence of a spleen with no other developmental defects (Bolze et al. 2013). RPSA is part of the small subunit of the ribosome and is ubiquitously expressed. It also participates in pre-ribosomal RNA processing (O'Donohue et al. 2010), but no ostensible processing defects could be identified in activated lymphocytes from patients with *RPSA* mutations. It is not known whether translation capacity is reduced in the cells of these patients. Patients with *RPSA* mutation do not present with anemia or any other morphological defects associated with DBA. Conversely, spleen abnormalities have not been frequently reported in patients with DBA. Thus, the function of RPSA appears distinct from ribosomal proteins associated with DBA. A role for RPSA in spleen development is not known, but it potentially aids in the translation of spleen essential transcripts. A variety of transcription factors control spleen development and further investigation will be required to elucidate the nature of the ribosomal defect in isolated congenital asplenia (Bolze et al. 2013).

Future studies will be required to delineate if additional ribosomal proteins promote specialized translation of select transcripts and thus provide another layer of control of gene

expression (Xue and Barna 2012; Xue et al. 2015). Still, it remains unclear whether these findings represent a specialized ribosome with individual ribosomal proteins allowing for the efficient translation of a subset of mRNAs. Alternatively, ribosome composition may not significantly differ across tissues, but tissue-specific defects due to altered translation may simply reflect tissue-specific expression of these transcripts. Proteomics based approaches may aid in the identification of differentially composed ribosomes (Reschke et al. 2013) and presents an interesting question for future studies of translation in physiology and disease.

4.3 Ribosomal Haploinsufficiency, Remission and Oncogenesis

Somatic mutations in ribosomal protein genes, including *RPL5*, *RPL10*, *RPL11* and *RPL22*, have been identified in cases of leukemia and solid tumors (Wang et al. 2011; Rao et al. 2012; De Keersmaecker et al. 2013; De Keersmaecker et al. 2015). Interestingly, congenital mutations in *RPL5* and *RPL11* are also found in DBA and a predisposition to neoplasia has been reported in patients with DBA (Vlachos et al. 2012). It is not known how ribosomal haploinsufficiency may present with a hypoproliferative defect (i.e. anemia) but also contribute to oncogenesis and hyperproliferation (De Keersmaecker et al. 2015). Changes in translation of specific transcripts may alter a cell's characteristics to proliferate or respond to environmental stimuli. Alternatively, it may acquire molecular properties, which repair or bypass the primary lesion. Interestingly, some DBA patients may go into remission without requiring further corticosteroid treatment or regular transfusions (Vlachos et al. 2008; Ruggero and Shimamura 2014). Potentially, compensatory somatic mutations in specific genes in hematopoietic stem cells may compensate for pathologic ribosomal protein mutations. These single clones may present with a proliferative advantage, especially in an otherwise hypo-proliferative bone marrow and eventually lead to clonal hematopoiesis, as has been described in the development of hematologic cancers (Genovese et al. 2014a; Jaiswal et al. 2014). In a genetic skin disorder named ichthyosis such somatic loss of disease-causing mutations have been described to revert phenotypes (Choate et al. 2010; Choate et al. 2015). In a yeast genetic model, a R98S mutation in *RPL10* as identified in human leukemia led to ribosome biogenesis defects and impaired proliferation (Sulima et al. 2014). Growth over many generations eventually selected for cells with normal growth, which had acquired mutations in a ribosome biogenesis factor. Interestingly, the second site mutation rescued hypo-proliferation but did not correct all the underlying defects conferred by the original mutation (Sulima et al. 2014; De Keersmaecker et al. 2015). Thus, it seems plausible that

similar mutations in DBA may contribute to alleviate the anemia, but ultimately also predispose to oncogenesis. Currently, no such modifier genes have been identified in DBA and the comparison of the exome/genome of hematopoietic cells of individual patients before and after remission as well as tumor and non-tumor cells may reveal such compensatory genetic lesions relevant to overcome ribosomal defects (De Keersmaecker et al. 2015).

4.4 Ribosomal Proteins, *GATA1*, *TSR2*…

Despite extensive whole-exome sequencing efforts the underlying genetic defect in a significant number of DBA patients remain unidentified (Sankaran et al. 2012a; Ludwig et al. 2014). The majority of patients present with lesions in ribosomal protein genes (Weiss et al. 2012) and two patients have been identified with *TSR2* mutations, which may potentially alter ribosome assembly (Gripp et al. 2014; Schutz et al. 2014). Overall, *GATA1* mutations are rarely observed in DBA and is probably reflective of the specific nature of the mutations: splice site and start codon mutations have been described in DBA and there may only be very few mutations leading to predominant or exclusive expression of GATA1s (Sankaran et al. 2012a; Klar et al. 2014; Ludwig et al. 2014; Parrella et al. 2014). Thus, it seems plausible that the remaining lesions may ultimately affect ribosomal proteins or biogenesis (Farrar et al. 2014). Whole-exome sequencing does not cover distal regulatory regions activating transcription of ribosomal protein genes and additional mutations or deletions may ultimately decrease ribosomal protein expression and subsequently lead to DBA. Another intriguing feature is that sixteen ribosomal genes have been associated with DBA, but present only a fraction of the 81 ribosomal proteins which make up the human ribosome (De Keersmaecker et al. 2015). Some mutations also appear more frequent than others. For example, genetic defects in *RPS19* can be identified in up to 25 % of patients with DBA (Weiss et al. 2012). Also, many of the genes associated to ribosomopathies encode proteins located on the cytoplasmic/solvent-accessible surfaces of the ribosome (Ben-Shem et al. 2011; Sulima et al. 2014; De Keersmaecker et al. 2015). They do not appear to stabilize interdomain ribosomal RNA interactions or locate in regions, which mediate mRNA decoding or peptidyl transfer leaving the ribosome functional enough to allow viability (De Keersmaecker et al. 2015). In contrast, the knockdown of a number of "core" ribosomal proteins may have more detrimental consequences and would be less compatible with cell survival and life. In fact, it is not known whether a significant frequency of miscarriages may be attributed to disorders of the

ribosome, and patients that present postnatally with DBA and other syndromes may only represent one end of the clinical spectrum of ribosomopathies.

4.5 The Role of Cyclin D3 in Erythropoiesis

While GWAS have revealed numerous loci associated with thousands of human traits, only marginal insight into the biology of these associations has been provided. The outlined findings indicate that variable levels of cyclin D3 determine the number of cell divisions in terminal erythropoiesis without compromising differentiation. Thus, red blood cell count and volume are inversely regulated. By consolidating chromatin immunoprecipitation data from human primary erythroid cells, rs9349205 was identified within an erythroid-specific enhancer element bound by GATA1 and TAL1 upstream of *CCND3* (M2 Figure 1A). A chromatin conformation capture (3C) assay verified binding to the promoter of the *CCND3* gene, suggesting that rs9349205 in fact regulates the expression of cyclin D3. Although reporter assays in erythroid cells demonstrate differential activity of the two alleles of rs9349205, no differential binding of TAL1 at this locus could be ascertained using chromatin immunoprecipitation from individuals heterozygous for this variant (M2 Figure 1B). Potentially, TAL1 may be recruited by GATA1, binding in its immediate vicinity, but is unable to activate transcription to the same extent, potentially due to conformational constraints. The enhancer element appeared to be erythroid specific as the reporter constructs proved to be inactive in non-erythroid cells, consistent with the binding of erythroid transcription factors GATA1 and TAL1 (M2 Figure 1C, D) (Pevny et al. 1995; Shivdasani et al. 1995). Directed deletion of this element using genome editing approaches would help to ascertain the specificity of this enhancer for erythroid cells (Bauer et al. 2013). When specific, cyclin D3 expression would be reduced upon deletion of this region in erythroid cells, whereas the same deletion in a different cell type would have no effect.

In vivo analysis of *Ccnd3*[-/-] mice demonstrates macrocytosis (erythrocytes with an increased volume) throughout different developmental stages of red blood cell production (M2 Supplementary Figure 11-13). These finding suggest a conserved role of cyclin D3 in primitive and both fetal and adult definitive erythropoiesis. Despite significantly reduced red blood cell counts, their larger volume is consistent with increased hemoglobin levels per cell and preserves overall hematocrit and hemoglobin levels in the adult mice. Thus, the anemia is mild, and as no severe differentiation defects have been observed, there appears no requirement to further stimulate erythroid output by means of compensatory extramedullary

erythropoiesis in the spleen. Consistent with this, $Ccnd3^{-/-}$ mice do not present with splenomegaly or increased erythropoiesis in the spleen, as can be observed in mice with severe erythroid differentiation defects (Socolovsky et al. 2001).

D-type cyclins bind to CDK4 and CDK6 to promote G1 to S phase progression (Malumbres and Barbacid 2009) and thus the inhibition of CDK4/6 mimicked aspects of knockdown/knockout of cyclin D3 (M2 Supplemental Figure 8). In the same fashion a recent study deleted $Cdk2$ and $Cdk4$ in murine hematopoietic cells, which led to increased erythrocyte size and reduced red blood cell count that phenocopied major characteristics of CDK4/6 inhibitor treatment and $Ccnd3^{-/-}$ mice presented here (Jayapal et al. 2015). Variants close to the cyclin D2 gene have also been associated with erythroid traits (van der Harst et al. 2012), and considering the severe anemia of $Ccnd2^{-/-}Ccnd3^{-/-}$ double knockout mice (Ciemerych et al. 2002), it seems plausible to assume that varying levels of expression of cyclin D2 may similarly influence MCV and RBC in humans. However, the role of cyclin D3 may be more dominant as $Ccnd2^{-/-}$ mice do not present with a reduced RBC or elevated MCV, at least in mice (M2 Supplementary Figure 5) (Sicinski et al. 1996). Thus, cyclin D3 may compensate for the loss of cyclin D2 but not *vice versa* as only $Ccnd3^{-/-}$ mice present with an obvious erythroid phenotype. Interestingly, common genetic variants within the cyclin A2 gene have also been associated with MCV, and depletion of cyclin A2 in erythroid cells leads to an increase red blood cell size (Ludwig et al. 2015). Although cyclin A2 appears to act at the transition through cytokinesis, these findings point to an unexpected common role of cell cycle regulators in influencing human erythroid traits.

A common feature of D-type cyclins is their regulation by environmental stimuli (Ciemerych et al. 2002). Cyclin D2 is a follicle stimulating hormone (FSH) responsive gene in ovarian granulosa cells (Sicinski et al. 1996), and in neutrophils the granulocyte colony-stimulating factor (G-CSF) signaling pathway impacts the cell cycle through cyclin D3 (Sicinska et al. 2006). In spermatogonia stem cell factor (SCF) and its receptor (KIT) induce cyclin D3 expression to promote cell cycle progression (Feng et al. 2000). The SCF/KIT pathway is also known to play an important role in the regulation of hematopoiesis and erythropoiesis (Felli et al. 2005; Hattangadi et al. 2011). Loss of function mutations in murine *Kit* lead to macrocytic anemia and oncogenic *Kit* mutants cause microcytic erythrocytosis (Bosbach et al. 2012). In agreement, tyrosine kinase inhibitors Sunitinib and Imatinib, also known to target human KIT, lead to macrocytosis in treated patients (Billemont et al. 2007;

Gillessen et al. 2007). Similarly, common genetic variants near *KIT* have been associated with red blood cell count (van der Harst et al. 2012). These studies implicate a possible role of this pathway in the regulation of cyclin D3 in erythroid differentiation and await further experimental validation.

4.6 Future Directions for functional Studies of GWAS

GWAS have revealed at least 75 loci to be associated with erythroid traits (van der Harst et al. 2012). Functional follow-up of these studies have mostly focused on the role of single genes (Sankaran et al. 2012b; Thom et al. 2014; Ludwig et al. 2015). One study conducted a loss of function screen of GWAS candidates in zebrafish to reveal new regulators of hematopoiesis, but provide no specific mechanistic insight in to the role of the identified genes and the underlying initial association (Bielczyk-Maczynska et al. 2014). Many human genes that are in these loci also do not have clear orthologues in zebrafish and limit the application of this model system. Identifying the causal variant and the affected gene remains a particularly challenging task as most variants are intergenic and the modulated gene remains unrecognized (Consortium 2012; Hnisz et al. 2013). Often the closest gene is analyzed, which has led to controversial findings as regulatory elements may act at megabase distances (Dina et al. 2007; Frayling et al. 2007; Smemo et al. 2014). In this study, analysis of transcription factor binding, reporter assays and chromatin conformation capture assisted in the identification of the causal variant regulating cyclin D3 expression. With respect to a more global analysis of GWAS derived variants, a number of systematic approaches may be applied to identify causal variants. Candidate genes may be identified by consolidating transcriptional data to find transcripts expressed nearby the associated loci, but only provides a rough estimate. Expression quantitative trait loci (eQTL) analysis allows identifying the linkage between genetic polymorphisms and variation in gene expression and has been successfully applied to identify genes of disease associations (Emilsson et al. 2008; Westra et al. 2013). Massively parallel reporter assays allow measuring of the transcriptional activity of gene regulatory variants and their respective alleles within associated loci to aid to identify transcriptionally active SNPs (Melnikov et al. 2012; Melnikov et al. 2014). Chromatin conformation capture assays, including high-resolution methods such as Hi-C have shown intergenic GWAS SNPs to be significantly enriched within promoter-interacting fragments (Rao et al. 2014; Mifsud et al. 2015) and may allow detection of interactions between associated loci and their target gene(s). These approaches and combinations thereof will likely

aid in the identification of causal variants and their candidate genes. Genetic perturbation of candidates may then be conducted on a larger scale using individual or pooled CRISPR or shRNA screens to assess their role within the respective system of interest (Chevrier et al. 2011; Possemato et al. 2011; Shalem et al. 2014; Wang et al. 2014), but profound mechanistic insight will require thorough examination on a more individual basis.

4.7 Outlook

The functional follow up of genetic studies has led to major insight into human physiology and disease. The identification of *GATA1* mutations in Diamond-Blackfan anemia elucidated a previously unrecognized defect due to ribosomal haploinsufficiency and opened up a new layer of complexity to the pathophysiology in this syndrome. GATA1 is a well-studied hematopoietic transcription factor, but a number of open questions with respect to its biology remain. The mechanistic nature and binding partners of the N-terminal transactivation domain of GATA1 remain to be identified. Generally, the delineation of the exact function of the individual protein isoforms of GATA1 in different hematopoietic cells and the role of GATA1 short in Down syndrome associated leukemia present challenging questions for future studies. The role of p53 with respect to the pathophysiology in DBA may have to be reassessed and its interplay with GATA1 presents an outstanding question warranting further investigation. The importance of *GATA1* in erythropoiesis and DBA provides a new rationale for novel treatment approaches to modulate GATA1 levels by small molecules or gene therapy approaches. With respect to all ribosomopathies a global and detailed understanding of how translation is modulated by genetic lesions suggested to influence ribosome biogenesis and function remains largely unknown. Furthermore, how these defects lead to tissue-specific phenotypes and potentially contribute to oncogenesis remain outstanding questions and provide an excellent incentive for further research. Subsequent mechanistically oriented studies will enable a more profound understanding of the basic cellular process of translation in physiology and disease and will likely aid in the identification of novel avenues to targeted therapies. Likewise, the follow-up of a common genetic variant revealed an unexpected role for ordinary cell cycle regulators to influence human quantitative erythroid traits. Numerous genome-wide association studies have been conducted and identified thousands of genetic variants to modulate disease susceptibility and human traits. The elucidation of the molecular mechanisms underlying these associations presents the major challenge for upcoming studies, but will inevitably reveal novel and thrilling aspects across the entire spectrum of biology and human disease.

5. Summary

Advances in genomic technology have led to the identification of numerous genetic variants that are associated with specific phenotypes. Whole-exome sequencing is often applied to identify rare variants in the context of Mendelian disorders, whereas genome-wide association studies (GWAS) have identified thousands of common variants associated with specific, quantifiable human traits. In both instances, identifying the specific causal variant and subsequently determining its mechanism of action, even on a small scale, remains difficult. In this thesis, we took advantage of the recent identification of the first non-ribosomal protein gene mutation found in a family with Diamond-Blackfan anemia (DBA). We followed up on the molecular etiology of these patients and discovered new insights with respect to the pathophysiology of DBA. Furthermore, we functionally investigated the role of common genetic variation on red blood cell traits, leading to new biological insights of cell cycle regulators in terminal erythroid differentiation.

In contrast to the more subtle effects in phenotype due to common genetic variation, rare genetic variants associated with Mendelian disorders are often deleterious and may require a significant amount of clinical attention. DBA is a rare congenital disease and is considered a disorder of the ribosome, as over 50 % of patients have heterozygous loss of function mutations in ribosomal protein genes, resulting in ribosomal haploinsufficiency. Although ribosomal proteins are ubiquitously expressed, the cell-type specific defect in the form of the anemia has puzzled researchers. Recently, in rare cases of DBA, mutations in the essential erythroid transcription factor GATA1 were described, but it remained unclear whether these patients shared a common pathophysiology as those with ribosomal haploinsufficiency. We show that ribosomal haploinsufficiency leads to impaired translation of *GATA1* mRNA and connect these two seemingly disparate sets of molecular lesions. We suggest that the amount of functional ribosomes is depleted in DBA, decreasing the overall translation initiation capacity. This is supported by specific inhibition of translation initiation by the small molecule 4EGI-1, phenocopying many *in vitro* aspects of DBA. We attribute the apparent selective translational defect to a highly structured 5' untranslated region of the *GATA1* mRNA, restricting its efficient translation when the cellular translation initiation potential is impaired. Most importantly, analysis of transcriptional signatures of primary erythroid patient cells with ribosomal protein mutations supports impairment in GATA1

transcriptional activity in DBA, and ectopic GATA1 expression in these cells greatly improved impaired erythropoiesis. These findings have led to new insights into the pathophysiology of DBA and provide not only an explanation for the specificity of the observed defect but suggest potential therapeutic avenues such as *GATA1* gene therapy approaches. More generally, this work implicates selectively impaired translation of specific transcripts as a mechanism of human disease.

GWASs have been enormously successful at identifying genetic associations but usually provide limited immediate insight to the underlying biology of these associations. By following up on genetic variants associated with red blood cell count and the average volume of a red blood cell, we identified an unexpected role in terminal erythropoiesis for a common cell cycle regulator. The causal variant mapped to a transcription factor binding site bound by erythroid transcription factors. Reporter gene assays showed differential enhancer activity for the associated alleles, suggesting altered transcriptional regulation of the nearby gene, cyclin D3. By following up on cyclin D3, a detailed analysis showed that downregulation of cyclin D3 levels lead to premature cell cycle arrest, without ostensible defects in differentiation. As a result, erythroid progenitor cells underwent fewer cell divisions and remained larger in size, preserving overall hemoglobin levels. Collectively, these data suggest that varying levels of cyclin D3 in terminal erythropoiesis contribute to the natural variation observed in red blood cell parameters in the human population.

Together, both studies illustrate how the functional follow up of rare and common human genetic variation leads to new insights and a deeper understanding of the underlying biology that effect basic cellular processes such as cell cycle regulation and translation in physiology and disease.

6. Zusammenfassung

Die Fortschritte in Genotypisierungs- und Sequenzierungstechnologien haben zur Entdeckung von zahlreichen humanen genetischen Varianten geführt, welche mit spezifischen Phänotypen bzw. Pathologien assoziiert sind. Die Exomsequenzierung wird häufig für die Identifizierung von seltenen genetischen Varianten in monogenetischen Erkrankungen genutzt. Im Gegensatz dazu haben genomweite Assoziationsstudien (GWAS) tausende von häufigen genetischen Varianten entdeckt, welche mit spezifischen humanen Phänotypen und Merkmalen assoziiert sind. Üblicherweise stellt in beiden Fällen die Identifizierung der kausalen Variante und die anschließende Aufklärung der damit verbundenen mechanistischen Grundlagen eine große Hürde dar. In der vorgelegten Dissertation haben wir auf Basis der jüngsten Entdeckung der ersten nicht-ribosomalen Mutation in einer Familie mit Diamond-Blackfan Anämie (DBA) den zu Grunde liegenden molekularen Mechanismus untersucht und neue Erkenntnisse über die damit verbundene Pathophysiologie dieser Erkrankung gewonnen. Darüber hinaus haben wir die Rolle von häufigen genetischen Varianten mit Assoziationen zu humanen Erythrozyten-Merkmalen untersucht und konnten so eine neue und ungeahnte Rolle für Zellzyklus-Regulatoren in der terminalen erythropoetischen Differenzierung aufdecken.

Häufige genetische Varianten haben in der Regel eine geringen Einfluss auf die Ausprägung eines (Krankheits-)Merkmals, wohingegen monogenetische Erkrankungen oft mit starker Beeinträchtigung sowie intensiver klinischer Betreuung einhergehen. DBA ist ein seltenes kongenitales Syndrom und ist assoziiert mit einer Funktionsstörung des Ribosoms. Über 50 % der Patienten haben heterozygote Funktionsverlustmutationen in Genen ribosomaler Proteine, die zu ribosomaler Haploinsuffizienz führen. Obwohl ribosomale Proteine ubiquitär exprimiert werden, stellte der zelltyp-spezifische Defekt in Form der Anämie Wissenschaftler lange Zeit vor ein Rätsel. Vor kurzem wurden in seltenen Fällen von DBA Mutationen im Gen des für die Erythropoese essentiellen Transkriptionsfaktors GATA1 beschrieben. Es blieb jedoch zunächst unklar, ob diesen Patienten, sowie Patienten mit ribosomaler Haploinsuffizienz eine gemeinsame Pathophysiologie zugrunde liegt. Im Rahmen dieser Arbeit konnten wir zeigen, dass ribosomale Haploinsuffizienz die Translation von *GATA1* mRNA beeinträchtigt und somit eine Verbindung zwischen den beiden scheinbar ungleichen molekularen Läsionen herstellt. Wir gehen davon aus, dass in DBA die Anzahl funktioneller Ribosomen reduziert und somit auch das Translationsinitiations-Potential der Zelle verringert ist. Diese Annahme wird durch die Beobachtung gestützt, dass die spezifische

Inhibition der Translationinitiation mit dem Molekül 4-EGI1 viele *in vitro* Charakteristika von DBA nachbildet. Wir konnten die scheinbar selektiv beeinträchtigte Translation auf eine ausgeprägt strukturierte 5' untranslatierte Region von *GATA1* mRNA zurückführen, welche ihre effiziente Translation unter Bedingungen reduzierter Translationsinitiations-Kapazität der Zelle einschränkt. Die Bedeutung von GATA1 in DBA wird durch eine signifikante Verminderung der transkriptionellen Aktivität von GATA1 in primären Zellen von DBA Patienten mit Mutationen in ribosomalen Proteinen untermauert. Entsprechend konnte die ektope Expression von GATA1 in diesen Zellen die beeinträchtigte Erythropoese stark verbessern. Diese Beobachtungen haben neue Einblicke in die Pathophysiologie von DBA geschaffen und erklären nicht nur die Spezifität des Defekts, sondern bereiten auch den Weg für neue therapeutische Ansätze wie die Gentherapie mittels *GATA1*. Ferner legen unsere Erkenntnisse die selektiv beeinträchtigte Translation spezifischer Transkripte als Grundlage für eine humane Pathologie dar.

GWAS haben eine Vielzahl neuer genetischer Assoziationen entdeckt, liefern unmittelbar jedoch nur geringe Erkenntnisse über die zugrunde liegende Biologie dieser Assoziationen. Durch Untersuchung von genetischen Varianten, welche mit Erythrozyten-Merkmalen wie der Anzahl und dem durchschnittlichen Volumen von roten Blutkörperchen assoziiert sind, haben wir Zellzyklus-Regulatoren eine bisher unbekannte Rolle zugeordnet. Die kausale Variante konnte in einer Bindungsstelle von erythropoetischen Transkriptionsfaktoren lokalisiert werden. Reportergen-Analysen zeigten unterschiedliche Enhancer-Aktivität für die assoziierten Allele und lassen variierende Mengen an Transkripten als Ursache für die Beeinflussung von Erythrozyten-Merkmalen vermuten. Wir identifizierten das durch die kausale Variante regulierte Gen Cyclin D3, und zeigen, dass geringere Cyclin D3-Level zu einem frühzeitigen Stopp des Zellzyklus führen, ohne die Differenzierung wesentlich zu beeinflussen. Das führt dazu, dass sich erythropoetischen Vorläuferzellen seltener teilen und sich so größere Erythrozyten bilden, welche somit den Hämoglobinspiegel aufrecht erhalten. Allgemein tragen so unterschiedliche Mengen an Cyclin D3 in der terminalen Erythropoese zu der natürlich vorkommenden Variation in Erythrozyten-Merkmalen in der menschlichen Bevölkerung bei.

Zusammen illustrieren beide Studien, wie die funktionelle Untersuchung von seltenen und häufigen genetischen Varianten tiefere Einblicke in die zugrunde liegenden molekularen Mechanismen liefern, welche fundamentale zelluläre Prozesse wie die Zellzyklus-Regulation und Translation in Physiologie und Krankheit beeinflussen.

7. Abbreviation List

3C	chromatin conformation capture
BFU-E	burst forming unit erythroid
bp	base pair
CAGE	cap analysis gene expression
CCND3	cyclin D3
CD	cluster of differentiation
CDK	cyclin dependent kinase
cDNA	complementary DNA
CFU-E	colony forming unit erythroid
ChIP	chromatin immunoprecipitation
CRISPR	clustered regularly interspaced short palindromic repeat
DBA	Diamond-Blackfan anemia
DNA	deoxyribonucleic acid
E14.5	embryonic day 14.5
eIF	eukaryotic initiation factor
ENCODE	encyclopedia of DNA elements
EPO	erythropoietin
EPOR	erythropoietin receptor
eQTL	expression quantitative trait loci
ExAC	exome aggregation consortium
FACS	fluorescence activated cell sorting
FL	fetal liver
FSC	forward scatter
GATA1 FL	full length isoform of GATA1
GATA1s	short isoform of GATA1
G-CSF	granulocyte colony-stimulating factor
GFP	green fluorescent protein
GSEA	gene set enrichment analysis
GWAS	genome-wide association study
Hb	hemoglobin
Hct	hematocrit
HSC	hematopoietic stem cell
Ig	immunoglobulin
IL	interleukin
IP	immunoprecipitation
IRES	internal ribosomal entry site
JAK	Janus kinase
KO	knockout
LD	linkage disequilibrium
Luc	luciferase

MCH	mean corpuscular hemoglobin
MCHC	mean corpuscular hemoglobin concentration
MCV	mean corpuscular volume
MDM2	mouse double minute 2
MEP	megakaryocyte and erythroid progenitor
mRNA	messenger RNA
mTOR	mammalian target of rapamycin
PCR	polymerase chain reaction
qRT-PCR	quantitative reverse transcriptase PCR
RACE	rapid amplification of cDNA ends
RBC	red blood cell count
RNA	ribonucleic acid
RP	ribosomal protein
rRNA	ribosomal RNA
rs	reference SNP
SCF	stem cell factor
SE	super enhancer
seq	sequencing
shRNA	short hairpin RNA
SNP	single nucleotide polymorphism
SSC	side scatter
STAT	signal transducer and activator of transcription
TAL1	T-cell acute lymphocytic leukemia-1
T-ALL	T-cell acute lymphoblastic leukemia
TD	transactivation domain
TE	traditional enhancer
TOP	terminal oligopyrimidine tract
uORF	upstream open reading frame
U	units
UTR	untranslated region
WES	whole-exome sequencing
WT	wild-type

8. Eidesstattliche Erklärung

Ich versichere hiermit, dass ich die vorliegende Dissertation selbstständig verfasst, die für diese Arbeit benutzten Hilfsmittel genannt und die Ergebnisse anderer klar gekennzeichnet habe.

Berlin, den 31. Mai 2015   Leif Ludwig

8. Eidesstattliche Erklärung

9. Bibliography

Aiuti A, Biasco L, Scaramuzza S, Ferrua F, Cicalese MP, Baricordi C, Dionisio F, Calabria A, Giannelli S, Castiello MC et al. 2013. Lentiviral hematopoietic stem cell gene therapy in patients with Wiskott-Aldrich syndrome. *Science* **341**: 1233151.

An X, Schulz VP, Li J, Wu K, Liu J, Xue F, Hu J, Mohandas N, Gallagher PG. 2014. Global transcriptome analyses of human and murine terminal erythroid differentiation. *Blood* **123**: 3466-3477.

Arlet JB, Ribeil JA, Guillem F, Negre O, Hazoume A, Marcion G, Beuzard Y, Dussiot M, Moura IC, Demarest S et al. 2014. HSP70 sequestration by free alpha-globin promotes ineffective erythropoiesis in beta-thalassaemia. *Nature* **514**: 242-246.

Babendure JR, Babendure JL, Ding JH, Tsien RY. 2006. Control of mammalian translation by mRNA structure near caps. *Rna* **12**: 851-861.

Badhai J, Frojmark AS, E JD, Schuster J, Dahl N. 2009. Ribosomal protein S19 and S24 insufficiency cause distinct cell cycle defects in Diamond-Blackfan anemia. *Biochimica et biophysica acta* **1792**: 1036-1042.

Balarajan Y, Ramakrishnan U, Ozaltin E, Shankar AH, Subramanian SV. 2011. Anaemia in low-income and middle-income countries. *Lancet* **378**: 2123-2135.

Ball SE, McGuckin CP, Jenkins G, Gordon-Smith EC. 1996. Diamond-Blackfan anaemia in the U.K.: analysis of 80 cases from a 20-year birth cohort. *British journal of haematology* **94**: 645-653.

Barlow JL, Drynan LF, Hewett DR, Holmes LR, Lorenzo-Abalde S, Lane AL, Jolin HE, Pannell R, Middleton AJ, Wong SH et al. 2010. A p53-dependent mechanism underlies macrocytic anemia in a mouse model of human 5q- syndrome. *Nature medicine* **16**: 59-66.

Bauer DE, Kamran SC, Lessard S, Xu J, Fujiwara Y, Lin C, Shao Z, Canver MC, Smith EC, Pinello L et al. 2013. An erythroid enhancer of BCL11A subject to genetic variation determines fetal hemoglobin level. *Science* **342**: 253-257.

Ben-Shem A, Garreau de Loubresse N, Melnikov S, Jenner L, Yusupova G, Yusupov M. 2011. The structure of the eukaryotic ribosome at 3.0 A resolution. *Science* **334**: 1524-1529.

Benyamin B, Ferreira MA, Willemsen G, Gordon S, Middelberg RP, McEvoy BP, Hottenga JJ, Henders AK, Campbell MJ, Wallace L et al. 2009. Common variants in TMPRSS6 are associated with iron status and erythrocyte volume. *Nature genetics* **41**: 1173-1175.

Bibikova E, Youn MY, Danilova N, Ono-Uruga Y, Konto-Ghiorghi Y, Ochoa R, Narla A, Glader B, Lin S, Sakamoto KM. 2014. TNF-mediated inflammation represses GATA1 and activates p38 MAP kinase in RPS19-deficient hematopoietic progenitors. *Blood* **124**: 3791-3798.

Bielczyk-Maczynska E, Serbanovic-Canic J, Ferreira L, Soranzo N, Stemple DL, Ouwehand WH, Cvejic A. 2014. A loss of function screen of identified genome-wide association study Loci reveals new genes controlling hematopoiesis. *PLoS Genet* **10**: e1004450.

Billemont B, Izzedine H, Rixe O. 2007. Macrocytosis due to treatment with sunitinib. *The New England journal of medicine* **357**: 1351-1352; author reply 1352.

Bjornson ZB, Nolan GP, Fantl WJ. 2013. Single-cell mass cytometry for analysis of immune system functional states. *Current opinion in immunology* **25**: 484-494.

Bolze A, Mahlaoui N, Byun M, Turner B, Trede N, Ellis SR, Abhyankar A, Itan Y, Patin E, Brebner S et al. 2013. Ribosomal protein SA haploinsufficiency in humans with isolated congenital asplenia. *Science* **340**: 976-978.

Bosbach B, Deshpande S, Rossi F, Shieh JH, Sommer G, de Stanchina E, Veach DR, Scandura JM, Manova-Todorova K, Moore MA et al. 2012. Imatinib resistance and microcytic erythrocytosis in a KitV558Delta;T669I/+ gatekeeper-mutant mouse model of gastrointestinal stromal tumor. *Proc Natl Acad Sci U S A* **109**: E2276-2283.

Boultwood J, Pellagatti A. 2014. Reduced translation of GATA1 in Diamond-Blackfan anemia. *Nature medicine* **20**: 703-704.

Boultwood J, Yip BH, Vuppusetty C, Pellagatti A, Wainscoat JS. 2013. Activation of the mTOR pathway by the amino acid (L)-leucine in the 5q- syndrome and other ribosomopathies. *Advances in biological regulation* **53**: 8-17.

Boylan JW, Van Liew JB, Feig PU. 1991. Inverse changes in erythroid cell volume and number regulate the hematocrit in newborn genetically hypertensive rats. *Proc Natl Acad Sci U S A* **88**: 9848-9852.

Bunn HF. 1997. Pathogenesis and treatment of sickle cell disease. *The New England journal of medicine* **337**: 762-769.

Bursac S, Brdovcak MC, Pfannkuchen M, Orsolic I, Golomb L, Zhu Y, Katz C, Daftuar L, Grabusic K, Vukelic I et al. 2012. Mutual protection of ribosomal proteins L5 and L11 from degradation is essential for p53 activation upon ribosomal biogenesis stress. *Proc Natl Acad Sci U S A* **109**: 20467-20472.

Bushby K, Finkel R, Wong B, Barohn R, Campbell C, Comi GP, Connolly AM, Day JW, Flanigan KM, Goemans N et al. 2014. Ataluren treatment of patients with nonsense mutation dystrophinopathy. *Muscle & nerve* **50**: 477-487.

Buszczak M, Signer RA, Morrison SJ. 2014. Cellular differences in protein synthesis regulate tissue homeostasis. *Cell* **159**: 242-251.

Byrska-Bishop M, VanDorn D, Campbell AE, Betensky M, Arca PR, Yao Y, Gadue P, Costa FF, Nemiroff RL, Blobel GA et al. 2015. Pluripotent stem cells reveal erythroid-specific activities of the GATA1 N-terminus. *The Journal of clinical investigation* **125**: 993-1005.

Calis JC, Phiri KS, Faragher EB, Brabin BJ, Bates I, Cuevas LE, de Haan RJ, Phiri AI, Malange P, Khoka M et al. 2008. Severe anemia in Malawian children. *The New England journal of medicine* **358**: 888-899.

Calligaris R, Bottardi S, Cogoi S, Apezteguia I, Santoro C. 1995. Alternative translation initiation site usage results in two functionally distinct forms of the GATA-1 transcription factor. *Proc Natl Acad Sci U S A* **92**: 11598-11602.

Calvo SE, Pagliarini DJ, Mootha VK. 2009. Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. *Proc Natl Acad Sci U S A* **106**: 7507-7512.

Campagna DR, de Bie CI, Schmitz-Abe K, Sweeney M, Sendamarai AK, Schmidt PJ, Heeney MM, Yntema HG, Kannengiesser C, Grandchamp B et al. 2014. X-linked sideroblastic anemia due to ALAS2 intron 1 enhancer element GATA-binding site mutations. *American journal of hematology* **89**: 315-319.

Campbell AE, Wilkinson-White L, Mackay JP, Matthews JM, Blobel GA. 2013. Analysis of disease-causing GATA1 mutations in murine gene complementation systems. *Blood* **121**: 5218-5227.

Cantor AB, Iwasaki H, Arinobu Y, Moran TB, Shigematsu H, Sullivan MR, Akashi K, Orkin SH. 2008. Antagonism of FOG-1 and GATA factors in fate choice for the mast cell lineage. *The Journal of experimental medicine* **205**: 611-624.

Casanova JL, Conley ME, Seligman SJ, Abel L, Notarangelo LD. 2014. Guidelines for genetic studies in single patients: lessons from primary immunodeficiencies. *The Journal of experimental medicine* **211**: 2137-2149.

Chambers JC, Zhang W, Li Y, Sehmi J, Wass MN, Zabaneh D, Hoggart C, Bayele H, McCarthy MI, Peltonen L et al. 2009. Genome-wide association study identifies variants in TMPRSS6 associated with hemoglobin levels. *Nature genetics* **41**: 1170-1172.

Cheng Y, Wu W, Kumar SA, Yu D, Deng W, Tripic T, King DC, Chen KB, Zhang Y, Drautz D et al. 2009. Erythroid GATA1 function revealed by genome-wide analysis of transcription factor occupancy, histone modifications, and mRNA expression. *Genome research* **19**: 2172-2184.

Chevrier N, Mertins P, Artyomov MN, Shalek AK, Iannacone M, Ciaccio MF, Gat-Viks I, Tonti E, DeGrace MM, Clauser KR et al. 2011. Systematic discovery of TLR signaling components delineates viral-sensing circuits. *Cell* **147**: 853-867.

Chlon TM, McNulty M, Goldenson B, Rosinski A, Crispino JD. 2015. Global transcriptome and chromatin occupancy analysis reveal the short isoform of GATA1 is deficient for erythroid specification and gene expression. *Haematologica* **100**: 575-584.

Choate KA, Lu Y, Zhou J, Choi M, Elias PM, Farhi A, Nelson-Williams C, Crumrine D, Williams ML, Nopper AJ et al. 2010. Mitotic recombination in patients with ichthyosis causes reversion of dominant mutations in KRT10. *Science* **330**: 94-97.

Choate KA, Lu Y, Zhou J, Elias PM, Zaidi S, Paller AS, Farhi A, Nelson-Williams C, Crumrine D, Milstone LM et al. 2015. Frequent somatic reversion of KRT1 mutations in ichthyosis with confetti. *The Journal of clinical investigation* **125**: 1703-1707.

Choi YJ, Li X, Hydbring P, Sanda T, Stefano J, Christie AL, Signoretti S, Look AT, Kung AL, von Boehmer H et al. 2012. The requirement for cyclin D function in tumor maintenance. *Cancer cell* **22**: 438-451.

Chung J, Bauer DE, Ghamari A, Nizzi CP, Deck KM, Kingsley PD, Yien YY, Huston NC, Chen C, Schultz IJ et al. 2015. The mTORC1/4E-BP pathway coordinates hemoglobin production with L-leucine availability. *Science signaling* **8**: ra34.

Church C, Moir L, McMurray F, Girard C, Banks GT, Teboul L, Wells S, Bruning JC, Nolan PM, Ashcroft FM et al. 2010. Overexpression of Fto leads to increased food intake and results in obesity. *Nature genetics* **42**: 1086-1092.

Ciemerych MA, Kenney AM, Sicinska E, Kalaszczynska I, Bronson RT, Rowitch DH, Gardner H, Sicinski P. 2002. Development of mice expressing a single D-type cyclin. *Genes & development* **16**: 3277-3289.

Cmejlova J, Dolezalova L, Pospisilova D, Petrtylova K, Petrak J, Cmejla R. 2006. Translational efficiency in patients with Diamond-Blackfan anemia. *Haematologica* **91**: 1456-1464.

Consortium EP. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**: 57-74.

Cooper AB, Sawai CM, Sicinska E, Powers SE, Sicinski P, Clark MR, Aifantis I. 2006. A unique function for cyclin D3 in early B cell development. *Nature immunology* **7**: 489-497.

Cooper DN, Chen JM, Ball EV, Howells K, Mort M, Phillips AD, Chuzhanova N, Krawczak M, Kehrer-Sawatzki H, Stenson PD. 2010. Genes, mutations, and human inherited disease at the dawn of the age of personalized genomics. *Human mutation* **31**: 631-655.

Couzin-Frankel J. 2010. Major heart disease genes prove elusive. *Science* **328**: 1220-1221.

Crispino JD, Weiss MJ. 2014. Erythro-megakaryocytic transcription factors associated with hereditary anemia. *Blood* **123**: 3080-3088.

D'Andrea AD, Lodish HF, Wong GG. 1989. Expression cloning of the murine erythropoietin receptor. *Cell* **57**: 277-285.

Danilova N, Sakamoto KM, Lin S. 2008. Ribosomal protein S19 deficiency in zebrafish leads to developmental abnormalities and defective erythropoiesis through activation of p53 protein family. *Blood* **112**: 5228-5237.

De Keersmaecker K, Atak ZK, Li N, Vicente C, Patchett S, Girardi T, Gianfelici V, Geerdens E, Clappier E, Porcu M et al. 2013. Exome sequencing identifies mutation in CNOT3 and ribosomal genes RPL5 and RPL10 in T-cell acute lymphoblastic leukemia. *Nature genetics* **45**: 186-190.

De Keersmaecker K, Sulima SO, Dinman JD. 2015. Ribosomopathies and the paradox of cellular hypo- to hyperproliferation. *Blood* **125**: 1377-1382.

De Maria R, Zeuner A, Eramo A, Domenichelli C, Bonci D, Grignani F, Srinivasula SM, Alnemri ES, Testa U, Peschle C. 1999. Negative regulation of erythropoiesis by caspase-mediated cleavage of GATA-1. *Nature* **401**: 489-493.

Deshpande A, Sicinski P, Hinds PW. 2005. Cyclins and cdks in development and cancer: a perspective. *Oncogene* **24**: 2909-2915.

Diamond L, Blackfan K. 1938. Hypoplastic anemia. *Am J Dis Child*: 464-467.

Dina C, Meyre D, Gallina S, Durand E, Korner A, Jacobson P, Carlsson LM, Kiess W, Vatin V, Lecoeur C et al. 2007. Variation in FTO contributes to childhood obesity and severe adult obesity. *Nature genetics* **39**: 724-726.

Doherty L, Sheen MR, Vlachos A, Choesmel V, O'Donohue MF, Clinton C, Schneider HE, Sieff CA, Newburger PE, Ball SE et al. 2010. Ribosomal protein genes RPS10 and RPS26 are commonly mutated in Diamond-Blackfan anemia. *American journal of human genetics* **86**: 222-228.

Doulatov S, Notta F, Laurenti E, Dick JE. 2012. Hematopoiesis: a human perspective. *Cell stem cell* **10**: 120-136.

Draptchinskaia N, Gustavsson P, Andersson B, Pettersson M, Willig TN, Dianzani I, Ball S, Tchernia G, Klar J, Matsson H et al. 1999. The gene encoding ribosomal protein S19 is mutated in Diamond-Blackfan anaemia. *Nature genetics* **21**: 169-175.

Du X, She E, Gelbart T, Truksa J, Lee P, Xia Y, Khovananth K, Mudd S, Mann N, Moresco EM et al. 2008. The serine protease TMPRSS6 is required to sense iron deficiency. *Science* **320**: 1088-1092.

Dussiot M, Maciel TT, Fricot A, Chartier C, Negre O, Veiga J, Grapton D, Paubelle E, Payen E, Beuzard Y et al. 2014. An activin receptor IIA ligand trap corrects ineffective erythropoiesis in beta-thalassemia. *Nature medicine* **20**: 398-407.

Dutt S, Narla A, Lin K, Mullally A, Abayasekara N, Megerdichian C, Wilson FH, Currie T, Khanna-Gupta A, Berliner N et al. 2011. Haploinsufficiency for ribosomal protein genes causes selective activation of p53 in human erythroid progenitor cells. *Blood* **117**: 2567-2576.

Dvir S, Velten L, Sharon E, Zeevi D, Carey LB, Weinberger A, Segal E. 2013. Deciphering the rules by which 5'-UTR sequences affect protein expression in yeast. *Proc Natl Acad Sci U S A* **110**: E2792-2801.

Ebert BL, Lee MM, Pretz JL, Subramanian A, Mak R, Golub TR, Sieff CA. 2005. An RNA interference model of RPS19 deficiency in Diamond-Blackfan anemia recapitulates defective hematopoiesis and rescue by dexamethasone: identification of dexamethasone-responsive genes by microarray. *Blood* **105**: 4620-4626.

Ebert BL, Pretz J, Bosco J, Chang CY, Tamayo P, Galili N, Raza A, Root DE, Attar E, Ellis SR et al. 2008. Identification of RPS14 as a 5q- syndrome gene by RNA interference screen. *Nature* **451**: 335-339.

Edwards SL, Beesley J, French JD, Dunning AM. 2013. Beyond GWASs: illuminating the dark road from association to function. *American journal of human genetics* **93**: 779-797.

Elagib KE, Racke FK, Mogass M, Khetawat R, Delehanty LL, Goldfarb AN. 2003. RUNX1 and GATA-1 coexpression and cooperation in megakaryocytic differentiation. *Blood* **101**: 4333-4341.

Emilsson V, Thorleifsson G, Zhang B, Leonardson AS, Zink F, Zhu J, Carlson S, Helgason A, Walters GB, Gunnarsdottir S et al. 2008. Genetics of gene expression and its effect on disease. *Nature* **452**: 423-428.

Fagioli F, Quarello P, Zecca M, Lanino E, Corti P, Favre C, Ripaldi M, Ramenghi U, Locatelli F, Prete A. 2014. Haematopoietic stem cell transplantation for Diamond Blackfan anaemia: a report from the Italian Association of Paediatric Haematology and Oncology Registry. *British journal of haematology* **165**: 673-681.

Fang S, Jensen JP, Ludwig RL, Vousden KH, Weissman AM. 2000. Mdm2 is a RING finger-dependent ubiquitin protein ligase for itself and p53. *The Journal of biological chemistry* **275**: 8945-8951.

Farh KK, Marson A, Zhu J, Kleinewietfeld M, Housley WJ, Beik S, Shoresh N, Whitton H, Ryan RJ, Shishkin AA et al. 2015. Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* **518**: 337-343.

Farrar JE, Quarello P, Fisher R, O'Brien KA, Aspesi A, Parrella S, Henson AL, Seidel NE, Atsidaftos E, Prakash S et al. 2014. Exploiting pre-rRNA processing in Diamond Blackfan anemia gene discovery and diagnosis. *American journal of hematology* **89**: 985-991.

Felli N, Fontana L, Pelosi E, Botta R, Bonci D, Facchiano F, Liuzzi F, Lulli V, Morsilli O, Santoro S et al. 2005. MicroRNAs 221 and 222 inhibit normal erythropoiesis and erythroleukemic cell growth via kit receptor down-modulation. *Proc Natl Acad Sci U S A* **102**: 18081-18086.

Feng LX, Ravindranath N, Dym M. 2000. Stem cell factor/c-kit up-regulates cyclin D3 and promotes cell cycle progression via the phosphoinositide 3-kinase/p70 S6 kinase pathway in spermatogonia. *The Journal of biological chemistry* **275**: 25572-25576.

Ferreira R, Ohneda K, Yamamoto M, Philipsen S. 2005. GATA1 function, a paradigm for transcription factors in hematopoiesis. *Molecular and cellular biology* **25**: 1215-1227.

Fischer J, Koch L, Emmerling C, Vierkotten J, Peters T, Bruning JC, Ruther U. 2009. Inactivation of the Fto gene protects from obesity. *Nature* **458**: 894-898.

Flygare J, Aspesi A, Bailey JC, Miyake K, Caffrey JM, Karlsson S, Ellis SR. 2007. Human RPS19, the gene mutated in Diamond-Blackfan anemia, encodes a ribosomal protein required for the maturation of 40S ribosomal subunits. *Blood* **109**: 980-986.

Flygare J, Kiefer T, Miyake K, Utsugisawa T, Hamaguchi I, Da Costa L, Richter J, Davey EJ, Matsson H, Dahl N et al. 2005. Deficiency of ribosomal protein S19 in CD34+ cells generated by siRNA blocks erythroid development and mimics defects seen in Diamond-Blackfan anemia. *Blood* **105**: 4627-4634.

Flygare J, Olsson K, Richter J, Karlsson S. 2008. Gene therapy of Diamond Blackfan anemia CD34(+) cells leads to improved erythroid development and engraftment following transplantation. *Experimental hematology* **36**: 1428-1435.

Flygare J, Rayon Estrada V, Shin C, Gupta S, Lodish HF. 2011. HIF1alpha synergizes with glucocorticoids to promote BFU-E progenitor self-renewal. *Blood* **117**: 3435-3444.

Frayling TM, Timpson NJ, Weedon MN, Zeggini E, Freathy RM, Lindgren CM, Perry JR, Elliott KS, Lango H, Rayner NW et al. 2007. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* **316**: 889-894.

Fugger L, McVean G, Bell JI. 2012. Genomewide association studies and common disease--realizing clinical utility. *The New England journal of medicine* **367**: 2370-2371.

Fujiwara Y, Browne CP, Cunniff K, Goff SC, Orkin SH. 1996. Arrested development of embryonic red cell precursors in mouse embryos lacking transcription factor GATA-1. *Proc Natl Acad Sci U S A* **93**: 12355-12358.

Ganesh SK, Zakai NA, van Rooij FJ, Soranzo N, Smith AV, Nalls MA, Chen MH, Kottgen A, Glazer NL, Dehghan A et al. 2009. Multiple loci influence erythrocyte phenotypes in the CHARGE Consortium. *Nature genetics* **41**: 1191-1198.

Garcon L, Ge J, Manjunath SH, Mills JA, Apicella M, Parikh S, Sullivan LM, Podsakoff GM, Gadue P, French DL et al. 2013. Ribosomal and hematopoietic defects in induced pluripotent stem cells derived from Diamond Blackfan anemia patients. *Blood* **122**: 912-921.

Gazda HT, Sheen MR, Vlachos A, Choesmel V, O'Donohue MF, Schneider H, Darras N, Hasman C, Sieff CA, Newburger PE et al. 2008. Ribosomal protein L5 and L11 mutations are associated with cleft palate and abnormal thumbs in Diamond-Blackfan anemia patients. *American journal of human genetics* **83**: 769-780.

Gazda HT, Zhong R, Long L, Niewiadomska E, Lipton JM, Ploszynska A, Zaucha JM, Vlachos A, Atsidaftos E, Viskochil DH et al. 2004. RNA and protein evidence for haplo-insufficiency in Diamond-Blackfan anaemia patients with RPS19 mutations. *British journal of haematology* **127**: 105-113.

Genovese G, Kahler AK, Handsaker RE, Lindberg J, Rose SA, Bakhoum SF, Chambert K, Mick E, Neale BM, Fromer M et al. 2014a. Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *The New England journal of medicine* **371**: 2477-2487.

Genovese P, Schiroli G, Escobar G, Di Tomaso T, Firrito C, Calabria A, Moi D, Mazzieri R, Bonini C, Holmes MC et al. 2014b. Targeted genome editing in human repopulating haematopoietic stem cells. *Nature* **510**: 235-240.

Gentner B, Visigalli I, Hiramatsu H, Lechman E, Ungari S, Giustacchini A, Schira G, Amendola M, Quattrini A, Martino S et al. 2010. Identification of hematopoietic stem cell-specific miRNAs enables gene therapy of globoid cell leukodystrophy. *Science translational medicine* **2**: 58ra84.

Gieger C Radhakrishnan A Cvejic A Tang W Porcu E Pistis G Serbanovic-Canic J Elling U Goodall AH Labrune Y et al. 2011. New gene functions in megakaryopoiesis and platelet formation. *Nature* **480**: 201-208.

147

Gillessen S, Graf L, Korte W, Cerny T. 2007. Macrocytosis and cobalamin deficiency in patients treated with sunitinib. *The New England journal of medicine* **356**: 2330-2331.

Gismondi A, Caldarola S, Lisi G, Juli G, Chellini L, Iadevaia V, Proud CG, Loreni F. 2014. Ribosomal stress activates eEF2K-eEF2 pathway causing translation elongation inhibition and recruitment of terminal oligopyrimidine (TOP) mRNAs on polysomes. *Nucleic acids research* **42**: 12668-12680.

Gkogkas CG, Khoutorsky A, Ran I, Rampakakis E, Nevarko T, Weatherill DB, Vasuta C, Yee S, Truitt M, Dallaire P et al. 2013. Autism-related deficits via dysregulated eIF4E-dependent translational control. *Nature* **493**: 371-377.

Gnirke A, Melnikov A, Maguire J, Rogov P, LeProust EM, Brockman W, Fennell T, Giannoukos G, Fisher S, Russ C et al. 2009. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nature biotechnology* **27**: 182-189.

Goldfarb AN. 2009. Megakaryocytic programming by a transcriptional regulatory loop: A circle connecting RUNX1, GATA-1, and P-TEFb. *Journal of cellular biochemistry* **107**: 377-382.

Gotz R, Wiese S, Takayama S, Camarero GC, Rossoll W, Schweizer U, Troppmair J, Jablonka S, Holtmann B, Reed JC et al. 2005. Bag1 is essential for differentiation and survival of hematopoietic and neuronal cells. *Nature neuroscience* **8**: 1169-1178.

Greene ME, Mundschau G, Wechsler J, McDevitt M, Gamis A, Karp J, Gurbuxani S, Arceci R, Crispino JD. 2003. Mutations in GATA1 in both transient myeloproliferative disorder and acute megakaryoblastic leukemia of Down syndrome. *Blood cells, molecules & diseases* **31**: 351-356.

Gripp KW, Curry C, Olney AH, Sandoval C, Fisher J, Chong JX, Genomics UWCfM, Pilchman L, Sahraoui R, Stabley DL et al. 2014. Diamond-Blackfan anemia with mandibulofacial dystostosis is heterogeneous, including the novel DBA genes TSR2 and RPS28. *American journal of medical genetics Part A* **164A**: 2240-2249.

Gripp KW, McDonald-McGinn DM, La Rossa D, McGain D, Federman N, Vlachos A, Glader BE, McKenzie SE, Lipton JM, Zackai EH. 2001. Bilateral microtia and cleft palate in cousins with Diamond-Blackfan anemia. *American journal of medical genetics* **101**: 268-274.

Guo H, Ingolia NT, Weissman JS, Bartel DP. 2010. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* **466**: 835-840.

Gustavsson P, Skeppner G, Johansson B, Berg T, Gordon L, Kreuger A, Dahl N. 1997. Diamond-Blackfan anaemia in a girl with a de novo balanced reciprocal X;19 translocation. *Journal of medical genetics* **34**: 779-782.

Haas M, Vlcek V, Balabanov P, Salmonson T, Bakchine S, Markey G, Weise M, Schlosser-Weber G, Brohmann H, Yerro CP et al. 2015. European Medicines Agency review of ataluren for the treatment of ambulant patients aged 5 years and older with Duchenne muscular dystrophy resulting from a nonsense mutation in the dystrophin gene. *Neuromuscular disorders : NMD* **25**: 5-13.

Hattangadi SM, Wong P, Zhang L, Flygare J, Lodish HF. 2011. From stem cell to red cell: regulation of erythropoiesis at multiple levels by multiple proteins, RNAs, and chromatin modifications. *Blood* **118**: 6258-6268.

Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA. 2009. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A* **106**: 9362-9367.

Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-Andre V, Sigova AA, Hoke HA, Young RA. 2013. Super-enhancers in the control of cell identity and disease. *Cell* **155**: 934-947.

Hollanda LM, Lima CS, Cunha AF, Albuquerque DM, Vassallo J, Ozelo MC, Joazeiro PP, Saad ST, Costa FF. 2006. An inherited mutation leading to production of only the short isoform of GATA-1 is associated with impaired erythropoiesis. *Nature genetics* **38**: 807-812.

Horos R, Ijspeert H, Pospisilova D, Sendtner R, Andrieu-Soler C, Taskesen E, Nieradka A, Cmejla R, Sendtner M, Touw IP et al. 2012. Ribosomal deficiencies in Diamond-Blackfan anemia impair translation of transcripts essential for differentiation of murine and human erythroblasts. *Blood* **119**: 262-272.

Horos R, von Lindern M. 2012. Molecular mechanisms of pathology and treatment in Diamond Blackfan Anaemia. *British journal of haematology* **159**: 514-527.

Hsieh AC, Liu Y, Edlind MP, Ingolia NT, Janes MR, Sher A, Shi EY, Stumpf CR, Christensen C, Bonham MJ et al. 2012. The translational landscape of mTOR signalling steers cancer initiation and metastasis. *Nature* **485**: 55-61.

Hu J, Liu J, Xue F, Halverson G, Reid M, Guo A, Chen L, Raza A, Galili N, Jaffray J et al. 2013. Isolation and functional characterization of human erythroblasts at distinct stages: implications for understanding of normal and disordered erythropoiesis in vivo. *Blood* **121**: 3246-3253.

Ingolia NT, Lareau LF, Weissman JS. 2011. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* **147**: 789-802.

Iskander D, Psaila B, Gerrard G, Chaidos A, En Foong H, Harrington Y, Karnik LC, Roberts I, de la Fuente J, Karadimitris A. 2015. Elucidation of the EP defect in Diamond-Blackfan anemia by characterization and prospective isolation of human EPs. *Blood* **125**: 2553-2557.

Iwasaki H, Mizuno S, Wells RA, Cantor AB, Watanabe S, Akashi K. 2003. GATA-1 converts lymphoid and myelomonocytic progenitors into the megakaryocyte/erythrocyte lineages. *Immunity* **19**: 451-462.

Jaako P, Debnath S, Olsson K, Bryder D, Flygare J, Karlsson S. 2012. Dietary L-leucine improves the anemia in a mouse model for Diamond-Blackfan anemia. *Blood* **120**: 2225-2228.

Jaako P, Debnath S, Olsson K, Modlich U, Rothe M, Schambach A, Flygare J, Karlsson S. 2014. Gene therapy cures the anemia and lethal bone marrow failure in a mouse model of RPS19-deficient Diamond-Blackfan anemia. *Haematologica* **99**: 1792-1798.

Jaako P, Debnath S, Olsson K, Zhang Y, Flygare J, Lindstrom MS, Bryder D, Karlsson S. 2015. Disruption of the 5S RNP-Mdm2 interaction significantly improves the erythroid defect in a mouse model for Diamond-Blackfan anemia. *Leukemia*.

Jaako P, Flygare J, Olsson K, Quere R, Ehinger M, Henson A, Ellis S, Schambach A, Baum C, Richter J et al. 2011. Mice with ribosomal protein S19 deficiency develop bone marrow failure and symptoms like patients with Diamond-Blackfan anemia. *Blood* **118**: 6087-6096.

Jaiswal S, Fontanillas P, Flannick J, Manning A, Grauman PV, Mar BG, Lindsley RC, Mermel CH, Burtt N, Chavez A et al. 2014. Age-related clonal hematopoiesis associated with adverse outcomes. *The New England journal of medicine* **371**: 2488-2498.

Jayapal SR, Wang CQ, Bisteau X, Caldez MJ, Lim S, Tergaonkar V, Osato M, Kaldis P. 2015. Hematopoiesis specific loss of Cdk2 and Cdk4 results in increased erythrocyte size and delayed platelet recovery following stress. *Haematologica* **100**: 431-438.

Jeffreys AJ, Kauppi L, Neumann R. 2001. Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. *Nature genetics* **29**: 217-222.

Jelkmann W. 2013. Physiology and pharmacology of erythropoietin. *Transfusion medicine and hemotherapy : offizielles Organ der Deutschen Gesellschaft fur Transfusionsmedizin und Immunhamatologie* **40**: 302-309.

Jia Q, Zhang Q, Zhang Z, Wang Y, Zhang W, Zhou Y, Wan Y, Cheng T, Zhu X, Fang X et al. 2013. Transcriptome analysis of the zebrafish model of Diamond-Blackfan anemia from RPS19 deficiency via p53-dependent and -independent pathways. *PLoS one* **8**: e71782.

Josephs H. 1936. Anemia of infancy and early childhood. *Medicine* **15**: 307-451.

Kadauke S, Blobel GA. 2013. Mitotic bookmarking by transcription factors. *Epigenetics & chromatin* **6**: 6.

Kadauke S, Udugama MI, Pawlicki JM, Achtman JC, Jain DP, Cheng Y, Hardison RC, Blobel GA. 2012. Tissue-specific mitotic bookmarking by hematopoietic transcription factor GATA1. *Cell* **150**: 725-737.

Kadri Z, Shimizu R, Ohneda O, Maouche-Chretien L, Gisselbrecht S, Yamamoto M, Romeo PH, Leboulch P, Chretien S. 2009. Direct binding of pRb/E2F-2 to GATA-1 regulates maturation and terminal cell division during erythropoiesis. *PLoS biology* **7**: e1000123.

Kamatani Y, Matsuda K, Okada Y, Kubo M, Hosono N, Daigo Y, Nakamura Y, Kamatani N. 2010. Genome-wide association study of hematological and biochemical traits in a Japanese population. *Nature genetics* **42**: 210-215.

Kaneko K, Furuyama K, Fujiwara T, Kobayashi R, Ishida H, Harigae H, Shibahara S. 2014. Identification of a novel erythroid-specific enhancer for the ALAS2 gene and its loss-of-function mutation which is associated with congenital sideroblastic anemia. *Haematologica* **99**: 252-261.

Kassebaum NJ, Jasrasaria R, Naghavi M, Wulf SK, Johns N, Lozano R, Regan M, Weatherall D, Chou DP, Eisele TP et al. 2014. A systematic analysis of global anemia burden from 1990 to 2010. *Blood* **123**: 615-624.

Kathiresan S, Srivastava D. 2012. Genetics of human cardiovascular disease. *Cell* **148**: 1242-1257.

Kerem E, Konstan MW, De Boeck K, Accurso FJ, Sermet-Gaudelus I, Wilschanski M, Elborn JS, Melotti P, Bronsveld I, Fajac I et al. 2014. Ataluren for the treatment of nonsense-mutation cystic fibrosis: a randomised, double-blind, placebo-controlled phase 3 trial. *The Lancet Respiratory medicine* **2**: 539-547.

Kerenyi MA, Orkin SH. 2010. Networking erythropoiesis. *The Journal of experimental medicine* **207**: 2537-2541.

Klar J, Khalfallah A, Arzoo PS, Gazda HT, Dahl N. 2014. Recurrent GATA1 mutations in Diamond-Blackfan anaemia. *British journal of haematology* **166**: 949-951.

Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, Henning AK, SanGiovanni JP, Mane SM, Mayne ST et al. 2005. Complement factor H polymorphism in age-related macular degeneration. *Science* **308**: 385-389.

Knight ZA, Schmidt SF, Birsoy K, Tan K, Friedman JM. 2014. A critical role for mTORC1 in erythropoiesis and anemia. *eLife* **3**: e01913.

Kobayashi M, Nishikawa K, Yamamoto M. 2001. Hematopoietic regulatory domain of gata1 gene is positively regulated by GATA1 protein in zebrafish embryos. *Development* **128**: 2341-2350.

Kondrashov N, Pusic A, Stumpf CR, Shimizu K, Hsieh AC, Xue S, Ishijima J, Shiroishi T, Barna M. 2011. Ribosome-mediated specificity in Hox mRNA translation and vertebrate tissue patterning. *Cell* **145**: 383-397.

Kong J, Lasko P. 2012. Translational control in cellular and developmental processes. *Nature reviews Genetics* **13**: 383-394.

Kozak M. 1991. An analysis of vertebrate mRNA sequences: intimations of translational control. *The Journal of cell biology* **115**: 887-903.

Kozar K, Ciemerych MA, Rebel VI, Shigematsu H, Zagozdzon A, Sicinska E, Geng Y, Yu Q, Bhattacharya S, Bronson RT et al. 2004. Mouse development and cell proliferation in the absence of D-cyclins. *Cell* **118**: 477-491.

Lander ES. 2011. Initial impact of the sequencing of the human genome. *Nature* **470**: 187-197.

Lander ES Linton LM Birren B Nusbaum C Zody MC Baldwin J Devon K Dewar K Doyle M FitzHugh W et al. 2001. Initial sequencing and analysis of the human genome. *Nature* **409**: 860-921.

Laplante M, Sabatini DM. 2012. mTOR signaling in growth control and disease. *Cell* **149**: 274-293.

Lee HY, Gao X, Barrasa MI, Li H, Elmes RR, Peters LL, Lodish HF. 2015. PPAR-alpha and glucocorticoid receptor synergize to promote erythroid progenitor self-renewal. *Nature*.

Li J, Hale J, Bhagia P, Xue F, Chen L, Jaffray J, Yan H, Lane J, Gallagher PG, Mohandas N et al. 2014a. Isolation and transcriptome analyses of human erythroid progenitors: BFU-E and CFU-E. *Blood* **124**: 3636-3645.

Li L, Lee JY, Gross J, Song SH, Dean A, Love PE. 2010. A requirement for Lim domain binding protein 1 in erythropoiesis. *The Journal of experimental medicine* **207**: 2543-2550.

Li X, Lu YC, Dai K, Torregroza I, Hla T, Evans T. 2014b. Elavl1a regulates zebrafish erythropoiesis via posttranscriptional control of gata1. *Blood* **123**: 1384-1392.

Lipton JM, Atsidaftos E, Zyskind I, Vlachos A. 2006. Improving clinical care and elucidating the pathophysiology of Diamond Blackfan anemia: an update from the Diamond Blackfan Anemia Registry. *Pediatric blood & cancer* **46**: 558-564.

Lipton JM, Ellis SR. 2009. Diamond-Blackfan anemia: diagnosis, treatment, and molecular pathogenesis. *Hematology/oncology clinics of North America* **23**: 261-282.

Ludwig LS, Cho H, Wakabayashi A, Eng JC, Ulirsch JC, Fleming MD, Lodish HF, Sankaran VG. 2015. Genome-wide association study follow-up identifies cyclin A2 as a regulator of the transition through cytokinesis during terminal erythropoiesis. *American journal of hematology* **90**: 386-391.

Ludwig LS, Gazda HT, Eng JC, Eichhorn SW, Thiru P, Ghazvinian R, George TI, Gotlib JR, Beggs AH, Sieff CA et al. 2014. Altered translation of GATA1 in Diamond-Blackfan anemia. *Nature medicine* **20**: 748-753.

Lujan E, Zunder ER, Ng YH, Goronzy IN, Nolan GP, Wernig M. 2015. Early reprogramming regulators identified by prospective isolation and mass cytometry. *Nature* **521**: 352-356.

Lupski JR, Belmont JW, Boerwinkle E, Gibbs RA. 2011. Clan genomics and the complex architecture of human disease. *Cell* **147**: 32-43.

Macias E, Jin A, Deisenroth C, Bhat K, Mao H, Lindstrom MS, Zhang Y. 2010. An ARF-independent c-MYC-activated tumor suppression pathway mediated by ribosomal protein-Mdm2 Interaction. *Cancer cell* **18**: 231-243.

Maeda T, Ito K, Merghoub T, Poliseno L, Hobbs RM, Wang G, Dong L, Maeda M, Dore LC, Zelent A et al. 2009. LRF is an essential downstream target of GATA1 in erythroid development and regulates BIM-dependent apoptosis. *Developmental cell* **17**: 527-540.

Majewski IJ, Metcalf D, Mielke LA, Krebs DL, Ellis S, Carpinelli MR, Mifsud S, Di Rago L, Corbin J, Nicola NA et al. 2006. A mutation in the translation initiation codon of Gata-1 disrupts megakaryocyte maturation and causes thrombocytopenia. *Proc Natl Acad Sci U S A* **103**: 14146-14151.

Malumbres M, Barbacid M. 2009. Cell cycle, CDKs and cancer: a changing paradigm. *Nature reviews Cancer* **9**: 153-166.

Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A et al. 2009. Finding the missing heritability of complex diseases. *Nature* **461**: 747-753.

Marechal V, Elenbaas B, Piette J, Nicolas JC, Levine AJ. 1994. The ribosomal L5 protein is associated with mdm-2 and mdm-2-p53 complexes. *Molecular and cellular biology* **14**: 7414-7420.

Matsson H, Davey EJ, Draptchinskaia N, Hamaguchi I, Ooka A, Leveen P, Forsberg E, Karlsson S, Dahl N. 2004. Targeted disruption of the ribosomal protein S19 gene is lethal prior to implantation. *Molecular and cellular biology* **24**: 4032-4037.

Matsson H, Davey EJ, Frojmark AS, Miyake K, Utsugisawa T, Flygare J, Zahou E, Byman I, Landin B, Ronquist G et al. 2006. Erythropoiesis in the Rps19 disrupted mouse: Analysis of erythropoietin

response and biochemical markers for Diamond-Blackfan anemia. *Blood cells, molecules & diseases* **36**: 259-264.

McCann KL, Baserga SJ. 2013. Genetics. Mysterious ribosomopathies. *Science* **341**: 849-850.

McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JP, Hirschhorn JN. 2008. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nature reviews Genetics* **9**: 356-369.

McDevitt MA, Shivdasani RA, Fujiwara Y, Yang H, Orkin SH. 1997. A "knockdown" mutation created by cis-element gene targeting reveals the dependence of erythroid cell maturation on the level of transcription factor GATA-1. *Proc Natl Acad Sci U S A* **94**: 6781-6785.

McElroy SP, Nomura T, Torrie LS, Warbrick E, Gartner U, Wood G, McLean WH. 2013. A lack of premature termination codon read-through efficacy of PTC124 (Ataluren) in a diverse array of reporter assays. *PLoS biology* **11**: e1001593.

McGrath K, Palis J. 2008. Ontogeny of erythropoiesis in the mammalian embryo. *Current topics in developmental biology* **82**: 1-22.

Melnikov A, Murugan A, Zhang X, Tesileanu T, Wang L, Rogov P, Feizi S, Gnirke A, Callan CG, Jr., Kinney JB et al. 2012. Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. *Nature biotechnology* **30**: 271-277.

Melnikov A, Zhang X, Rogov P, Wang L, Mikkelsen TS. 2014. Massively parallel reporter assays in cultured mammalian cells. *Journal of visualized experiments : JoVE*.

Mifsud B, Tavares-Cadete F, Young AN, Sugar R, Schoenfelder S, Ferreira L, Wingett SW, Andrews S, Grey W, Ewels PA et al. 2015. Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nature genetics*.

Modell B, Darlison M. 2008. Global epidemiology of haemoglobin disorders and derived service indicators. *Bulletin of the World Health Organization* **86**: 480-487.

Moerke NJ, Aktas H, Chen H, Cantel S, Reibarkh MY, Fahmy A, Gross JD, Degterev A, Yuan J, Chorev M et al. 2007. Small-molecule inhibition of the interaction between the translation initiation factors eIF4E and eIF4G. *Cell* **128**: 257-267.

Motakis E, Guhl S, Ishizu Y, Itoh M, Kawaji H, de Hoon M, Lassmann T, Carninci P, Hayashizaki Y, Zuberbier T et al. 2014. Redefinition of the human mast cell transcriptome by deep-CAGE sequencing. *Blood* **123**: e58-67.

Mugishima H, Ohga S, Ohara A, Kojima S, Fujisawa K, Tsukimoto I, Aplastic Anemia Committee of the Japanese Society of Pediatric H. 2007. Hematopoietic stem cell transplantation for Diamond-Blackfan anemia: a report from the Aplastic Anemia Committee of the Japanese Society of Pediatric Hematology. *Pediatric transplantation* **11**: 601-607.

Mukhopadhyay R, Ray PS, Arif A, Brady AK, Kinter M, Fox PL. 2008. DAPK-ZIPK-L13a axis constitutes a negative-feedback module regulating inflammatory gene expression. *Molecular cell* **32**: 371-382.

Musunuru K, Strong A, Frank-Kamenetsky M, Lee NE, Ahfeldt T, Sachs KV, Li X, Li H, Kuperwasser N, Ruda VM et al. 2010. From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. *Nature* **466**: 714-719.

Narla A, Ebert BL. 2010. Ribosomopathies: human disorders of ribosome dysfunction. *Blood* **115**: 3196-3205.

Narla A, Payne EM, Abayasekara N, Hurst SN, Raiser DM, Look AT, Berliner N, Ebert BL, Khanna-Gupta A. 2014. L-Leucine improves the anaemia in models of Diamond Blackfan anaemia and the 5q- syndrome in a TP53-independent way. *British journal of haematology* **167**: 524-528.

Nathan DG, Clarke BJ, Hillman DG, Alter BP, Housman DE. 1978a. Erythroid precursors in congenital hypoplastic (Diamond-Blackfan) anemia. *The Journal of clinical investigation* **61**: 489-498.

Nathan DG, Hillman DG, Chess L, Alter BP, Clarke BJ, Breard J, Housman DE. 1978b. Normal erythropoietic helper T cells in congenital hypoplastic (Diamond-Blackfan) anemia. *The New England journal of medicine* **298**: 1049-1051.

Nei Y, Obata-Ninomiya K, Tsutsui H, Ishiwata K, Miyasaka M, Matsumoto K, Nakae S, Kanuka H, Inase N, Karasuyama H. 2013. GATA-1 regulates the generation and function of basophils. *Proc Natl Acad Sci U S A* **110**: 18620-18625.

Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA et al. 2010. Exome sequencing identifies the cause of a mendelian disorder. *Nature genetics* **42**: 30-35.

Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, Shaffer T, Wong M, Bhattacharjee A, Eichler EE et al. 2009. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* **461**: 272-276.

Nishikawa K, Kobayashi M, Masumi A, Lyons SE, Weinstein BM, Liu PP, Yamamoto M. 2003. Self-association of Gata1 enhances transcriptional activity in vivo in zebra fish embryos. *Molecular and cellular biology* **23**: 8295-8305.

O'Brien KA, Anderson SM, Farrar J, Vlachos A, Atsidaftos E, Lipton JM, Ellis S, Bodine DM. 2014. In Vitro Analysis of Erythroid Cells Derived from the Culture of Diamond Blackfan Anemia Patient CD34+ Cells. *Blood* **124**: 744-744.

O'Donohue MF, Choesmel V, Faubladier M, Fichant G, Gleizes PE. 2010. Functional dichotomy of ribosomal proteins during the synthesis of mammalian 40S ribosomal subunits. *The Journal of cell biology* **190**: 853-866.

Orkin SH, Zon LI. 2008. Hematopoiesis: an evolving paradigm for stem cell biology. *Cell* **132**: 631-644.

Palis J. 2014. Primitive and definitive erythropoiesis in mammals. *Frontiers in physiology* **5**: 3.

Palis J, Malik J, McGrath KE, Kingsley PD. 2010. Primitive erythropoiesis in the mammalian embryo. *The International journal of developmental biology* **54**: 1011-1018.

Pan X, Ohneda O, Ohneda K, Lindeboom F, Iwata F, Shimizu R, Nagano M, Suwabe N, Philipsen S, Lim KC et al. 2005. Graded levels of GATA-1 expression modulate survival, proliferation, and differentiation of erythroid progenitors. *The Journal of biological chemistry* **280**: 22385-22394.

Parrella S, Aspesi A, Quarello P, Garelli E, Pavesi E, Carando A, Nardi M, Ellis SR, Ramenghi U, Dianzani I. 2014. Loss of GATA-1 full length as a cause of Diamond-Blackfan anemia phenotype. *Pediatric blood & cancer* **61**: 1319-1321.

Paul DS, Albers CA, Rendon A, Voss K, Stephens J, HaemGen C, van der Harst P, Chambers JC, Soranzo N, Ouwehand WH et al. 2013. Maps of open chromatin highlight cell type-restricted patterns of regulatory sequence variation at hematological trait loci. *Genome research* **23**: 1130-1141.

Pauling L, Itano HA, et al. 1949. Sickle cell anemia a molecular disease. *Science* **110**: 543-548.

Payne EM, Virgilio M, Narla A, Sun H, Levine M, Paw BH, Berliner N, Look AT, Ebert BL, Khanna-Gupta A. 2012. L-Leucine improves the anemia and developmental defects associated with Diamond-Blackfan anemia and del(5q) MDS by activating the mTOR pathway. *Blood* **120**: 2214-2224.

Pearson P, Francomano C, Foster P, Bocchini C, Li P, McKusick V. 1994. The status of online Mendelian inheritance in man (OMIM) medio 1994. *Nucleic acids research* **22**: 3470-3473.

Perdahl EB, Naprstek BL, Wallace WC, Lipton JM. 1994. Erythroid failure in Diamond-Blackfan anemia is characterized by apoptosis. *Blood* **83**: 645-650.

Pereboom TC, Bondt A, Pallaki P, Klasson TD, Goos YJ, Essers PB, Groot Koerkamp MJ, Gazda HT, Holstege FC, Costa LD et al. 2014. Translation of branched-chain aminotransferase-1 transcripts is impaired in cells haploinsufficient for ribosomal protein genes. *Experimental hematology* **42**: 394-403 e394.

Perrotta S, Gallagher PG, Mohandas N. 2008. Hereditary spherocytosis. *Lancet* **372**: 1411-1426.

Pevny L, Lin CS, D'Agati V, Simon MC, Orkin SH, Costantini F. 1995. Development of hematopoietic cells lacking transcription factor GATA-1. *Development* **121**: 163-172.

Piel FB, Weatherall DJ. 2014. The alpha-thalassemias. *The New England journal of medicine* **371**: 1908-1916.

Pishesha N, Thiru P, Shi J, Eng JC, Sankaran VG, Lodish HF. 2014. Transcriptional divergence and conservation of human and mouse erythropoiesis. *Proc Natl Acad Sci U S A* **111**: 4103-4108.

Pospisilova D, Cmejlova J, Hak J, Adam T, Cmejla R. 2007. Successful treatment of a Diamond-Blackfan anemia patient with amino acid leucine. *Haematologica* **92**: e66-67.

Possemato R, Marks KM, Shaul YD, Pacold ME, Kim D, Birsoy K, Sethumadhavan S, Woo HK, Jang HG, Jha AK et al. 2011. Functional genomics reveal that the serine synthesis pathway is essential in breast cancer. *Nature* **476**: 346-350.

Pott S, Lieb JD. 2015. What are super-enhancers? *Nature genetics* **47**: 8-12.

Ramenghi U, Garelli E, Valtolina S, Campagnoli MF, Timeus F, Crescenzio N, Mair M, Varotto S, D'Avanzo M, Nobili B et al. 1999. Diamond-Blackfan anaemia in the Italian population. *British journal of haematology* **104**: 841-848.

Rao S, Lee SY, Gutierrez A, Perrigoue J, Thapa RJ, Tu Z, Jeffers JR, Rhodes M, Anderson S, Oravecz T et al. 2012. Inactivation of ribosomal protein L22 promotes transformation by induction of the stemness factor, Lin28B. *Blood* **120**: 3764-3773.

Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES et al. 2014. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**: 1665-1680.

Raychaudhuri S. 2011. Mapping rare and common causal alleles for complex human diseases. *Cell* **147**: 57-69.

Reschke M, Clohessy JG, Seitzer N, Goldstein DP, Breitkopf SB, Schmolze DB, Ala U, Asara JM, Beck AH, Pandolfi PP. 2013. Characterization and analysis of the composition and dynamics of the mammalian riboproteome. *Cell reports* **4**: 1276-1287.

Ribeil JA, Zermati Y, Vandekerckhove J, Cathelin S, Kersual J, Dussiot M, Coulon S, Moura IC, Zeuner A, Kirkegaard-Sorensen T et al. 2007. Hsp70 regulates erythropoiesis by preventing caspase-3-mediated cleavage of GATA-1. *Nature* **445**: 102-105.

Robb L, Lyons I, Li R, Hartley L, Kontgen F, Harvey RP, Metcalf D, Begley CG. 1995. Absence of yolk sac hematopoiesis from mice with a targeted disruption of the scl gene. *Proc Natl Acad Sci U S A* **92**: 7075-7079.

Robledo S, Idol RA, Crimmins DL, Ladenson JH, Mason PJ, Bessler M. 2008. The role of human ribosomal proteins in the maturation of rRNA and ribosome production. *Rna* **14**: 1918-1929.

Roy V, Perez WS, Eapen M, Marsh JC, Pasquini M, Pasquini R, Mustafa MM, Bredeson CN, Non-Malignant Marrow Disorders Working Committee of the International Bone Marrow Transplant R. 2005. Bone marrow transplantation for diamond-blackfan anemia. *Biology of blood and marrow transplantation : journal of the American Society for Blood and Marrow Transplantation* **11**: 600-608.

Ruggero D, Shimamura A. 2014. Marrow failure: a window into ribosome biology. *Blood* **124**: 2784-2792.

Sankaran VG, Ghazvinian R, Do R, Thiru P, Vergilio JA, Beggs AH, Sieff CA, Orkin SH, Nathan DG, Lander ES et al. 2012a. Exome sequencing identifies GATA1 mutations resulting in Diamond-Blackfan anemia. *The Journal of clinical investigation* **122**: 2439-2443.

Sankaran VG, Ludwig LS, Sicinska E, Xu J, Bauer DE, Eng JC, Patterson HC, Metcalf RA, Natkunam Y, Orkin SH et al. 2012b. Cyclin D3 coordinates the cell cycle during differentiation to regulate erythrocyte size and number. *Genes & development* **26**: 2075-2087.

Sankaran VG, Menne TF, Xu J, Akie TE, Lettre G, Van Handel B, Mikkola HK, Hirschhorn JN, Cantor AB, Orkin SH. 2008. Human fetal hemoglobin expression is regulated by the developmental stage-specific repressor BCL11A. *Science* **322**: 1839-1842.

Sankaran VG, Orkin SH. 2013. Genome-wide association studies of hematologic phenotypes: a window into human hematopoiesis. *Current opinion in genetics & development* **23**: 339-344.

Sankaran VG, Weiss MJ. 2015. Anemia: progress in molecular mechanisms and therapies. *Nature medicine* **21**: 221-230.

Sankaran VG, Xu J, Ragoczy T, Ippolito GC, Walkley CR, Maika SD, Fujiwara Y, Ito M, Groudine M, Bender MA et al. 2009. Developmental and species-divergent globin switching are driven by BCL11A. *Nature* **460**: 1093-1097.

Santini E, Huynh TN, MacAskill AF, Carter AG, Pierre P, Ruggero D, Kaphzan H, Klann E. 2013. Exaggerated translation causes synaptic and behavioural aberrations associated with autism. *Nature* **493**: 411-415.

Schutz S, Fischer U, Altvater M, Nerurkar P, Pena C, Gerber M, Chang Y, Caesar S, Schubert OT, Schlenstedt G et al. 2014. A RanGTP-independent mechanism allows ribosomal protein nuclear import for ribosome assembly. *eLife* **3**: e03473.

Schwanhausser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, Chen W, Selbach M. 2011. Global quantification of mammalian gene expression control. *Nature* **473**: 337-342.

Sermet-Gaudelus I, Boeck KD, Casimir GJ, Vermeulen F, Leal T, Mogenet A, Roussel D, Fritsch J, Hanssens L, Hirawat S et al. 2010. Ataluren (PTC124) induces cystic fibrosis transmembrane conductance regulator protein expression and activity in children with nonsense mutation cystic fibrosis. *American journal of respiratory and critical care medicine* **182**: 1262-1272.

Service S, DeYoung J, Karayiorgou M, Roos JL, Pretorious H, Bedoya G, Ospina J, Ruiz-Linares A, Macedo A, Palha JA et al. 2006. Magnitude and distribution of linkage disequilibrium in population isolates and implications for genome-wide association studies. *Nature genetics* **38**: 556-560.

Shah P, Ding Y, Niemczyk M, Kudla G, Plotkin JB. 2013. Rate-limiting steps in yeast protein translation. *Cell* **153**: 1589-1601.

Shalem O, Sanjana NE, Hartenian E, Shi X, Scott DA, Mikkelsen TS, Heckl D, Ebert BL, Root DE, Doench JG et al. 2014. Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343**: 84-87.

Shapiro GI. 2006. Cyclin-dependent kinase pathways as targets for cancer treatment. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **24**: 1770-1783.

Sherr CJ, Roberts JM. 1999. CDK inhibitors: positive and negative regulators of G1-phase progression. *Genes & development* **13**: 1501-1512.

Shimizu R, Hasegawa A, Ottolenghi S, Ronchi A, Yamamoto M. 2013. Verification of the in vivo activity of three distinct cis-acting elements within the Gata1 gene promoter-proximal enhancer in mice. *Genes to cells : devoted to molecular & cellular mechanisms* **18**: 1032-1041.

Shiraki T, Kondo S, Katayama S, Waki K, Kasukawa T, Kawaji H, Kodzius R, Watahiki A, Nakamura M, Arakawa T et al. 2003. Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. *Proc Natl Acad Sci U S A* **100**: 15776-15781.

Shivdasani RA, Fujiwara Y, McDevitt MA, Orkin SH. 1997. A lineage-selective knockout establishes the critical role of transcription factor GATA-1 in megakaryocyte growth and platelet development. *The EMBO journal* **16**: 3965-3973.

Shivdasani RA, Mayer EL, Orkin SH. 1995. Absence of blood formation in mice lacking the T-cell leukaemia oncoprotein tal-1/SCL. *Nature* **373**: 432-434.

Sicinska E, Aifantis I, Le Cam L, Swat W, Borowski C, Yu Q, Ferrando AA, Levin SD, Geng Y, von Boehmer H et al. 2003. Requirement for cyclin D3 in lymphocyte development and T cell leukemias. *Cancer cell* **4**: 451-461.

Sicinska E, Lee YM, Gits J, Shigematsu H, Yu Q, Rebel VI, Geng Y, Marshall CJ, Akashi K, Dorfman DM et al. 2006. Essential role for cyclin D3 in granulocyte colony-stimulating factor-driven expansion of neutrophil granulocytes. *Molecular and cellular biology* **26**: 8052-8060.

Sicinski P, Donaher JL, Geng Y, Parker SB, Gardner H, Park MY, Robker RL, Richards JS, McGinnis LK, Biggers JD et al. 1996. Cyclin D2 is an FSH-responsive gene involved in gonadal cell proliferation and oncogenesis. *Nature* **384**: 470-474.

Sicinski P, Donaher JL, Parker SB, Li T, Fazeli A, Gardner H, Haslam SZ, Bronson RT, Elledge SJ, Weinberg RA. 1995. Cyclin D1 provides a link between development and oncogenesis in the retina and breast. *Cell* **82**: 621-630.

Siegfried NA, Busan S, Rice GM, Nelson JA, Weeks KM. 2014. RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nature methods* **11**: 959-965.

Signer RA, Magee JA, Salic A, Morrison SJ. 2014. Haematopoietic stem cells require a highly regulated protein synthesis rate. *Nature* **509**: 49-54.

Singh SA, Goldberg TA, Henson AL, Husain-Krautter S, Nihrane A, Blanc L, Ellis SR, Lipton JM, Liu JM. 2014. p53-Independent cell cycle and erythroid differentiation defects in murine embryonic stem cells haploinsufficient for Diamond Blackfan anemia-proteins: RPS19 versus RPL5. *PloS one* **9**: e89098.

Sjogren SE, Flygare J. 2012. Progress towards mechanism-based treatment for Diamond-Blackfan anemia. *TheScientificWorldJournal* **2012**: 184362.

Slatkin M. 2008. Linkage disequilibrium--understanding the evolutionary past and mapping the medical future. *Nature reviews Genetics* **9**: 477-485.

Slowikowski K, Hu X, Raychaudhuri S. 2014. SNPsea: an algorithm to identify cell types, tissues and pathways affected by risk loci. *Bioinformatics* **30**: 2496-2497.

Smemo S, Tena JJ, Kim KH, Gamazon ER, Sakabe NJ, Gomez-Marin C, Aneas I, Credidio FL, Sobreira DR, Wasserman NF et al. 2014. Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature* **507**: 371-375.

Socolovsky M, Fallon AE, Wang S, Brugnara C, Lodish HF. 1999. Fetal anemia and apoptosis of red cell progenitors in Stat5a-/-5b-/- mice: a direct role for Stat5 in Bcl-X(L) induction. *Cell* **98**: 181-191.

Socolovsky M, Nam H, Fleming MD, Haase VH, Brugnara C, Lodish HF. 2001. Ineffective erythropoiesis in Stat5a(-/-)5b(-/-) mice due to decreased survival of early erythroblasts. *Blood* **98**: 3261-3273.

Solis C, Aizencang GI, Astrin KH, Bishop DF, Desnick RJ. 2001. Uroporphyrinogen III synthase erythroid promoter mutations in adjacent GATA1 and CP2 elements cause congenital erythropoietic porphyria. *The Journal of clinical investigation* **107**: 753-762.

Sonenberg N, Hinnebusch AG. 2009. Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell* **136**: 731-745.

Soranzo N, Spector TD, Mangino M, Kuhnel B, Rendon A, Teumer A, Willenborg C, Wright B, Chen L, Li M et al. 2009. A genome-wide meta-analysis identifies 22 loci associated with eight hematological parameters in the HaemGen consortium. *Nature genetics* **41**: 1182-1190.

Stonestrom AJ, Hsu SC, Jahn KS, Huang P, Keller CA, Giardine BM, Kadauke S, Campbell AE, Evans P, Hardison RC et al. 2015. Functions of BET proteins in erythroid gene expression. *Blood* **125**: 2825-2834.

Stranger BE, Stahl EA, Raj T. 2011. Progress and promise of genome-wide association studies for human complex trait genetics. *Genetics* **187**: 367-383.

Subtelny AO, Eichhorn SW, Chen GR, Sive H, Bartel DP. 2014. Poly(A)-tail profiling reveals an embryonic switch in translational control. *Nature* **508**: 66-71.

Sulima SO, Patchett S, Advani VM, De Keersmaecker K, Johnson AW, Dinman JD. 2014. Bypass of the pre-60S ribosomal quality control as a pathway to oncogenesis. *Proc Natl Acad Sci U S A* **111**: 5640-5645.

Suragani RN, Cadena SM, Cawley SM, Sako D, Mitchell D, Li R, Davies MV, Alexander MJ, Devine M, Loveday KS et al. 2014. Transforming growth factor-beta superfamily ligand trap ACE-536 corrects anemia by promoting late-stage erythropoiesis. *Nature medicine* **20**: 408-414.

Takai J, Moriguchi T, Suzuki M, Yu L, Ohneda K, Yamamoto M. 2013. The Gata1 5' region harbors distinct cis-regulatory modules that direct gene activation in erythroid cells and gene inactivation in HSCs. *Blood* **122**: 3450-3460.

Teng T, Mercer CA, Hexley P, Thomas G, Fumagalli S. 2013. Loss of tumor suppressor RPL5/RPL11 does not induce cell cycle arrest but impedes proliferation due to reduced ribosome content and translation capacity. *Molecular and cellular biology* **33**: 4660-4671.

Teslovich TM Musunuru K Smith AV Edmondson AC Stylianou IM Koseki M Pirruccello JP Ripatti S Chasman DI Willer CJ et al. 2010. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**: 707-713.

Thapa M, Bommakanti A, Shamsuzzaman M, Gregory B, Samsel L, Zengel JM, Lindahl L. 2013. Repressed synthesis of ribosomal proteins generates protein-specific cell cycle and morphological phenotypes. *Molecular biology of the cell* **24**: 3620-3633.

Thom CS, Traxler EA, Khandros E, Nickas JM, Zhou OY, Lazarus JE, Silva AP, Prabhu D, Yao Y, Aribeana C et al. 2014. Trim58 degrades Dynein and regulates terminal erythropoiesis. *Developmental cell* **30**: 688-700.

Thoreen CC, Chantranupong L, Keys HR, Wang T, Gray NS, Sabatini DM. 2012. A unifying model for mTORC1-mediated regulation of mRNA translation. *Nature* **485**: 109-113.

Thoreen CC, Kang SA, Chang JW, Liu Q, Zhang J, Gao Y, Reichling LJ, Sim T, Sabatini DM, Gray NS. 2009. An ATP-competitive mammalian target of rapamycin inhibitor reveals rapamycin-resistant functions of mTORC1. *The Journal of biological chemistry* **284**: 8023-8032.

Torihara H, Uechi T, Chakraborty A, Shinya M, Sakai N, Kenmochi N. 2011. Erythropoiesis failure due to RPS19 deficiency is independent of an activated Tp53 response in a zebrafish model of Diamond-Blackfan anaemia. *British journal of haematology* **152**: 648-654.

Tournamille C, Colin Y, Cartron JP, Le Van Kim C. 1995. Disruption of a GATA motif in the Duffy gene promoter abolishes erythroid gene expression in Duffy-negative individuals. *Nature genetics* **10**: 224-228.

Trainor CD, Mas C, Archambault P, Di Lello P, Omichinski JG. 2009. GATA-1 associates with and inhibits p53. *Blood* **114**: 165-173.

Tsang AP, Fujiwara Y, Hom DB, Orkin SH. 1998. Failure of megakaryopoiesis and arrested erythropoiesis in mice lacking the GATA-1 transcriptional cofactor FOG. *Genes & development* **12**: 1176-1188.

Tzoneva G, Perez-Garcia A, Carpenter Z, Khiabanian H, Tosello V, Allegretta M, Paietta E, Racevskis J, Rowe JM, Tallman MS et al. 2013. Activating mutations in the NT5C2 nucleotidase gene drive chemotherapy resistance in relapsed ALL. *Nature medicine* **19**: 368-371.

Uda M, Galanello R, Sanna S, Lettre G, Sankaran VG, Chen W, Usala G, Busonero F, Maschio A, Albai G et al. 2008. 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* **105**: 1620-1625.

Ulirsch JC, Lacy JN, An X, Mohandas N, Mikkelsen TS, Sankaran VG. 2014. Altered chromatin occupancy of master regulators underlies evolutionary divergence in the transcriptional landscape of erythroid differentiation. *PLoS Genet* **10**: e1004890.

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. 2012. Seventy-five genetic loci influencing the human red blood cell. *Nature* **492**: 369-375.

van Wijk R, van Solinge WW. 2005. The energy-less red blood cell is lost: erythrocyte enzyme abnormalities of glycolysis. *Blood* **106**: 4034-4042.

Visscher PM, Brown MA, McCarthy MI, Yang J. 2012. Five years of GWAS discovery. *American journal of human genetics* **90**: 7-24.

Vlachos A, Ball S, Dahl N, Alter BP, Sheth S, Ramenghi U, Meerpohl J, Karlsson S, Liu JM, Leblanc T et al. 2008. Diagnosing and treating Diamond Blackfan anaemia: results of an international clinical consensus conference. *British journal of haematology* **142**: 859-876.

Vlachos A, Rosenberg PS, Atsidaftos E, Alter BP, Lipton JM. 2012. Incidence of neoplasia in Diamond Blackfan anemia: a report from the Diamond Blackfan Anemia Registry. *Blood* **119**: 3815-3819.

Walkley CR, Sankaran VG, Orkin SH. 2008. Rb and hematopoiesis: stem cells to anemia. *Cell division* **3**: 13.

Wall L, deBoer E, Grosveld F. 1988. The human beta-globin gene 3' enhancer contains multiple binding sites for an erythroid-specific protein. *Genes & development* **2**: 1089-1100.

Wang K, Kan J, Yuen ST, Shi ST, Chu KM, Law S, Chan TL, Kan Z, Chan AS, Tsui WY et al. 2011. Exome sequencing identifies frequent mutation of ARID1A in molecular subtypes of gastric cancer. *Nature genetics* **43**: 1219-1223.

Wang R, Yoshida K, Toki T, Sawada T, Uechi T, Okuno Y, Sato-Otsubo A, Kudo K, Kamimaki I, Kanezaki R et al. 2015. Loss of function mutations in RPL27 and RPS27 identified by whole-exome sequencing in Diamond-Blackfan anaemia. *British journal of haematology* **168**: 854-864.

Wang T, Wei JJ, Sabatini DM, Lander ES. 2014. Genetic screens in human cells using the CRISPR-Cas9 system. *Science* **343**: 80-84.

Warner JR, McIntosh KB. 2009. How common are extraribosomal functions of ribosomal proteins? *Molecular cell* **34**: 3-11.

Warren AJ, Colledge WH, Carlton MB, Evans MJ, Smith AJ, Rabbitts TH. 1994. The oncogenic cysteine-rich LIM domain protein rbtn2 is essential for erythroid development. *Cell* **78**: 45-57.

Wechsler J, Greene M, McDevitt MA, Anastasi J, Karp JE, Le Beau MM, Crispino JD. 2002. Acquired mutations in GATA1 in the megakaryoblastic leukemia of Down syndrome. *Nature genetics* **32**: 148-152.

Weiss MJ, Mason PJ, Bessler M. 2012. What's in a name? *The Journal of clinical investigation* **122**: 2346-2349.

Weiss MJ, Orkin SH. 1995. Transcription factor GATA-1 permits survival and maturation of erythroid precursors by preventing apoptosis. *Proc Natl Acad Sci U S A* **92**: 9623-9627.

Welch EM, Barton ER, Zhuo J, Tomizawa Y, Friesen WJ, Trifillis P, Paushkin S, Patel M, Trotta CR, Hwang S et al. 2007. PTC124 targets genetic disorders caused by nonsense mutations. *Nature* **447**: 87-91.

155

Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, Klemm A, Flicek P, Manolio T, Hindorff L et al. 2014. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic acids research* **42**: D1001-1006.

Westra HJ, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, Christiansen MW, Fairfax BP, Schramm K, Powell JE et al. 2013. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nature genetics* **45**: 1238-1243.

Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH, Rahl PB, Lee TI, Young RA. 2013. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**: 307-319.

Wianny F, Real FX, Mummery CL, Van Rooijen M, Lahti J, Samarut J, Savatier P. 1998. G1-phase regulators, cyclin D1, cyclin D2, and cyclin D3: up-regulation at gastrulation and dynamic expression during neurulation. *Developmental dynamics : an official publication of the American Association of Anatomists* **212**: 49-62.

Willig TN, Niemeyer CM, Leblanc T, Tiemann C, Robert A, Budde J, Lambiliotte A, Kohne E, Souillet G, Eber S et al. 1999. Identification of new prognosis factors from the clinical and epidemiologic analysis of a registry of 229 Diamond-Blackfan anemia patients. DBA group of Societe d'Hematologie et d'Immunologie Pediatrique (SHIP), Gesellshaft fur Padiatrische Onkologie und Hamatologie (GPOH), and the European Society for Pediatric Hematology and Immunology (ESPHI). *Pediatric research* **46**: 553-561.

Wilschanski M, Miller LL, Shoseyov D, Blau H, Rivlin J, Aviram M, Cohen M, Armoni S, Yaakov Y, Pugatsch T et al. 2011. Chronic ataluren (PTC124) treatment of nonsense mutation cystic fibrosis. *The European respiratory journal* **38**: 59-69.

Wood AR Esko T Yang J Vedantam S Pers TH Gustafsson S Chu AY Estrada K Luan J Kutalik Z et al. 2014. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature genetics* **46**: 1173-1186.

Wu H, Liu X, Jaenisch R, Lodish HF. 1995. Generation of committed erythroid BFU-E and CFU-E progenitors does not require erythropoietin or the erythropoietin receptor. *Cell* **83**: 59-67.

Xue S, Barna M. 2012. Specialized ribosomes: a new frontier in gene regulation and organismal biology. *Nature reviews Molecular cell biology* **13**: 355-369.

Xue S, Tian S, Fujii K, Kladwang W, Das R, Barna M. 2015. RNA regulons in Hox 5' UTRs confer ribosome specificity to gene regulation. *Nature* **517**: 33-38.

Yadav GV, Chakraborty A, Uechi T, Kenmochi N. 2014. Ribosomal protein deficiency causes Tp53-independent erythropoiesis failure in zebrafish. *The international journal of biochemistry & cell biology* **49**: 1-7.

Yang Y, Muzny DM, Reid JG, Bainbridge MN, Willis A, Ward PA, Braxton A, Beuten J, Xia F, Niu Z et al. 2013. Clinical whole-exome sequencing for the diagnosis of mendelian disorders. *The New England journal of medicine* **369**: 1502-1511.

Yang Y, Muzny DM, Xia F, Niu Z, Person R, Ding Y, Ward P, Braxton A, Wang M, Buhay C et al. 2014. Molecular findings among patients referred for clinical whole-exome sequencing. *Jama* **312**: 1870-1879.

Yien YY, Robledo RF, Schultz IJ, Takahashi-Makise N, Gwynn B, Bauer DE, Dass A, Yi G, Li L, Hildick-Smith GJ et al. 2014. TMEM14C is required for erythroid mitochondrial heme metabolism. *The Journal of clinical investigation* **124**: 4294-4304.

Yip BH, Vuppusetty C, Attwood M, Giagounidis A, Germing U, Lamikanra AA, Roberts DJ, Maciejewski JP, Vandenberghe P, Mecucci C et al. 2013. Activation of the mTOR signaling pathway by L-leucine in 5q- syndrome and other RPS14-deficient erythroblasts. *Leukemia* **27**: 1760-1763.

Yoshida K, Toki T, Okuno Y, Kanezaki R, Shiraishi Y, Sato-Otsubo A, Sanada M, Park MJ, Terui K, Suzuki H et al. 2013. The landscape of somatic mutations in Down syndrome-related myeloid disorders. *Nature genetics* **45**: 1293-1299.

Yu L, Ji W, Zhang H, Renda MJ, He Y, Lin S, Cheng EC, Chen H, Krause DS, Min W. 2010. SENP1-mediated GATA1 deSUMOylation is critical for definitive erythropoiesis. *The Journal of experimental medicine* **207**: 1183-1195.

Zhang L, Prak L, Rayon-Estrada V, Thiru P, Flygare J, Lim B, Lodish HF. 2013a. ZFP36L2 is required for self-renewal of early burst-forming unit erythroid progenitors. *Nature* **499**: 92-96.

Zhang Y, Duc AC, Rao S, Sun XL, Bilbee AN, Rhodes M, Li Q, Kappes DJ, Rhodes J, Wiest DL. 2013b. Control of hematopoietic stem cell emergence by antagonistic functions of ribosomal protein paralogs. *Developmental cell* **24**: 411-425.

Zhang Y, Ear J, Yang Z, Morimoto K, Zhang B, Lin S. 2014. Defects of protein production in erythroid cells revealed in a zebrafish Diamond-Blackfan anemia model for mutation in RPS19. *Cell death & disease* **5**: e1352.

Zhang Y, Lu H. 2009. Signaling to p53: ribosomal proteins find their way. *Cancer cell* **16**: 369-377.

10. Publications

Publications presented within the scope of this thesis

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.

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.

* These authors contributed equally to this work

Additional Publications

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.

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.

* These authors contributed equally to this work

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.

Under revision

Giani FG, **Ludwig LS**, Wakabayashi A, 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.