

Mutations in linker-2 of KLF1 impair expression of membrane transporters and cytoskeletal proteins causing hemolysis

Received: 13 March 2023

Accepted: 9 July 2024

Published online: 15 August 2024

 Check for updatesStephen Huang^{1,2}, Casie Reed³, Melissa Ilsley^{1,2}, Graham Magor^{1,3}, Michael Tallack¹, Michael Landsberg⁴, Helen Mitchell³, Kevin Gillinder^{1,3} & Andrew Perkins^{1,2,3,5,6} ✉

The SP/KLF family of transcription factors harbour three C-terminal C2H2 zinc fingers interspersed by two linkers which confers DNA-binding to a 9-10 bp motif. Mutations in *KLF1*, the founding member of the family, are common. Missense mutations in linker two result in a mild phenotype. However, when co-inherited with loss-of-function mutations, they result in severe non-spherocytic hemolytic anemia. We generate a mouse model of this disease by crossing *Klf1*^{+/-} mice with *Klf1*^{H350R/+} mice that harbour a missense mutation in linker-2. *Klf1*^{H350R/-} mice exhibit severe hemolysis without thalassemia. RNA-seq demonstrate loss of expression of genes encoding transmembrane and cytoskeletal proteins, but not globins. ChIP-seq show no change in DNA-binding specificity, but a global reduction in affinity, which is confirmed using recombinant proteins and in vitro binding assays. This study provides new insights into how linker mutations in zinc finger transcription factors result in different phenotypes to those caused by loss-of-function mutations.

There are >500 C2H2 zinc finger (ZF) proteins encoded in the mammalian genome^{1,2}. They can be sub-classified according to the number of ZFs, the presence of additional functional domains (e.g. a KRAB domain), and the presence or absence of highly-conserved linker sequences between individual ZFs. The presence of the linker, TGEKP, is highly predictive of a DNA-binding function rather than another function such as RNA or protein binding. For most C2H2 zinc finger proteins there is limited information about function, but this is not the case for the SP/KLF subfamily which are ancient^{3,4}. They play critical regulatory roles in the differentiation, proliferation, identity, and reprogramming of most cell types and tissues^{5,6}.

There are twenty-five SP/KLF proteins in humans and mice; they share three very similar C-terminal C2H2 ZFs interspaced by highly-conserved linkers^{3,7}. Linker 1 nearly always has the sequence, ¹TGE⁴K⁵P, whereas linker 2 is more variable (Supplementary Fig. 1A). Nevertheless, the threonine at position +1 (relative to the end of the preceding α -helix),

arginine or lysine at position +4, and proline at position +5, are highly conserved. NMR and crystallographic studies show linkers become structured in the presence of DNA, forming hydrogen bonds with the α -helix of the preceding ZF^{8,9}. In this way, the linker stabilizes the preceding α -helix and increases its length. This process is known as α -helix C-capping¹⁰. It is thought that the linkers also facilitate 'clamping' of the three ZFs to their extended 9 bp binding site in the major groove of DNA, and thereby increase DNA-binding affinity.

Many SP and KLF family members are broadly expressed whereas others are tissue specific⁵. *KLF1*, the founding member of the KLF family, is expressed uniquely in erythroid progenitor cells¹¹, and is essential for red blood cell production in mice and man¹²⁻¹⁴. ChIP-seq and RNA-seq experiments have shown KLF1 directly binds promoters and enhancers of genes which regulate the cell cycle, heme biosynthesis, globin production, and membrane/cytoskeletal integrity¹⁵⁻¹⁹. Missense mutations in the second linker of *KLF1* are quite common in

¹Mater Research Institute - UQ, The University of Queensland, St Lucia, Australia. ²School of Biomedical Sciences, The University of Queensland, St Lucia, Australia. ³Australian Centre for Blood Diseases, Monash University, Clayton, Australia. ⁴School of Chemistry and Molecular Biosciences, The University of Queensland, St Lucia, Australia. ⁵Department of Haematology, The Alfred Hospital, Melbourne, Australia. ⁶Biodiscovery Institute, Monash University, Clayton, Australia. ✉e-mail: Andrew.perkins@monash.edu

some populations such as those from Southern China and the Mediterranean (Supplementary Fig. 1B)^{20–22}. Missense mutations are less common in peoples of European ancestry so they are under-represented in large exome databases such as those accessible via gnomAD and the UK Biobank. The combination of a missense mutation in the second linker of *KLF1* with a loss-of-function mutation on the other allele results in non-spherocytic hemolytic anemia (NSHA) rather than β -thalassaemia²³, for reasons that are unclear.

Little is known about the biological functions of C2H2 linkers in vivo. We previously identified the mommed45 mutation in *Klf1* from a mouse ENU mutagenesis screen for epigenetic modifiers of expression of an α -globin transgene²⁴. The mutation results in a single amino acid substitution of histidine to arginine (H350R) at position +3 (underlined) in the second linker (¹TG³HRP) (Supplementary Fig. 1B). Interestingly, this histidine is not conserved between mouse and man; however, it is conserved in the closely related family members, KLF2 and KLF4 (Supplementary Fig. 1A). The +3 position in linker 2 (and in linker 1) is an acidic amino acid (glutamate or aspartate) for most KLF/SP family members (Supplementary Fig. 1A), so the basic amino acid (arginine) swap, H350R, is likely damaging to function. Indeed, in the homozygous state, the H350R mutation causes mild anemia with reticulocytosis and splenomegaly²⁴.

Herein, we established an animal model of human NSHA by crossing *Klf1* knockout mice¹² with mommed45 mice²⁴. The mouse model accurately phenocopies humans who harbor similar compound heterozygous mutations in *KLF1*^{20,21}. We gained insight into how this mutation results in disease by performing ChIP-seq and RNA-seq in fetal liver (FL) and adult splenic erythroblasts. We show just a small subset of the KLF1-dependent transcriptome is sensitive to the H350R mutation. The set includes genes that encode transmembrane proteins [e.g. *Slc2a1* (Glut1), *Slc2a4* (Glut4), *Slc1a5* (Asct2) and *Slc43a1* (Lat3)] and cytoskeletal proteins (e.g. *Ebp4.9/dematin*), but surprisingly not some well-established KLF1 target genes such as globins (e.g. *Hbb-b1*) and cell cycle regulators (e.g. *e2f2* and *e2f4*)^{11,12,25,26}. This is consistent with the NSHA phenotype rather than thalassaemia. We partially and transiently rescued the anemia with L-leucine supplementation in the drinking water. This provides a translational opportunity for humans with NSHA due to *KLF1* mutations. We show that the H350R mutation leads to reduced affinity of DNA-binding globally in vivo and in vitro, but unaltered sequence specificity, distinguishing it from other mutations in DNA-contacting residues of KLF1^{27–31}. We expand the in vitro analysis to other anemia-causing linker 2 mutations in humans. Similar mutations in linkers of other C2H2 zinc finger TFs have been reported in various human genetic disorders^{32,33}. Our studies provide a greater understanding of the global in vivo functions of the linkers of the C2H2 super family of TFs, and how these mutations cause diseases that are phenotypically distinct from diseases due to loss of function mutations.

Results

Compound heterozygous mice for null and missense mutations in linker 2 of *Klf1* display NSHA

We interbred mice carrying the mommed45 mutation in *Klf1* (*Klf1*^{H350R/+})²⁴ with mice carrying a null allele in *Klf1* (*Klf1*^{-/-})¹² to create *Klf1*^{H350R/-} mice and litter mates with five other genotypes. *Klf1*^{-/-} mice are embryonic lethal but *Klf1*^{H350R/-} mice were born at expected Mendelian ratios. However, they were severely pale and jaundiced. The phenotype resembles that which occurs in human patients who co-inherit loss-of-function and linker 2 missense mutations in *KLF1*^{20,21}. There was severe poikilocytosis, with prominent schistocytes, micro-spherocytes, pencil cells, and nucleated red blood cells (Fig. 1A; arrow) in blood smears from *Klf1*^{H350R/-} mice (target cells and stomatocytes were not prominent), whereas blood smears from heterozygote mice (*Klf1*^{+/-} and *Klf1*^{H350R/+}) and wild type mice (*Klf1*^{+/+}) were normal. Blood smears from *Klf1*^{H350R/H350R} mice showed slight polychromasia. There was a mild reduction in mean corpuscular volume

(MCV), mean corpuscular hemoglobin (MCH) and hematocrit (Hct), and a marked increase in circulating reticulocytes in *Klf1*^{H350R/-} mice compared to litter mates (Fig. 1B). We found an increase in normalized spleen weight (to body weight) (Fig. 1E), due to expansion of the red pulp (Fig. 1D). There was a marked increase in the number of CD71+ erythroid cells in the spleens of *Klf1*^{H350R/-} mice at the relative expense of B and T cells (Fig. 1F). Together, this indicates the splenomegaly was due to extra-medullary erythropoiesis. There was an increase in immature erythrocytes (CD71+TER119+) in the blood of *Klf1*^{H350R/-} mice compared to *Klf1*^{+/-} litter mates and a corresponding decrease in CD71/TER119+ mature erythrocytes (Fig. 1C). There was also a reduction in mature CD71/TER119+ erythroid cells in the bone marrow (Fig. 1C). Together with the blood findings, this suggests either impairment of terminal erythroid differentiation or destruction prior to bone marrow exit with premature release of immature erythrocytes into the circulation; i.e. a stress response.

Klf1^{H350R/-} embryos display reduced transcription of a small subset of KLF1-dependent genes

To understand the transcriptional consequences of the H350R mutation we performed mRNA-seq on *Klf1*^{+/-}, *Klf1*^{H350R/-}, and *Klf1*^{-/-} fetal liver (FLs). Crosses were undertaken so that these three genotypes were present within the same litters; thus, appropriate comparisons could be made between the allelic series in the context of one *Klf1* null allele. At E14.5, *Klf1*^{-/-} embryos were severely anemic as expected¹², but *Klf1*^{H350R/-} embryos were phenotypically indistinguishable from *Klf1*^{+/-} embryos. MGG-stained cytopins of the blood and fetal liver cells are shown in Supplementary Fig. 2. The circulating cells in *Klf1*^{-/-} mice are extremely fragile as reported³⁴. There are late erythroblasts in the fetal livers all genotypes (despite the lack of Ter119 expression) and some level of hemoglobinization, but the membranes are irregular in *Klf1*^{-/-} and *Klf1*^{H350R/-} compared with *Klf1*^{+/-} mice. We identified 1786 differentially expressed genes (DEGs) between *Klf1*^{+/-} and *Klf1*^{-/-} FLs to re-establish the KLF1-dependent transcriptome in this cross (Supplementary Table 3). The majority of these genes were down regulated in the absence of *Klf1* consistent with prior studies showing KLF1 acts primarily as a transcriptional activator in vivo¹⁵. As expected, this gene list is similar to previously published lists in BALB/c mice^{15,18,34}.

We next identified just 228 genes that are sensitive to the H350R missense mutation by determining DEGs between *Klf1*^{+/-} and *Klf1*^{H350R/-} FL RNAs (Fig. 2A); 168 of these overlap with the KLF1-dependent gene list and 60 do not (Fig. 2A). Most of the genes display reduced gene expression in *Klf1*^{H350R/-} FL, but some are increased (Fig. 2B). In fact, there is a range of transcriptional sensitivity to the H350R mutation, with many genes displaying slightly reduced expression levels, but failing to meet statistical cut-offs (Fig. 2B). Thus, many H350R-sensitive genes have an intermediate expression level between *Klf1*^{-/-} and *Klf1*^{+/-} as would be expected for a hypo-morphic mutation (Supplementary Fig. 3). For the significantly downregulated DEGs, we found enrichment of gene ontologies for the membrane cellular component and biological processes relating to the membrane (e.g. transmembrane transport and ion transport), and we found no significant differences in the expression of genes central to the function of KLF1 such as hemoglobin and cell cycle transcription factors (*E2f2*, *E2f4*)^{19,35}. We undertook qRT-PCR to validate H350R-sensitive genes in comparison with known *Klf1* target genes such as β -globin (*Hbb-b1*)^{11,15}. Consistent with the RNA-seq data, *Hbb-b1*, *Gata2*, *Gata1* and *Fli1* are normally expressed in the FLs of *Klf1*^{H350R/-} mice (Fig. 2B, D). *E2f2*, *E2f4*, *Tspo2*, and *Dusp1*^{15,18,26}, were expressed at intermediate levels in *Klf1*^{H350R/-} compared to *Klf1*^{+/-} and *Klf1*^{-/-} FL RNA; i.e. only mildly reduced (Fig. 2D). However, many genes that encode transmembrane and cytoskeletal proteins and enzymes, such as *Slc2a1*, *Slc43a1*, *Dmtn* (*Epb4.9*), *Slc1a5*, *Slc2a4*, *Fn3k*, and *Rgcc*, were expressed at very low levels in *Klf1*^{H350R/-} FL compared to *Klf1*^{+/-} FL (Fig. 2D). Thus, RNA-seq and qRT-PCR results support the genetics which shows the KLF1-H350R protein is more

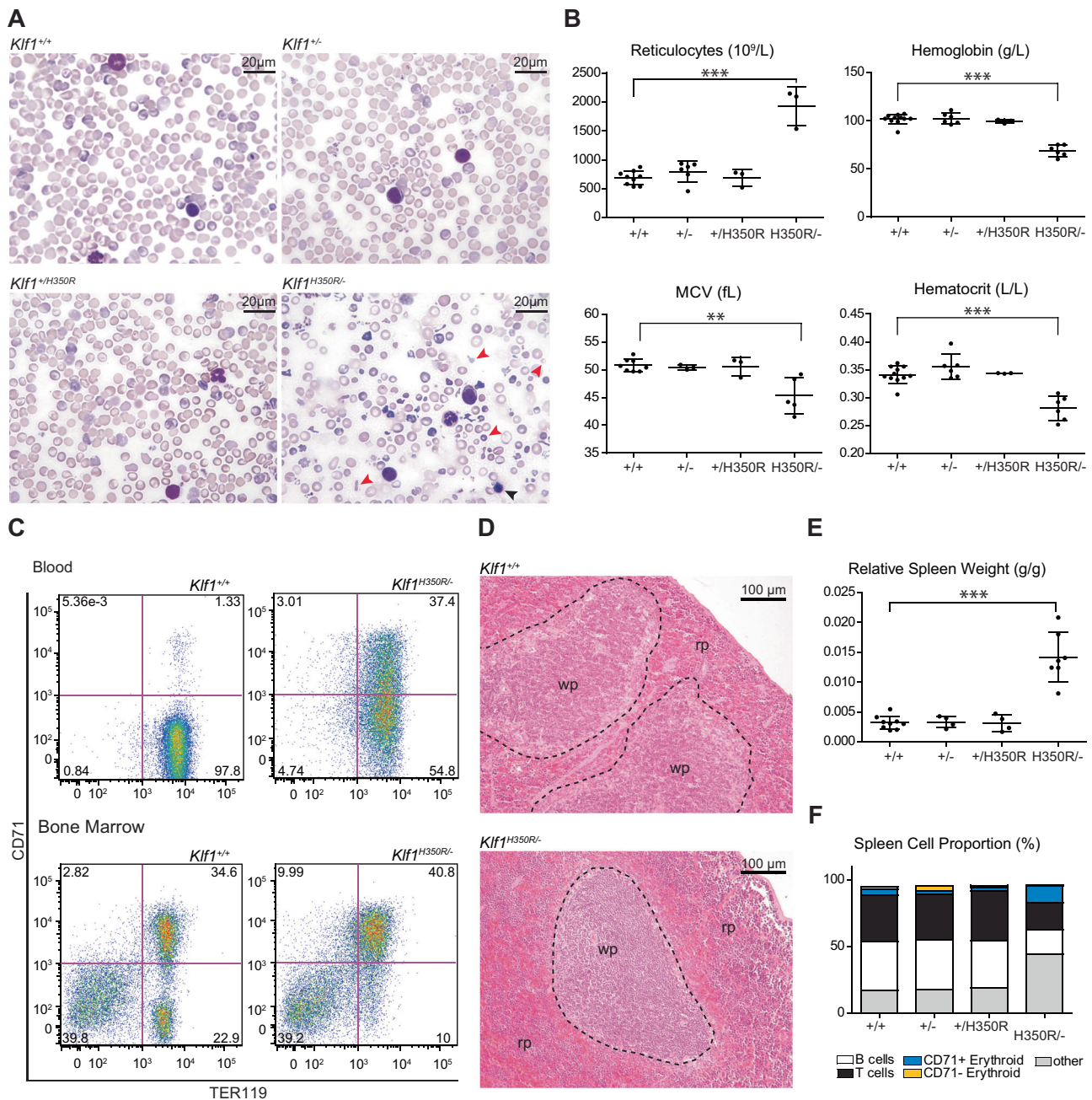


Fig. 1 | Non-spherocytic hemolytic anemia with increased reticulocytes and splenomegaly in *Klf1*^{H350R/-} mice. **A** Peripheral blood stained by May-Grünwald Giemsa (MGG) reveal hypochromic microcytic red blood cells (RBC) with anisopoikilocytosis (red arrows) and nucleated RBC (black arrow) in compound heterozygous mice (*Klf1*^{H350R/-}). **B** Compound heterozygote mice have increased reticulocytes, decreased mean corpuscular volume (MCV), decreased hemoglobin content, and decreased hematocrit levels compared to wild type. **C** Compound heterozygous animals display a significant increase in the proportion of CD71-positive cells in blood and a significant decrease of TER119-positive cells in the bone

marrow compared to normal mice. There is an expansion of red pulp (rp), but not white pulp (wp) in the spleens of compound heterozygous mice (**D**) together with an increase in relative spleen weight (**E**). **F** Spleens of compound heterozygous animals show a relative expansion of erythroid cells. In panels (**B**) and (**E**), each dot represents an independent animal ($n \geq 3$) with bars representing mean \pm SEM ($***P < 0.0005$, $**P < 0.05$). One-way ANOVA with Tukey post-hoc test was used for all statistical analysis. Images were taken on an Olympus BX50 with a DP26 camera using the cellSens Standard software. Scalebars represent either 20 μ m or 100 μ m as indicated.

defective at some target gene promoters and enhancers than others; it also explains the presence of NSHA rather than thalassemia.

Adult splenic erythroid cells from *Klf1*^{H350R/-} mice display similarly altered gene expression

Klf1^{H350R/-} mice survive into adulthood so we also undertook RNA-seq in FACS-sorted erythroid cells (CD71+, TER119+) from the spleens of adult mice (8–12 weeks of age) (See Methods). These cells are intermediate to late erythroblasts (S3 cells according to ref. 36). In this way we were

able to compare transcriptomes in phenotypically similar populations of purified erythroblasts from five different genotypes (+/+, +/H350R, +/-, H350R/H350R and H350R/-), and thereby generate gene expression profiles within an allelic series. *Klf1*^{-/-} mice die at E14.5 so we were not able to examine gene expression in adult erythroid cells of this genotype. There are few CD71+, TER119+ (S3) erythroid cells in the spleens of WT mice but a progressive increase in numbers in H350R/H350R mice and H350R/- mice consistent with the degree of splenomegaly and extramedullary erythropoiesis (Figs. 1F, 3A). The level of

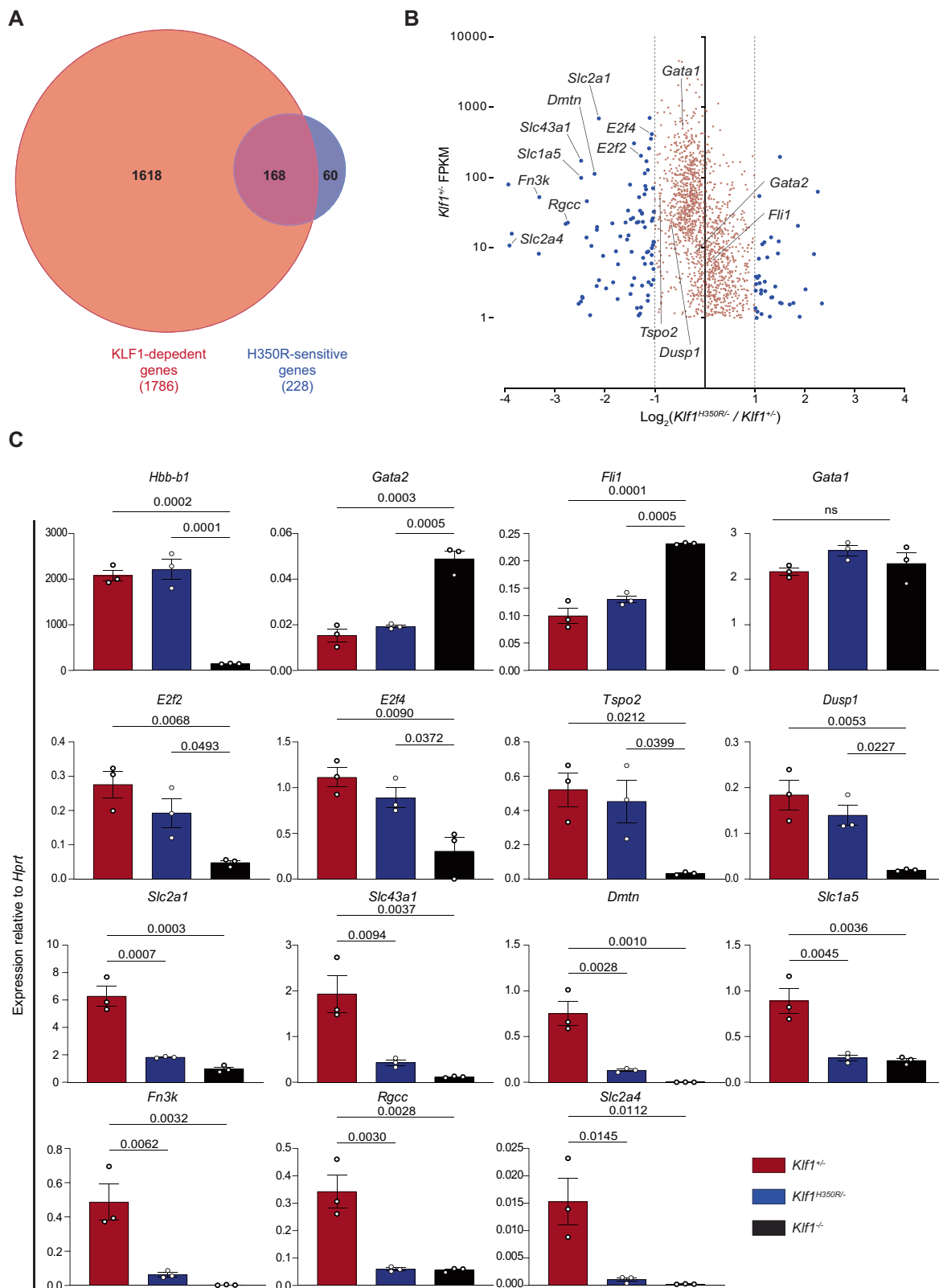


Fig. 2 | A subset of KLF1 regulated genes are sensitive to the H350R mutation. **A** Venn diagram of the 1786 DEGs in the KLF1 regulated transcriptome (*Klf1^{-/-}* vs *Klf1^{+/+}*) from E14.5 fetal liver. Of these, 168 are sensitive to the H350R mutation (*Klf1^{H350R/-}* vs *Klf1^{+/+}*). **B** MA plot of DEGs from *Klf1^{H350R/-}* embryos. Blue dots represent genes that are >2-fold down (left) or up (right) regulated in H350R/- versus +/-.

C qRT-PCR validation of DEGs identified from RNA-seq was performed with three biological replicates for each genotype and normalized to *Hprt*. The mean ± SEM is shown for each genotype and gene. One-way ANOVA with the Tukey post-hoc test was used for statistical analysis.

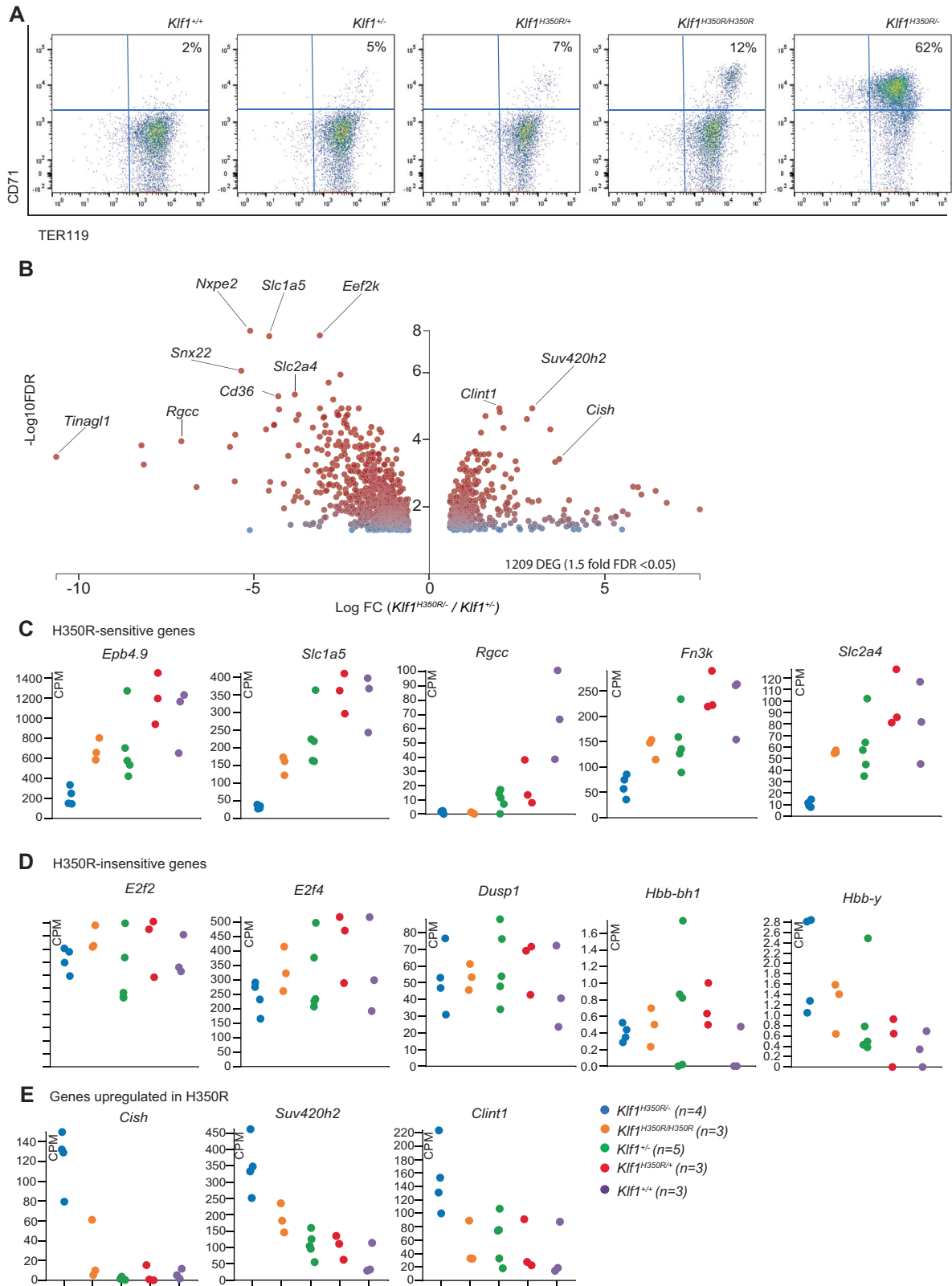


Fig. 3 | KLF1-H350R sensitive genes in splenic erythroid progenitor cells. **A** FACS plots of B220⁻, CD3⁻, CD11b⁻ cells in the spleens stained for CD71 and TER119. There is an increasing percentage of CD71⁺ + TER119⁺ (S3) pro-erythroblasts in the spleens of the allelic series of mice. **B** MDS plot of log fold change versus $-\log_{10}$ FDR in *Klf1*^{H350R/-} (*n* = 4) versus *Klf1*^{+/+} (*n* = 3) erythroid cells. Dots indicate

DEGs with >1.5 fold expression change and an FDR of <0.05. **C** Expression of KLF1-H350R-sensitive genes for five different genotypes. Each dot is a sample from a different mouse. The legend indicates genotypes and numbers. **D** Expression of KLF1-H350R-insensitive genes. **E** Expression of some EPO target genes is increased in *Klf1*^{H350R/-} mice.

surface expression of TER119 and CD71 (MFI) is slightly reduced in *Klf1*^{H350R/-} splenic erythroid cells compared with other genotypes (Figs. 1C, 3A); this contrasts with *Klf1*^{-/-} FL erythroid cells, which have markedly reduced expression of TER119 and moderately reduced expression of CD71³⁴.

We found 419 DEGs in erythroblasts from *Klf1*^{H350R/H350R} mice compared with WT mice (1.5 fold; FDR < 0.05) (Supplementary Fig. 4), and we found 1209 DEGs in *Klf1*^{H350R/-} mice compared with WT mice (Fig. 3B). In both comparisons, ~65% of genes were downregulated in the mutant mice and ~35% were upregulated. The latter were enriched for genes encoding proteins involved in EPOR signaling (e.g. *Clint1*, *Suw420h2*, *Podxl*, *Thbs1*, *Cish*, etc.)³⁷ or stress erythropoiesis (Fig. 3E); none of the top 25 upregulated genes had a KLF1 ChIP-seq peak nearby. Thus, we conclude these are upregulated indirectly in response to anemic stress. For the down-regulated genes, there was very strong overlap with the H350R-sensitive genes in the FL. For example, there was marked down regulation of *Epb4.9* (dematin) *Slc1a5* (Asct2), *Rgcc*, *Fn3k*, and *Slc2a4* (Glut4), and in *Klf1*^{H350R/-} erythroid cells from spleen and FL (Figs. 2, 3C). *Slc2a1* (Glut1) is not expressed in adult erythroid cells³⁸, so it was only found as a DEG in FL. In many cases, there was a clear gradient of gene expression level with respect to the allelic series/genotypes; i.e. H350R/- cells have the greatest perturbation of gene expression, followed by H350R/H350R cells, +/- cells, and H350R/+ cells in comparison with +/- cells (Fig. 3). Interestingly, expression of some well-known KLF1-target genes such as *Hbb-b1* (β-globin), *E2f2* and *E2f4*²⁶ was minimally perturbed in spleen erythroid cells from *Klf1*^{H350R/-} mice, as in the FL (Figs. 3D, 2). Embryonic and fetal globin genes (*Hbb-bh1*, *Hbb-y*) were expressed at very low levels in adult spleen and not significantly altered in *Klf1*^{H350R/-} mice (Fig. 3D); i.e. there was no evidence for upregulation of fetal and embryonic globin genes in *Klf1*^{H350R/-} mice. This contrasts with studies of mice with dominant NSHA from a missense mutation in zinc finger two that contacts DNA (the *nan* mouse)^{39,40}, and with elevated human γ-globin gene expression in *Klf1*^{-/-} mice that harbor a human β-globin locus YAC⁴¹ and KLF1^{+/-} humans⁴². In short, RNA-seq in adult splenic erythroid progenitor cells (S3 cells) detected a similar set of KLF1-H350R-sensitive genes to those found in the FL. We suggest this combination of expression changes underpins the observed NSHA phenotype rather than thalassemia (Fig. 1).

KLF1-H350R has reduced affinity but unaltered specificity for DNA in vivo

In order to determine the in vivo binding characteristics of the KLF1-H350R mutation, we performed ChIP-seq for KLF1 in *Klf1*^{+/+} versus *Klf1*^{H350R/H350R} FLs. We identified a total of 1144 KLF1-occupied sites in wild type FL using MACS2 (see Methods). This reflects improved technical expertise compared with previously published data¹⁶. High quality ChIP-seq data for murine KLF1 in primary cells has proven to be difficult to obtain, so this new data set provides a valuable new resource for the field (GSE94351). We found only 490 KLF1-occupied sites passed MACS2 peak calling in homozygous *Klf1*^{H350R/H350R} FL, most likely due to globally reduced DNA-binding affinity (see below). An overlap of 341 sites was observed between the two datasets based on a window of 100 bp either side of the peak summits (Fig. 4A).

An identical DNA consensus motif was preferred by both the wild type and H350R-mutant KLF1 proteins in vivo according to de novo motif discovery (Fig. 4B), and from counting of specific motifs closest to the center of each peak, as described²⁷ (Fig. 4C). Together, this shows unaltered DNA-binding specificity of KLF1-H350R in vivo and contrasts with ChIP-seq data for KLF1 mutations in DNA-contacting residues, such as those which cause dominant hemolytic anemia in the *Nan* mouse^{27,29} and CDA-IV in humans³⁰. The motif is consistent with the extended CACCC-box motif (CCM-CRC-CCN on the C-rich strand; or, NGG-GYG-KGG on the G-rich strand) bound in vivo by all SP/KLF family members studied to date^{4,16,43-47}. To further compare the in vivo

DNA-binding ability of wild-type KLF1 versus KLF1-H350R, we counted reads under each peak (+/-100bp from peak center). We found counts over most shared peaks were consistently greater for wild type KLF1 than for KLF1-H350R (Fig. 4D). The majority of H350R-sensitive genes as determined by RNA-seq have a KLF1 ChIP-seq peak at the promoter, intron or upstream of the TSS, so most are likely direct targets of KLF1 (Table 1). In short, the motifs within the cis-regulatory modules of KLF1-H350R-sensitive genes are indistinguishable from those in less sensitive genes globally (Fig. 4C). For example, dematin (*Epb4.9*) encodes a key linker protein between the cytoskeleton and the actin junctional transmembrane complex that includes GLUT1 and stomatin^{48,49}. There are three independent *Epb4.9* gene promoters³⁴. The second promoter and a putative intron enhancer in intron 2 are robustly bound by KLF1 in the FL (Fig. 5A). This binding is reduced in KLF1 H350R/H350R cells and mRNA is markedly downregulated. A previous report shows *Glut1* is regulated by SP3 in muscle cells via a proximal promoter CACC box element⁵⁰, but in FL there is a strong binding site at -1 kb upstream of the promoter (Fig. 5B). Binding and expression are markedly reduced in *Klf1*^{H350R/-} FLs. The same applies for promoters and putative intronic enhancers for *Slc2a4* (Glut4) and *Slc1a5* (Asct2) (Fig. 5 C-D), and many other KLF1 H350R-sensitive genes.

Recombinant linker 2-mutant KLF1 zinc finger domains have reduced DNA-binding affinity in vitro

It can be difficult to determine relative in vivo binding specificities of mutant transcription factors (TFs) from ChIP-seq data alone, due to variables which are difficult to control⁵¹. We therefore produced and purified GST-zinc finger domain fusion proteins in *E.coli* as previously described²⁷ (see Methods) for in vitro binding assays (Supplementary Fig. 5). The KLF1 binding sequence (CCM-CRC-CCN) is categorized by the presence of guanine (type 1) or an adenine (type 2) at the central position (underlined). The importance of a cytosine versus adenine at position 3 (M) for overall affinity has not been studied in detail. Using electrophoretic mobility shift assays we find both GST-KLF1-zf and GST-H350R-KLF1-zf bind to canonical type 1 motifs such as those found in the *Alas2* intronic enhancer and putative promoter and intronic enhancer of *Slc2a4* (Fig. 6A) and type 2 motifs, such as that which is present in the *E2f2* intronic enhancer²⁶ and the *Slc1a5* and *Slc43a1* putative intronic enhancers (Fig. 6B). However, the affinity of GST-H350R-KLF1zf for the Type 1 probe, CCC-CGC-CCT, and type 2 probe, CCC-CAC-CCT^{27,40}, are both ~4-fold reduced (Fig. 6). Similar results were found for the *Hbb* promoter (Supplementary Fig. 5B). This reduced in vitro binding affinity is consistent with the ChIP-seq data. It shows insertion of an arginine at position +3 in linker 2 of KLF1 reduces overall affinity of the zinc finger domain for all variants of the recognition motif. We made two additional mutations in linker 2, T348A and G349R, which are equivalent to human disease-associated mutations^{20-22,52}. These also result in ~4-fold reduced binding affinity for both type 1 and type 2 KLF1-binding sites in vitro (Fig. 6).

To seek an explanation for reduced binding affinity, we revisited the structure of KLF4 bound to DNA⁵³, and other solved zinc finger structures^{8,10,54}. In the KLF4 structure, threonine at +1 in linker 2 is part of the extended α-helix far from the DNA substrate (Supplementary Fig. 6, yellow box). The glycine at position +2 plays an important role in α-helix termination or C-capping. Similarly, for TFIIIA the glycine at position +2 in both linkers 1 and 2 makes an *i, i + 4* hydrogen bond with a carbonyl group on serine or leucine in the preceding α-helix (red line, Supplementary Fig. 6). In Zif268 and TFIIIA, the +4 lysine makes contact with the DNA backbone and this could also increase binding affinity¹⁰. It is difficult to speculate about the consequences of an arginine substitution at +3 in linker 2 (or G>R at +2 in the case of human patients), although it is worth noting that arginine does have a slightly stronger preference to adopt alpha helical secondary structure than either histidine or glycine⁵⁵. One explanation for the observed phenotype may therefore be that these mutations could modify the length

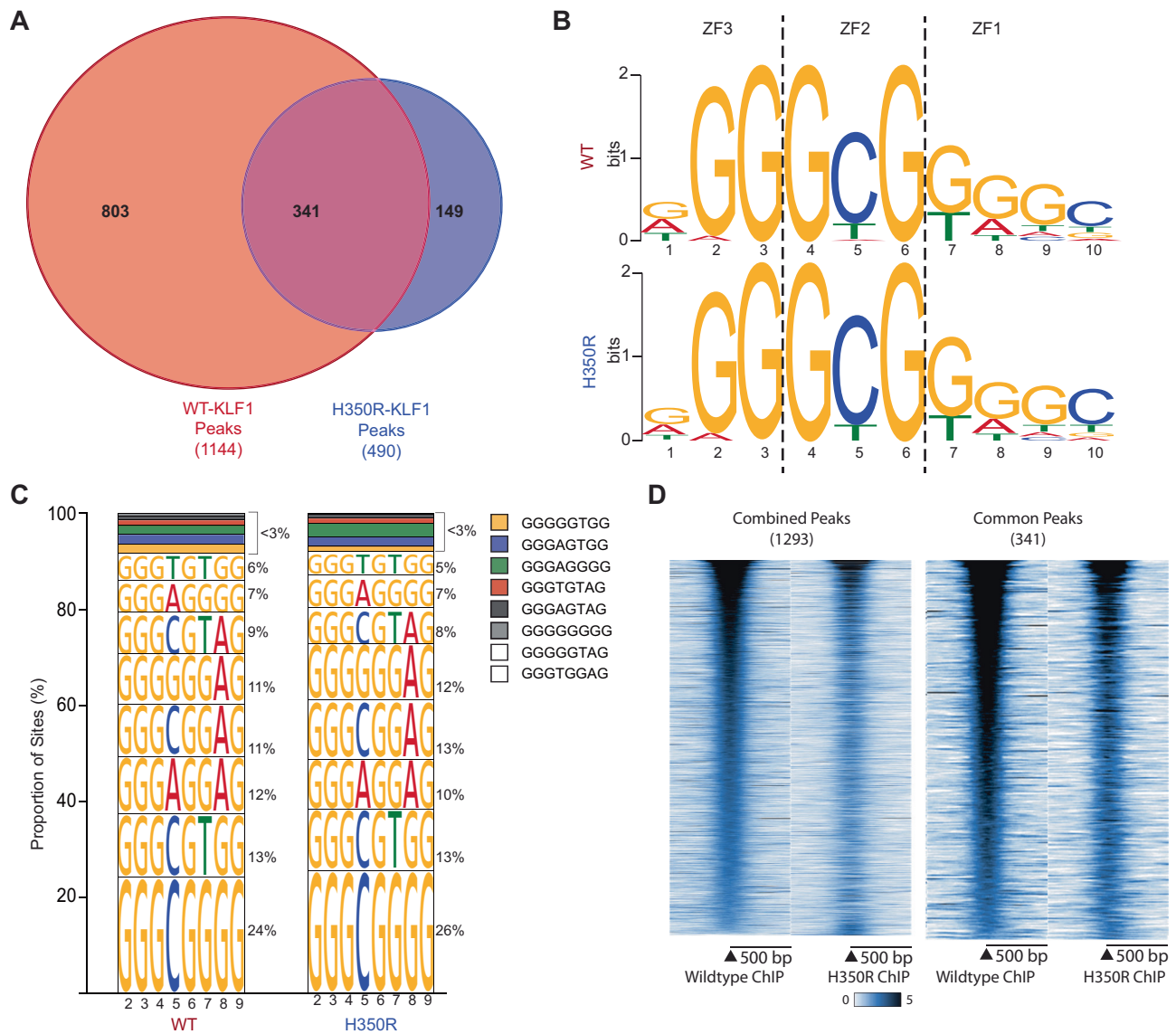


Fig. 4 | The H350R mutation does not alter DNA-sequence binding specificity in vivo. **A** 1144 KLF1-occupied sites and 490 KLF1-H350R-occupied sites were called by MACS2 analysis. An overlap of 372 peaks (common peaks) was identified using a 100 bp window between peaks summits. **B** De novo motif discovery found enrichment of the classical 9 bp CACCC motif (C-rich strand) in both WT and H350R datasets using 100 bp surrounding the peak summit. **C** Stacked histogram

illustrating the proportion of central motif sequence variants. Central motifs (CACCC motif to the nearest peak summit) were identified in 939 of the wild type peaks and 415 of the H350R peaks. Peaks with more than 1 identified binding site were omitted for this analysis. **D** Read density heat maps showing enrichment by ChIP surrounding the peak summit in both WT and H350R. Read density heat maps of common peaks display the shared regions (summits within 100 bp) only.

of the α -helix of finger 2, in both cases leading to reduction in DNA binding affinity. Alternatively, the close positioning of arginine side-chains at both linker positions +2/3 and +4 could result in repulsive forces between the sidechains that destabilize the important interaction between the +4 arginine and the DNA backbone.

Supplementation with L-leucine results in partial rescue of NSHA

One of the KLF1-H350R sensitive target genes is *Slc43a1* or LAT3. LAT3-deficient zebrafish have severe anemia that can be rescued by L-leucine supplementation in the water⁵⁶. Thus, we asked whether we could rescue the NSHA in *Klf1*^{H350R/-} mice by supplementation of the drinking water with L-leucine (1.5%). Indeed, there was a significant increase in Hb of -40 g/L (corrected p value < 0.005) and red blood cell count (RCC) of $-1.5 \times 10^9/L$ (corrected p value < 0.01) in *Klf1*^{H350R/-} mice, which was maximal after four weeks of supplementation, and was partially maintained at 6 weeks (18 g/L, p value < 0.01 for Hb), but mostly lost by

8 weeks. There was no change in Hb or RCC in *Klf1*^{+/-} mice (Fig. 7) (see Discussion).

Discussion

Mutations in *KLF1* are common in certain ethnic groups^{57,58}. Homozygous loss-of-function (LoF) mutations result in *hydrops fetalis*, which is fatal without intrauterine or perinatal blood transfusion support¹³, whereas heterozygous LoF mutations result in a mild phenotype which resembles β -thalassemia trait with mildly raised HbF and HbA2, raised ZPP plus blood serological abnormalities^{23,59}. Missense mutations in linker 2 result in very mild or absent phenotypes, but when co-inherited with loss of function mutations on the other allele, these lead to severe NSHA²¹. Lastly, dominant mutations at p.Glu325 in zinc finger 2 cause CDA type 4⁶⁰. This phenotypic variability is not fully understood and leads to diagnostic challenges.

We generated a murine model of NSHA by crossing loss of function¹² and linker 2 mutations in *Klf1*²⁴. The mice have severe

Table 1 | KLF Linker 2 (H350R)-sensitive genes (DEGs from mRNA-seq in fetal liver)

Symbol	H350R (FPKM)	WT (FPKM)	Fold change (H350R/WT)	ChIP Peak *	Function	Reference
Slc4a5	0.049	0.984	0.050	Yes (P, E, I)	Na ⁺ /HCO ³⁻ Co-transporter	
Nxpe2	5.187	79.20	0.065	No	Fam55b	
Slc2a4 (Glut4)	0.713	10.71	0.066	Yes (E, P, I)	Glut4; glucose transport	
Phospho1	1.083	15.77	0.068	Yes ^a	Part of ZFP652 unit	
Paqr6	0.820	8.203	0.099	Yes (P)	Progesterin receptor	
Fn3k	5.268	52.35	0.100	Yes (P)	Hb glycation	
Abi3	3.161	21.79	0.145	Yes ^a	Part of ZFP652 unit	
Rgcc	3.391	22.67	0.149	Yes (P, E)	enhances Cdk1 activity	
Phyhip	0.276	1.589	0.173	Yes (I)	Complex locus	
Nacad	0.306	1.711	0.179	Yes (I)		
Slc43a1 (Lat3)	30.99	172.4	0.179	Yes (P, I)	Leucine transport	
Slc1a5 (Asct2)	17.72	98.49	0.179	Yes (P, I)	Glutamine transport	
Il22	0.364	1.997	0.182	No	Cytokine	
lqcd	0.353	1.927	0.183	Yes ^b		
Sowaha (Ankrd43)	2.700	13.93	0.193	Yes (P, I)		
Fam53b (Simplet)	8.921	45.83	0.194	Yes (P, I)	Wnt signaling	
Apol11b	0.185	0.931	0.199	No		
Epb4.9	24.28	112.1	0.216	Yes (P, E, I)	Cytoskeleton; erythropoiesis	61
Wdr65	0.634	2.841	0.223	Yes (I)		
Mgll	4.005	17.91	0.223	Yes (I)		
Darc	4.417	19.69	0.224	Yes (P, I)	Duffy Antigen	
Slc2a1 (Glut1)	158.7	690.4	0.229	Yes (P, SE)	glucose transport	62
Sh3tc2	0.793	3.440	0.230	Yes (P)	Charcot-Marie-Tooth type 4c	
Kctd14	1.863	7.641	0.243	Yes (I)	K ⁺ channel	
Ano1 (Tmem16A)	0.898	3.208	0.280	No	Chloride channel	
B3gnt3	2.587	8.880	0.291	Yes (I)	Acetylglucosaminyltransferase	
Tac2	4.492	14.49	0.310	Yes (I)		
Ces2g	7.595	22.80	0.333	Yes (I)	Carboxylesterase	
Dnaja4	11.64	33.76	0.344	Yes (E)	Chaperone	
Marco	5.187	14.94	0.347	No	Macrophage receptor	
Ehbp1l1	27.95	78.83	0.354	Yes (P, I) ^c	endocytosis	
Unc5cl	0.781	2.190	0.356	No	IRAK activator	
Agpat4	9.097	25.39	0.358	Yes (I)	Acyltransferase	
Ctsf	3.158	8.719	0.362	Yes (P, I)	Cathepsin F	
Lingo3 (Lern3)	1.065	2.922	0.364	No		
Cast	9.802	26.50	0.369	Yes (P, I)	tenderization of meat	
Ubac1	114.3	305.2	0.374	Yes (P, I)	Ubiquitin ligase	
Gm12839	8.840	22.71	0.389	Yes (P,E)	Cyp4b1-ps2	
Hs6st1	9.258	23.24	0.398	Yes (I)	Heparan Sulfate Sulfotransferase	
Cntn3	0.433	1.085	0.399	No	Contactin 3	
Tmc6 (Ever1)	20.96	52.16	0.401	Yes (P) ^d	Barrier to HPV	
Tmc8 (Ever2)	13.67	33.57	0.407	Yes (P) ^d	Barrier to HPV	
Cadm3	0.469	1.152	0.407	Yes (E, I)	PI3K/ACK signaling	
E2f2	83.78	203.2	0.412	Yes (I)	DNA replication, KLF1 target	26
Eef2k	13.01	31.12	0.418	Yes, (P, E, I)	Threonine kinase	
Speg	12.71	30.37	0.418	None	Kinase	
Slc26a1	9.848	23.25	0.423	Yes (I)	Sulfate transporter	
M1ap	3.038	7.151	0.424	Yes (I)	Meiosis	
Rnf11	24.87	57.46	0.432	No	Ring finger involved in signaling	
Mpst	49.52	113.9	0.434	Yes, (P, E, I)	Anti-oxidant	
Artn	2.554	5.862	0.435	Yes (I)		
Fn3krp	8.508	19.53	0.435	Yes (P, E)		
Dck	33.11	75.59	0.438	Yes (I)	Deoxycytidine kinase	
Ttl12	29.84	67.47	0.442	Yes (P)		
Kcnn4	75.47	169.5	0.445	Yes (I)	Gardos channel; xerocytosis	71

Table 1 (continued) | KLF Linker 2 (H350R)-sensitive genes (DEGs from mRNA-seq in fetal liver)

Symbol	H350R (FPKM)	WT (FPKM)	Fold change (H350R/WT)	ChIP Peak *	Function	Reference
Ptp4a3 (Prl-3)	57.75	127.3	0.453	Yes (I)	Tyrosine phosphatase	
1700001L05Rik	0.950	2.090	0.454	Yes (P) ^e		
Lck	4.022	8.815	0.456	No	Src-type tyrosine kinase	
Nhlrc4	1.476	3.223	0.458	Yes (E) ^f		
Icam4	119.4	258.7	0.461	Yes (P)	Cell adhesion and blood island	72,73
Tmcc2	326.1	700.8	0.465	No		
Pklr	167.5	353.5	0.473	Yes (P, I)	Pyruvate Kinase, red blood cell Fx	20
Gdf3	2.843	5.957	0.477	No		
Nlrp1a	3.754	7.840	0.478	No		
Krt15	4.722	9.853	0.479	No		
E2f4	198.0	413.2	0.479	Yes (I)	Cell cycle; erythropoiesis	26
Slc22a4 (Octn1)	11.53	24.05	0.479	Yes (P, I)	Cation transporter	
Relt	12.68	26.10	0.486	No	TNF receptor	
Klf3	34.16	70.03	0.487	Yes (P, I)	Dampens KLF1 responses	74,75
Homer2	0.460	0.940	0.490	No		
Myo1d	2.434	4.965	0.490	No		
Hcn3	1.712	3.490	0.490	Yes (I)		
Lrrc56	5.952	12.11	0.491	No		
Bcl11a	15.30	30.62	0.499	No	Globin gene switching	
Xpo7	24.02	47.94	0.501	Distant	RBC enucleation	76
Ccdc71l	9.303	18.49	0.502	Yes (P)		
Cdkn2d (p19-INK4D)	16.78	33.36	0.503	No	Cell cycle inhibitor	
Pabpc4	195.3	387.3	0.504	Yes (P, I)		
Rcbtb2	10.41	20.61	0.505	No		
Mgst3	430.2	829.4	0.518	Yes (P, I)	Glutathione S-Transferase	
Lgals1	84.66	162.7	0.520	Yes (E)	Galectin -1	
Gm5088	2.729	5.225	0.522	No		
Acot11	6.565	12.56	0.522	No		
Tagln3	9.766	18.50	0.527	No		
Zfp341	2.874	5.396	0.532	Yes (P) ^g		
Adra2b	14.59	27.31	0.534	No		
Arhgdig	31.82	59.52	0.534	No		
Lmna	115.3	215.6	0.534	Yes (I)		
Sun1	14.20	26.46	0.536	Yes (P, E)	Nuclear envelope	
Tspo2	29.15	54.29	0.536	Yes (P, E, I)	Cholesterol binding	77
Cercam	41.74	77.71	0.537	Yes (P, I)	Cell adhesion	
Clk3	52.57	97.86	0.537	Yes (P, E)	Cdc-like kinase	
Dclk2	2.449	4.530	0.540	No	Doublecortin like kinase	
Sppl2b	31.94	58.82	0.543	Yes (P, I)	Pepidase	
5830417110Rik	4.237	7.794	0.543	No	Alternative transcript: Gon4-like	
Slc25a37	301.9	553.6	0.545	Yes (E)	Mitoferrin	
Tnfrsf21	13.12	23.95	0.548	Yes (I)	TNF receptor	

*Presence of KLF1 ChIP-seq peak (Yes or No) at promoter (P), intergenic enhancer (E), super enhancer (SE) or intronic region (I).

^aChIP-peak within intron ZFP652 gene which is part of Phospho1/Abi3 transcriptional unit.

^bChIP-seq peak in upstream gene.

^cChIP-seq peak also in neighboring genes, Fam89b and Kcnk7.

^dLinked genes.

^eBinds promoter of neighbor gene, Arfgap3.

^fBinds upstream of downstream gene, Pigq.

^gBinds promoter of Pxmp4.

fragmentary hemolysis rather than β -thalassemia, just like human patients with NSHA and KLF1 linker mutations^{22,23}. Using RNA-seq we show this is due to selective loss of a small subset of KLF1-dependent genes. Many of these encode transmembrane and cytoskeletal genes, which is consistent with the NSHA phenotype. One of the most down regulated genes is dematin (*Epb4.9*), which is a critical regulator of the cytoskeleton. Dematin knockout mice have severe fragmentary

hemolysis and red blood cell dysmorphology similar to *Klf1*^{H350R/-} mice⁶¹. We also found >10-fold downregulation of *Scl2a1* (*Glut1*), *Slc2a4* (*Glut4*), and *Slc1a5* (*ASCT2*) in the fetal liver (Fig. 2), and *Slc2a4* and *Slc1a5* in the spleen (Fig. 3). There is a developmental switch in expression of *Scl2a1* to *Slc2a4* between fetal and adult red blood cells that is not altered by the *Klf1* mutations³⁸. *Glut1* deficiency results in hemolytic anemia⁶², but the *Glut4* knockout mouse does not have a

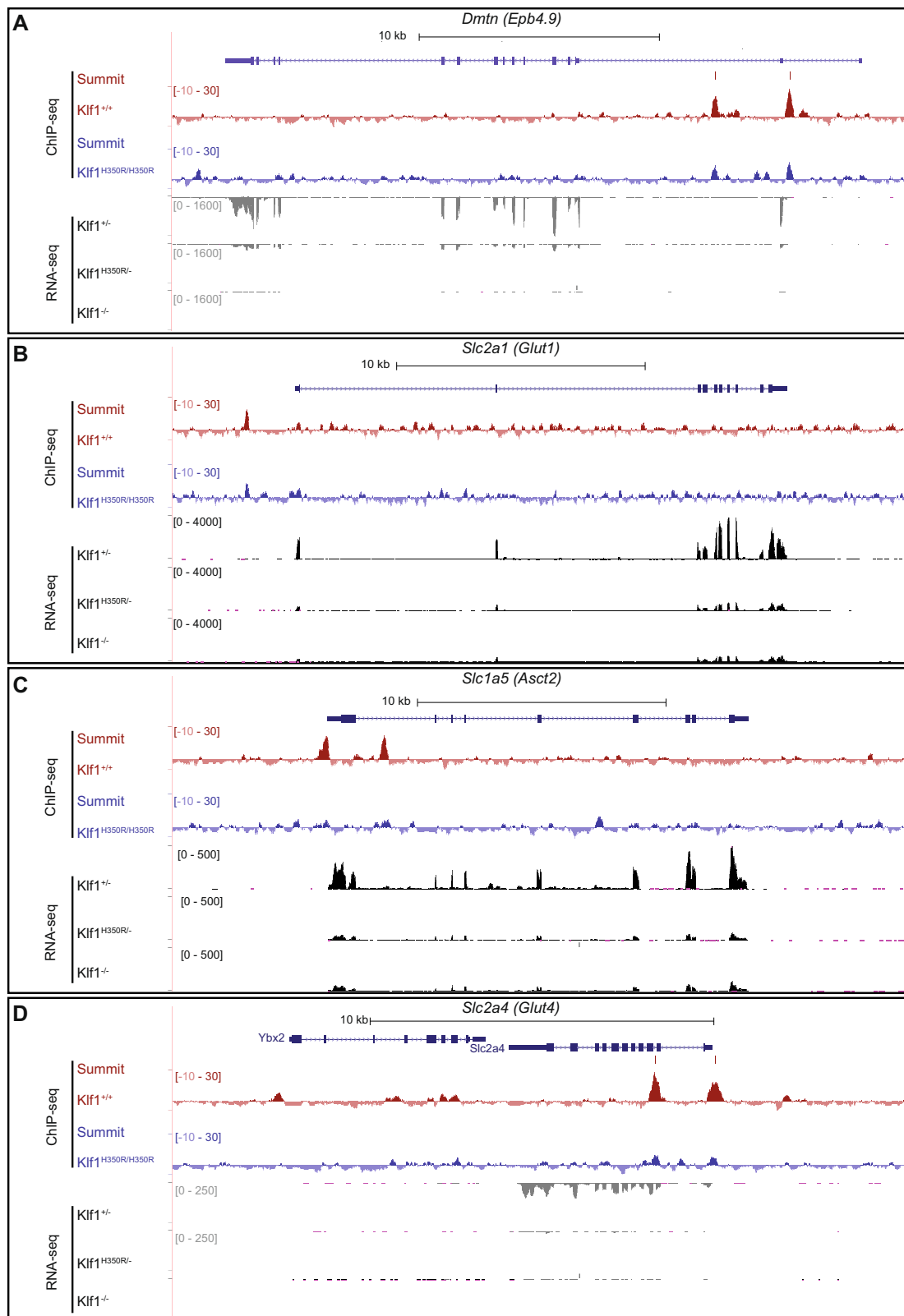


Fig. 5 | *Dmtn* (Epb4.9), *Slc2a1* (Glut1), *Slc1a5* (Asct2) and *Slc2a4* (Glut4) are all direct targets of KLF1 and sensitive to the KLF1-H350R mutation. UCSC Genomic Browser views of wiggle tracks of KLF1 ChIP-seq (red and blue) and mRNA-seq from the fetal liver (black and grey). ChIP-seq peak summits called by MACS2 are indicated with solid bars above the wiggle tracks of ChIP signal, which was generated by subtracting ChIP signal to background input DNA signal (*Klf1*^{+/+}: red, *Klf1*^{H350R/H350R}: blue). The mRNA-seq from the three fetal liver genotypes, *Klf1*^{+/+},

Klf1^{H350R/H350R} and *Klf1*^{-/-}, are shown below, forward strand reads are coloured black and reverse strand reads in grey. KLF1 binds **A** the second promoter and putative enhancer for *Dmtn*, **B** a putative promoter for *Slc2a1*, **C** the promoter and putative intronic enhancer for *Slc1a5*, **D** the promoter and putative intronic enhancer for *Slc2a4*. There is reduced ChIP-seq signal in *Klf1*^{H350R/H350R} mice, and reduced levels of RNA expression in *Klf1*^{H350R/H350R}.

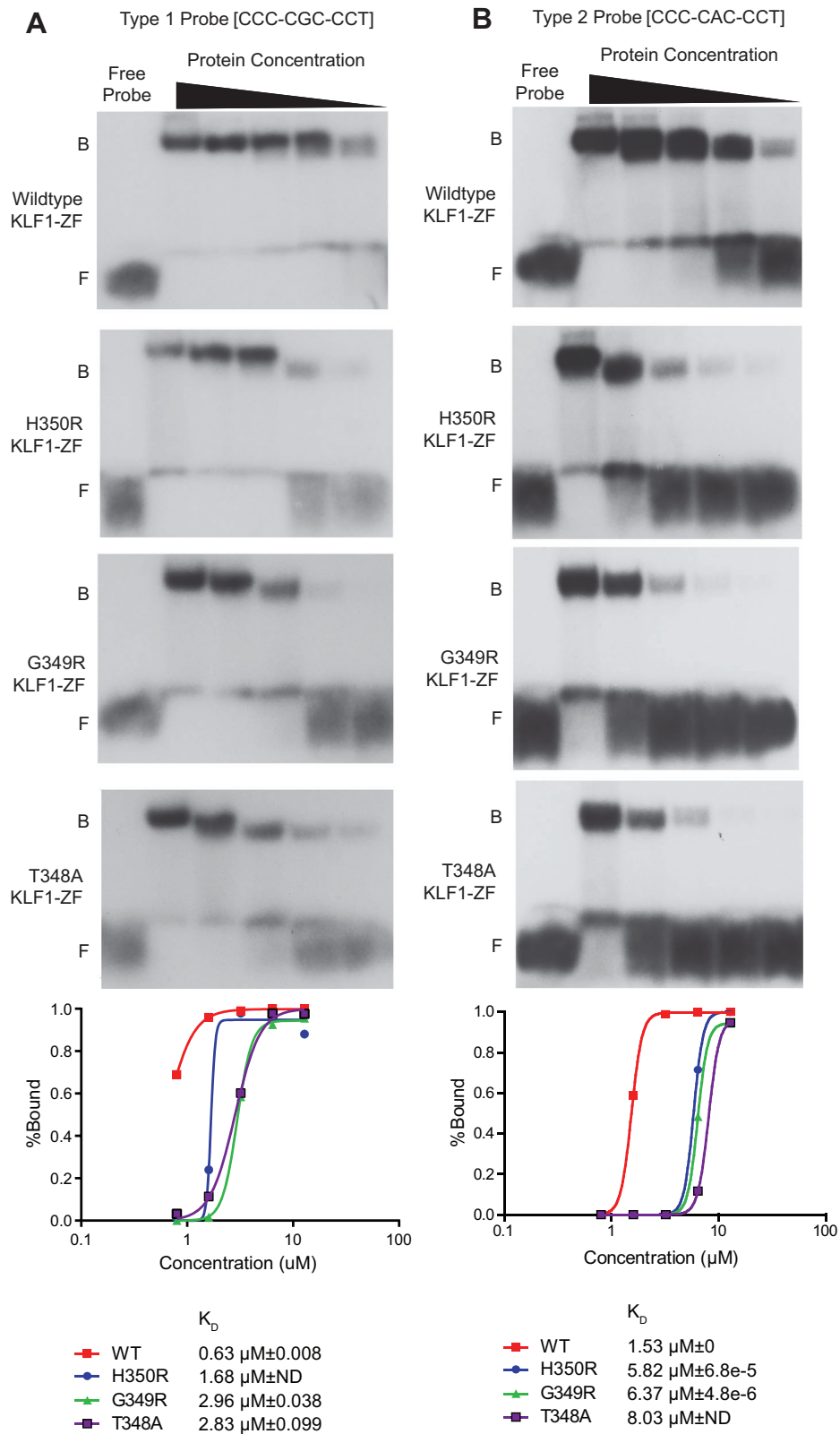


Fig. 6 | Mutations in the second linker of KLF1 cause reduced DNA-binding affinity. **A** Electrophoretic mobility shift assay comparing the binding affinity of three KLF1 linker-mutations (H350R, G349R, T348A) against a known KLF1 binding site in the first intron of *E2f2*. Protein concentrations were first equalized, and then

serially loaded using a 1:1 dilution with binding buffer, for each lane from left to right. The probe concentration was consistent across all 4 panels. **B** K_D values are calculated from quantification of each blot shown (**C**).

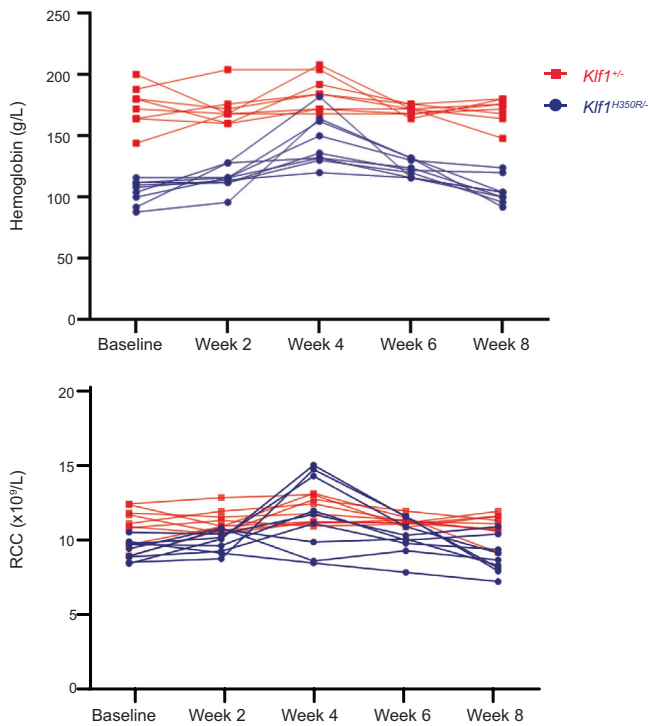


Fig. 7 | Partial transient rescue of anemia in $Klf1^{H350R/-}$ mice with L-leucine supplementation. Hemoglobin (Hb) levels and RBC counts (RCC) in mice at baseline and every second week following supplementation with L-leucine (1.5% weight/vol) in the drinking water for $Klf1^{+/+}$ (red; $n=8$) and $Klf1^{H350R/-}$ (blue; $n=9$) mice. There is no significant change in Hb levels or RCC in $Klf1^{+/+}$ mice, but a significant and variable upregulation of Hb and RCC in $Klf1^{H350R/-}$ mice at 4 weeks (** $p < 0.005$) and at 6 weeks (* $p < 0.01$) after L-leucine supplementation. Two-way ANOVA with a Tukey post-hoc test was used for statistical analysis.

major erythroid phenotype. ASCT2 is a glutamine transporter which is important for energy production and generation of erythroid cells from HSCs⁶³. GLUT1/4 and ASCT2 play critical roles in the uptake of glucose and glutamine, respectively. Glutamine is harnessed for protein biosynthesis, but is also a substrate for nucleotide biosynthesis via the TCA cycle⁶³ (Fig. 8).

Interestingly, exogenous delivery of nucleotides can rescue human erythroid differentiation from CD34 cells in which nucleotide synthesis from glutamine is blocked with 6-diazo-5-oxo-L-norleucine (DON)⁶³. So, reduced expression of both transporters in $Klf1^{H350R/-}$ mice is likely to result in a shortage of nucleotide production and contribute substantially to anemia (Fig. 8). This raises the possibility that exogenous nucleotides might be able to partially rescue the anemia in patients with linker mutations in *KLF1* if they could be delivered efficiently to erythroid progenitors in the bone marrow. Nucleotides are often added to infant formula preparations and they can be purchased as supplements in the belief they may enhance the health of the gut and immune system. Nucleotides are degraded to nucleosides by phosphatases in the gut and efficiently taken up by intestinal epithelial cells, but it is not clear if much enters the circulation to short cut purine and pyrimidine biosynthetic pathways. Hematopoietic stem and progenitor cells certainly express members of the ENT and CNT nucleoside transport families, so uptake from the diet is possible. Although this is not likely to be of major importance for normal nucleoside biosynthesis, it might provide an alternative source of nucleosides for *KLF1* mutant erythroid progenitors which are likely to be starved of nucleotides because of deficiency in ASCT2 and GLUT1 transporters. Also, supplementation of the diet with L-glutamine might benefit patients with NSHA due to *KLF1* mutations just as it does in patients with sickle cell disease⁶⁴. In this case the L-glutamine would need to enter erythroid progenitors despite low levels of expression of the receptor.

Slc43a1 (LAT3) is also markedly downregulated in $Klf1^{H350R/-}$ FL (Fig. 2). LAT3 is a neutral amino acid transporter with particular avidity for L-leucine, which is an essential building block for hemoglobin

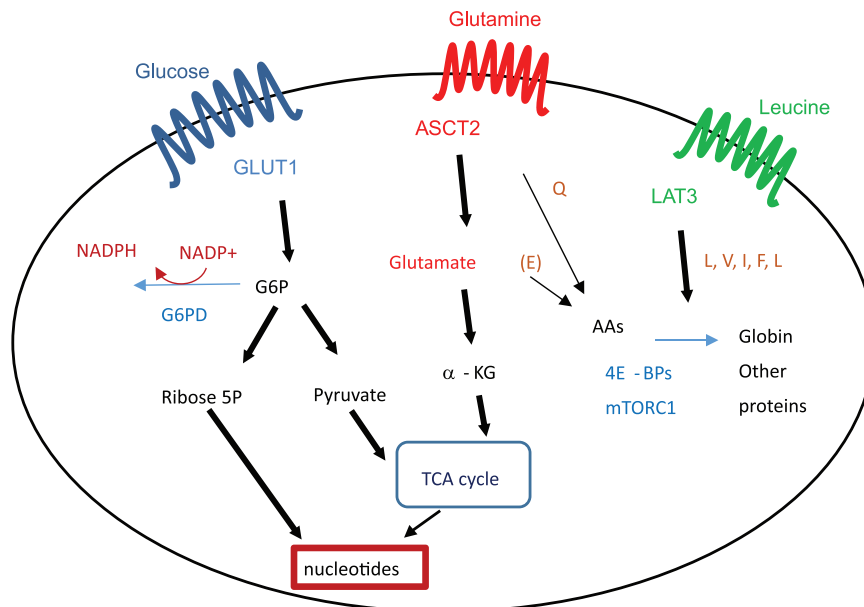


Fig. 8 | Model of metabolic defects in erythrocytes of $Klf1^{H350R/-}$ mice. The expression of three membrane transporters, *Slc2a1* (*Glut1*), *Slc1a5* (*Asct2*) and *Slc43a1* (*Lat3*) were significantly downregulated in $Klf1^{H350R/-}$ fetal livers when compared to $Klf1^{+/+}$ litter mates. These three transporters are important in the maturing erythrocyte and are involved in metabolic processes such as glucose

(GLUT1) and glutamine (ASCT2) uptake for nucleotide biosynthesis, as well as leucine and other neutral amino acid transport (LAT3) for hemoglobin synthesis. G6P is generated from glucose (imported by GLUT1) and is used for generation of NADPH from NADP⁺ by G6PD. This is critical for protection against oxidant stress.

synthesis in maturing marrow erythroblasts⁵⁶. So, expression at <10% of endogenous levels is also likely to contribute to the anemia in *Klf1*^{H350R/-} mice (Fig. 8). Interestingly, supplementation of L-leucine in the drinking water lead to a significant increase in hemoglobin and RCC in *Klf1*^{H350R/-} mice (Fig. 7), just as it does in LAT3 deficient zebrafish⁵⁶. The timing of improvement is consistent with improved translation of globins and other important genes in erythroid progenitor cells, but it is not clear why this benefit is not maintained. This suggests there could be an additional secondary translational problem in *Klf1*^{H350R/-} erythroid progenitor cells beyond transcriptional defects. There are other transmembrane transporters which are sensitive to the H350R mutation, so the NSHA phenotype is likely due to a combination of loss of many or all of these targets.

We employed ChIP-seq in primary FL cells to determine whether the DEGs are direct target genes of KLF1 and whether the H350R mutation results in altered DNA binding affinity or specificity. We found strong evidence for direct activation of most of the DEGs by KLF1. In particular, it binds the promoters and enhancers of *Slc5a1* (ASCT2), *Slc2a1* (GLUT1), *Slc2a4* (GLUT4), *Dmtn* (EBP4.9) and many other genes (Fig. 5), so these genes are direct KLF1 target genes (Table 1). We found a globally reduced ChIP signal in *Klf1*^{H350R/H350R} erythroid cells compared to wild type cells (Fig. 4) and confirmed ~4-fold reduced affinity at all KLF1-bound sequences using recombinant purified zinc finger domains and EMSA assays (Fig. 6). This is consistent with previous work which shows similar loss of DNA-binding affinity in vitro as a result of linker mutations in other transcription factors^{9,65}. Our work herein is the first to show this is also true in vivo. The DNA-binding specificity of H350R was unchanged in vivo and in vitro (Figs. 4, 6), as one might predict for a mutation far removed from the canonical DNA-binding domain (Supplementary Fig. 6). This contrasts with the change in DNA-binding specificity that accompanies mutations in the DNA-contacting amino acids of zinc finger 2; these cause dominant congenital dyserythropoietic anemia with marked erythroblastosis in mouse and man^{23,27,40,60}.

Given the DNA-binding affinity is globally reduced, it is unclear why some genes are particularly sensitive to the linker 2 mutation and some are not. There were no obvious differences in promoter/enhancer contexts (e.g. additional DNA-motifs such as those that bind GATA1) between H350R-sensitive and insensitive genes. Some KLF1 target genes are also dose sensitive in humans who are simple carriers for *KLF1* loss of function (LoF) mutations^{20,21,23,66,67}. So, the mechanism for the selective dependency of some genes remains uncertain.

Nevertheless, this study sheds light on the general mechanisms by which mutations in the linkers of zinc finger TFs cause disease. Zinc finger TFs are one of the largest gene families in the mammalian genome. Linker mutations in *ZBTB20* cause Primrose Syndrome³³, and similar mutations in *ZNF407* cause cognitive impairment³². One important lesson from this work is that missense mutations in TFs can result in different phenotypes from loss of function mutations because only a subset of target genes are affected. Thus, it can be difficult to suspect mutations in well known genes as likely disease-causing candidates based on clinical findings. These genes might not be considered for re-sequencing in such cases and the genetic etiology might remain a mystery.

Methods

Materials and methods

Mouse breeding and tissue collection. All animal procedures were conducted under approval by the University of Queensland and The Alfred Research Alliance Animal Ethics Committees. Animals were housed in Greenline techniplast IVC cages with corn cob bedding with 12/12 h dark/light cycle, 23 °C ambient temperature and humidity maintained at 70%. Animals were fed standard chow *ad libitum*. Heterozygous *mommed45* (D45) mice (*Klf1*^{H350R/-})²⁴, were interbred with *Klf1* heterozygous knockout (*Klf1*^{+/-}) mice¹². Both male and female

offspring were used in this study and sacrificed by cervical dislocation. Pregnant females were sacrificed at E14.5 and single cell suspensions of FL cells were lysed in TRIzol. For rescue experiments, L-leucine was added at 1.5% in the drinking water from 6 weeks of age for 4 weeks.

Flow cytometry. Flow cytometry analyses were performed on blood, spleen, and bone marrow collected from mice at 6–8 weeks of age. FACS of splenic erythroblasts was performed on spleens from animals between 8–12 weeks of age. Antibodies used in this study are listed in Supplementary Table 1. All samples were analyzed on the BD LSRII flow cytometer. Data and cytometry plots were analyzed and generated with FlowJo (v9) software.

Histology and hematological evaluation. Spleens were fixed in 4% PFA overnight at 4 °C, dehydrated in 70% ethanol and embedded in paraffin. 4μm sections were stained with Haematoxylin and Eosin. Blood smears were prepared from peripheral blood from two-week-old mice and stained with May-Grünwald-Giemsa. Peripheral blood from six-week-old mice was collected and analyzed using a Celdyn Sapphire or Celdyn Emerald (Abbott diagnostics) automated blood analyzer with mouse software.

Transcriptome profiling and gene expression analysis. RNA from female *Klf1*^{+/-}, *Klf1*^{H350R/-} and *Klf1*^{-/-} fetal livers (FLs) were extracted from TRIzol suspensions (ThermoFisher) as per manufacturer's protocol. Whole fetal liver samples were used because it is not possible to employ CD71 and TER119 surface expression to sort stages of erythroid gene expression in *Klf1*^{-/-} fetal livers^{18,34}. Samples with a Bioanalyser RIN Score >7 (Agilent Technologies) were filtered through the GLOBINclear™ kit (ThermoFisher) and mRNA was selected using the Dynabeads mRNA DIRECT kit (ThermoFisher). cDNA libraries were prepared from the filtered RNA using the Ion total RNA-seq Kit-v2 (ThermoFisher). These were sequenced on the Ion Proton as described²⁷. Reads were mapped to the mouse genome (mm9) using a combination of Tophat and TMAP²⁷. Cuffdiff was used to identify differential gene expression between the three genotypes: *Klf1*^{+/-}, *Klf1*^{H350R/-} and *Klf1*^{-/-}⁶⁸. Gene expression levels were validated by qRT-PCR using SYBR green PCR master mix (ThermoFisher). Primer sequences are available in Supplementary Table 2. Graphs were generated using RStudio with the ggplots2 package (ver. 0.98.1103) and Graphpad Prism (ver.5.01).

RNA from FACS-sorted CD71+ TER119+ splenic erythrocytes of five viable mouse genotypes was extracted from TRIzol. mRNA was isolated from 1 μg total RNA using the NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB; #E7490). RNA-seq libraries were prepared using the NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (NEB; #E7760) according to the manufacturer's instructions. Libraries were multiplexed and sequenced on an Illumina NovaSeq 6000 in paired end mode with 50 bp reads. Reads were aligned to the mouse genome (mm9) with hisat2(2.1.0). Picard (2.9.2) was used to remove optical duplicates. Gene level counts were obtained with subRead-featureCounts (1.6.4). Differential expression analysis was performed with Voom/Limma through the DEGUST web portal (4.2-dev).

Chromatin Immunoprecipitation with sequencing. ChIP was performed using a rabbit polyclonal antibody raised against the N-terminus of KLF1¹⁶ using FL cells from wild type and homozygous *Klf1*^{H350R/H350R} embryos. Two immunoprecipitations from FLs for each genotype were pooled to increase library complexity. Sequencing libraries were generated from ChIP-ed and input DNA using the NEBNext ChIP-seq Library prep reagent set (NEB #E6200) and NEBNext Multiplex oligo for Illumina (NEB #E7335). Libraries were sequenced on a Nextseq500 (Illumina). Reads were mapped to the mouse genome (mm9 and mm10) using Bowtie2 and filtered through a blacklist and whitelist (re-introducing α and β-globin multimapping regions). PCR

sequencing duplicates were removed using Picard tools (<https://broadinstitute.github.io/picard/>). MACS2 was used to identify peaks depicting wild type KLF1 and KLF1-H350R bound regions. DNA enrichment was validated by qPCR using primers spanning the peak and genomic regions -1 kb up and downstream of the peak (See Supplementary Table 2 for primer sequences). Graphs were generated using Rstudio with the ggplots2 package (ver. 0.98.1103). Heat maps of enriched sequencing read depth were created using EaSeq⁶⁹

Recombinant GST-ZF protein purification and in vitro binding affinity assays. The ZF domain of murine KLF1 (encoding amino acids 261 to 376) was amplified by RT-PCR from *Klf1*^{H350R/H350R} FL cDNA and cloned into pGEX-6p-1 in frame with GST, to generate KLF1-H350R-zf. Mutations were introduced by PCR mutagenesis and cloned into the same site of pGEX-6p-1 to generate KLF1-T348A-zf and KLF1-G349R-zf. Generation of the wild type GST-KLF-zf vector was previously reported²⁷. Rosetta (DE3) cells were transformed and selected in ampicillin and chloramphenicol. Expression of GST-fusion proteins was induced overnight at 16 °C in 1 mM IPTG. Fusion proteins were then extracted and purified on Glutathione-Sepharose 4B (GE Healthcare #17075601) beads, eluted with reduced glutathione in elution buffer as previously described²⁷. Protein purity and yield was monitored using SDS-PAGE with Coomassie Blue staining (Supplementary Fig. 5), and function verified by electrophoretic mobility shift assay (EMSA). Oligos (20mers) were end labeled with ³²P-γ-ATP and polynucleotide kinase, then annealed with excess reverse strand oligo and purified using Sephadex G-25 spin columns (Roche). The forward strand oligo sequences used are provided in Supplementary Table 2. EMSA was performed as described²⁷. Free and bound probes were quantified using a Typhoon Phosphor Imager and Image Quant (IQ) software. The K_D was calculated by plotting bound probe (shifted) versus total (bound plus free). A polynomial curve ($n = 3$) of best fit was generated in graphpad and the protein concentration at which 50% of the probe is bound (i.e. K_D) was assigned from this curve.

Structural analysis. The structure of KLF4 bound to DNA (PDB ID 5KE6)⁵³ was visualized and Supplementary Fig. 6 was prepared using the UCSF Chimera Software Package⁷⁰.

Statistics and reproducibility. Statistical analysis was performed in Graphpad Prism. One-way Anova with a tukey post-hoc test was used to compare the hematological parameters and qRT-PCR results.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All sequencing data generated in this manuscript are available through GEO under accessions GSE94351 and GSE240553. All data are available within the main text, figures and supplementary information. Source data are provided with this publication and requests for additional information should be directed to the corresponding author. Source data are provided with this paper.

References

- Ravasi, T. et al. Systematic characterization of the zinc-finger-containing proteins in the mouse transcriptome. *Genome Res.* **13**, 1430–1442 (2003).
- Bellefroid, E. J. et al. The human genome contains hundreds of genes coding for finger proteins of the Kruppel type. *DNA* **8**, 377–387 (1989).
- Presnell, J. S., Schnitzler, C. E. & Browne, W. E. KLF/SP transcription factor family evolution: expansion, diversification, and innovation in eukaryotes. *Genome Biol. Evol.* **7**, 2289–2309 (2015).
- Zhao, Y. et al. “Stripe” transcription factors provide accessibility to co-binding partners in mammalian genomes. *Mol. Cell* **82**, 3398–3411 e11 (2022).
- McConnell, B. B. & Yang, V. W. Mammalian Kruppel-like factors in health and diseases. *Physiol. Rev.* **90**, 1337–1381 (2010).
- Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).
- van Vliet, J. et al. Human KLF17 is a new member of the Sp/KLF family of transcription factors. *Genomics* **87**, 474–482 (2006).
- Laity, J. H., Dyson, H. J. & Wright, P. E. DNA-induced alpha-helix capping in conserved linker sequences is a determinant of binding affinity in Cys(2)-His(2) zinc fingers. *J. Mol. Biol.* **295**, 719–727 (2000).
- Choo, Y. & Klug, A. A role in DNA binding for the linker sequences of the first three zinc fingers of TFIIIA. *Nucleic Acids Res.* **21**, 3341–3346 (1993).
- Pavletich, N. P. & Pabo, C. O. Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å. *Science* **252**, 809–817 (1991).
- Miller, I. J. & Bieker, J. J. A novel, erythroid cell-specific murine transcription factor that binds to the CACCC element and is related to the Kruppel family of nuclear proteins. *Mol. Cell Biol.* **13**, 2776–2786 (1993).
- Perkins, A. C., Sharpe, A. H. & Orkin, S. H. Lethal beta-thalassaemia in mice lacking the erythroid CACCC- transcription factor EKLF. *Nature* **375**, 318–322 (1995).
- Magor, G. W. et al. KLF1-null neonates display hydrops fetalis and a deranged erythroid transcriptome. *Blood* **125**, 2405–2417 (2015).
- Nuez, B., Michalovich, D., Bygrave, A., Ploemacher, R. & Grosveld, F. Defective haematopoiesis in fetal liver resulting from inactivation of the EKLF gene. *Nature* **375**, 316–318 (1995).
- Tallack, M. R. et al. Novel roles for KLF1 in erythropoiesis revealed by mRNA-seq. *Genome Res.* **22**, 2385–2398 (2012).
- Tallack, M. R. et al. A global role for KLF1 in erythropoiesis revealed by ChIP-seq in primary erythroid cells. *Genome Res.* **20**, 1052–1063 (2010).
- Hay, D. et al. Genetic dissection of the alpha-globin super-enhancer in vivo. *Nat. Genet.* **48**, 895–903 (2016).
- Pilon, A. M. et al. Failure of terminal erythroid differentiation in EKLF-deficient mice is associated with cell cycle perturbation and reduced expression of E2F2. *Mol. Cell Biol.* **28**, 7394–7401 (2008).
- Gnanapragasam, M. N. et al. EKLF/KLF1-regulated cell cycle exit is essential for erythroblast enucleation. *Blood* **128**, 1631–1641 (2016).
- Viprakasit, V. et al. Mutations in Kruppel-like factor 1 cause transfusion-dependent hemolytic anemia and persistence of embryonic globin gene expression. *Blood* **123**, 1586–1595 (2014).
- Huang, J. et al. Compound heterozygosity for KLF1 mutations is associated with microcytic hypochromic anemia and increased fetal hemoglobin. *Eur. J. Hum. Genet.* **23**, 1341–1348 (2015).
- Gallienne, A. E., Dreau, H. M., Schuh, A., Old, J. M. & Henderson, S. Ten novel mutations in the erythroid transcription factor KLF1 gene associated with increased fetal hemoglobin levels in adults. *Hematologica* **97**, 340–343 (2012).
- Perkins, A. et al. Kruppeling erythropoiesis: an unexpected broad spectrum of human red blood cell disorders due to KLF1 variants. *Blood* **127**, 1856–1862 (2016).
- Sorolla, A. et al. Identification of novel hypomorphic and null mutations in *Klf1* derived from a genetic screen for modifiers of alpha-globin transgene variegation. *Genomics* **105**, 116–122 (2015).
- Nilson, D. G., Sabatino, D. E., Bodine, D. M. & Gallagher, P. G. Major erythrocyte membrane protein genes in EKLF-deficient mice. *Exp. Hematol.* **34**, 705–712 (2006).
- Tallack, M. R., Keys, J. R., Humbert, P. O. & Perkins, A. C. EKLF/KLF1 controls cell cycle entry via direct regulation of E2f2. *J. Biol. Chem.* **284**, 20966–20974 (2009).

27. Gillinder, K. R. et al. Promiscuous DNA-binding of a mutant zinc finger protein corrupts the transcriptome and diminishes cell viability. *Nucleic Acids Res.* **45**, 1130–1143 (2017).
28. Perkins, A. C. & Bieker, J. Congenital anemia phenotypes due to KLF1 mutations. *J. Pediatr. Hematol. Oncol.* **43**, e148–e149 (2020).
29. Planutis, A. et al. Neomorphic effects of the neonatal anemia (Nan-Eklf) mutation contribute to deficits throughout development. *Development* **144**, 430–440 (2017).
30. Ilesley, M. D. et al. Corrupted DNA-binding specificity and ectopic transcription underpin dominant neomorphic mutations in KLF/SP transcription factors. *BMC Genomics* **20**, 417 (2019).
31. Kulczynska, K., Bieker, J. J. & Siatecka, M. A Kruppel-like factor 1 (KLF1) Mutation Associated with Severe Congenital Dyserythropoietic Anemia Alters Its DNA-Binding Specificity. *Mol. Cell Biol.* **40**, e00444–19 (2020).
32. Kambouris, M. et al. Mutations in zinc finger 407 [ZNF407] cause a unique autosomal recessive cognitive impairment syndrome. *Orphanet J. Rare Dis.* **9**, 80 (2014).
33. Cordeddu, V. et al. Mutations in ZBTB20 cause Primrose syndrome. *Nat. Genet.* **46**, 815–817 (2014).
34. Hodge, D. et al. A global role for EKLF in definitive and primitive erythropoiesis. *Blood* **107**, 3359–3370 (2006).
35. Tallack, M. R. & Perkins, A. C. KLF1 directly coordinates almost all aspects of terminal erythroid differentiation. *IUBMB Life* **62**, 886–890 (2010).
36. Porpiglia, E., Hidalgo, D., Koulis, M., Tzafiriri, A. R. & Socolovsky, M. Stat5 signaling specifies basal versus stress erythropoietic responses through distinct binary and graded dynamic modalities. *PLoS Biol.* **10**, e1001383 (2012).
37. Gillinder, K. R. et al. Direct targets of pStat5 signalling in erythropoiesis. *PLoS ONE* **12**, e0180922 (2017).
38. Montel-Hagen, A. et al. The Glut1 and Glut4 glucose transporters are differentially expressed during perinatal and postnatal erythropoiesis. *Blood* **112**, 4729–4738 (2008).
39. Nebor, D. et al. Mutant KLF1 in adult anemic nan mice leads to profound transcriptome changes and disordered erythropoiesis. *Sci. Rep.* **8**, 12793 (2018).
40. Siatecka, M. et al. Severe anemia in the Nan mutant mouse caused by sequence-selective disruption of erythroid Kruppel-like factor. *Proc. Natl Acad. Sci. USA* **107**, 15151–15156 (2010).
41. Perkins, A. C., Gaensler, K. M. & Orkin, S. H. Silencing of human fetal globin expression is impaired in the absence of the adult beta-globin gene activator protein EKLF. *Proc. Natl Acad. Sci. USA* **93**, 12267–12271 (1996).
42. Borg, J. et al. Haploinsufficiency for the erythroid transcription factor KLF1 causes hereditary persistence of fetal hemoglobin. *Nat. Genet.* **42**, 801–805 (2010).
43. Burdach, J. et al. Regions outside the DNA-binding domain are critical for proper in vivo specificity of an archetypal zinc finger transcription factor. *Nucleic Acids Res.* **42**, 276–289 (2014).
44. Loft, A. et al. Browning of human adipocytes requires KLF11 and reprogramming of PPARgamma superenhancers. *Genes Dev.* **29**, 7–22 (2015).
45. Aksoy, I. et al. Klf4 and Klf5 differentially inhibit mesoderm and endoderm differentiation in embryonic stem cells. *Nat. Commun.* **5**, 3719 (2014).
46. Chen, X. et al. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133**, 1106–1117 (2008).
47. Terrados, G. et al. Genome-wide localization and expression profiling establish Sp2 as a sequence-specific transcription factor regulating vitally important genes. *Nucleic Acids Res.* **40**, 7844–7857 (2012).
48. Khan, A. A. et al. Dematin and adducin provide a novel link between the spectrin cytoskeleton and human erythrocyte membrane by directly interacting with glucose transporter-1. *J. Biol. Chem.* **283**, 14600–14609 (2008).
49. Rungaldier, S., Oberwagner, W., Salzer, U., Csaszar, E. & Prohaska, R. Stomatin interacts with GLUT1/SLC2A1, band 3/SLC4A1, and aquaporin-1 in human erythrocyte membrane domains. *Biochim Biophys. Acta* **1828**, 956–966 (2013).
50. Fandos, C. et al. GLUT1 glucose transporter gene transcription is repressed by Sp3. Evidence for a regulatory role of Sp3 during myogenesis. *J. Mol. Biol.* **294**, 103–119 (1999).
51. Lun, A. T. & Smyth, G. K. csaw: a Bioconductor package for differential binding analysis of ChIP-seq data using sliding windows. *Nucleic Acids Res.* **44**, e45 (2015).
52. Tepakhan, W. et al. Nine known and five novel mutations in the erythroid transcription factor KLF1 gene and phenotypic expression of fetal hemoglobin in hemoglobin E disorder. *Blood Cells Mol. Dis.* **59**, 85–91 (2016).
53. Hashimoto, H. et al. Distinctive Klf4 mutants determine preference for DNA methylation status. *Nucleic Acids Res.* **44**, 10177–10185 (2016).
54. Pavletich, N. P. & Pabo, C. O. Crystal structure of a five-finger GLI-DNA complex: new perspectives on zinc fingers. *Science* **261**, 1701–1707 (1993).
55. Doig, A. J. & Baldwin, R. L. N- and C-capping preferences for all 20 amino acids in alpha-helical peptides. *Protein Sci.* **4**, 1325–1336 (1995).
56. Chung, J. et al. The mTORC1/4E-BP pathway coordinates hemoglobin production with L-leucine availability. *Sci. Signal* **8**, ra34 (2015).
57. Liu, D. et al. KLF1 mutations are relatively more common in a thalassemia endemic region and ameliorate the severity of beta-thalassemia. *Blood* **124**, 803–811 (2014).
58. Perseu, L. et al. KLF1 gene mutations cause borderline HbA(2). *Blood* **118**, 4454–4458 (2011).
59. Singleton, B. K., Frayne, J. & Anstee, D. J. Blood group phenotypes resulting from mutations in erythroid transcription factors. *Curr. Opin. Hematol.* **19**, 486–493 (2012).
60. Arnaud, L. et al. A dominant mutation in the gene encoding the erythroid transcription factor KLF1 causes a congenital dyserythropoietic anemia. *Am. J. Hum. Genet.* **87**, 721–727 (2010).
61. Lu, Y. et al. Gene disruption of dematin causes precipitous loss of erythrocyte membrane stability and severe hemolytic anemia. *Blood* **128**, 93–103 (2016).
62. Flatt, J. F. et al. Stomatin-deficient cryohydrocytosis results from mutations in SLC2A1: a novel form of GLUT1 deficiency syndrome. *Blood* **118**, 5267–5277 (2011).
63. Oburoglu, L. et al. Glucose and glutamine metabolism regulate human hematopoietic stem cell lineage specification. *Cell Stem Cell* **15**, 169–184 (2014).
64. Niihara, Y. et al. A phase 3 trial of l-glutamine in sickle cell disease. *N. Engl. J. Med.* **379**, 226–235 (2018).
65. Clemens, K. R. et al. Relative contributions of the zinc fingers of transcription factor IIIA to the energetics of DNA binding. *J. Mol. Biol.* **244**, 23–35 (1994).
66. Singleton, B. K., Burton, N. M., Green, C., Brady, R. L. & Anstee, D. J. Mutations in EKLF/KLF1 form the molecular basis of the rare blood group In(Lu) phenotype. *Blood* **112**, 2081–2088 (2008).
67. Singleton, B. K. et al. Mutations in the second zinc finger of human EKLF reduce promoter affinity but give rise to benign and disease phenotypes. *Blood* **118**, 3137–3145 (2011).
68. Trapnell, C. et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7**, 562–578 (2012).
69. Lerdrup, M., Johansen, J. V., Agrawal-Singh, S. & Hansen, K. An interactive environment for agile analysis and visualization of ChIP-seq data. *Nat. Struct. Mol. Biol.* **23**, 349–357 (2016).
70. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

71. Glogowska, E., Lezon-Geyda, K., Maksimova, Y., Schulz, V. P. & Gallagher, P. G. Mutations in the Gardos channel (KCNN4) are associated with hereditary xerocytosis. *Blood* **126**, 1281–1284 (2015).
72. Xue, L., Galdass, M., Gnanapragasam, M. N., Manwani, D. & Bieker, J. J. Extrinsic and intrinsic control by EKLF (KLF1) within a specialized erythroid niche. *Development* **141**, 2245–2254 (2014).
73. Choi, H. S., Lee, E. M., Kim, H. O., Park, M. I. & Baek, E. J. Autonomous control of terminal erythropoiesis via physical interactions among erythroid cells. *Stem Cell Res.* **10**, 442–453 (2013).
74. Funnell, A. P. et al. Erythroid Kruppel-like factor directly activates the basic Kruppel-like factor gene in erythroid cells. *Mol. Cell Biol.* **27**, 2777–2790 (2007).
75. Ilesley, M. D. et al. Kruppel-like factors compete for promoters and enhancers to fine-tune transcription. *Nucleic Acids Res.* **45**, 6572–6588 (2017).
76. Hattangadi, S. M. et al. Histones to the cytosol: exportin 7 is essential for normal terminal erythroid nuclear maturation. *Blood* **124**, 1931–1940 (2014).
77. Kiatpakdee, B. et al. Cholesterol-binding protein TSPO2 coordinates maturation and proliferation of terminally differentiating erythroblasts. *J. Biol. Chem.* **295**, 8048–8063 (2020).

Acknowledgements

This work was supported by NHMRC grants APP1082439 and APP1183556.

Author contributions

S.H.: performed research, analyzed data, wrote the paper; K.G.: performed research, analyzed data, bioinformatics, wrote the paper; H.M.: performed research and provided bioinformatics support; C.R., M.I.: performed research; G.M.: provided bioinformatics support, analyzed data; M.T.: annotated and curated gene lists; M.L.: undertook structural modeling; A.P.: designed research, performed research, wrote the paper, funded the research

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-024-50579-4>.

Correspondence and requests for materials should be addressed to Andrew Perkins.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024