Rational targeting of a NuRD subcomplex guided by comprehensive in situ mutagenesis

Falak Sher1,2,†, Mir Hossain1,11, Davide Seruggia1,11, Vivien A. C. Schoonenberg1,3, Qiuming Yao1,4, Paolo Cifani5, Laura M. K. Dassama1, Mitchel A. Cole1, Chunyan Ren1, Divya S. Vinjamur1, Claudio Macias-Trevino1, Kevin Luk5, Connor McGuckin1, Patrick G. Schupp1, Matthew C. Canver1, Ryo Kurita7, Yukio Nakamura8, Yuko Fujiwara1, Scot A. Wolfe6, Luca Pinello4, Takahiro Maeda9, Alex Kentsis5, Stuart H. Orkin1,10 and Daniel E. Bauer1,10*

Developmental silencing of fetal globins serves as both a paradigm of spatiotemporal gene regulation and an opportunity for therapeutic intervention of β-hemoglobinopathy. The nucleosome remodeling and deacetylase (NuRD) chromatin complex participates in γ-globin repression. We used pooled CRISPR screening to disrupt NuRD protein coding sequences comprehensively in human adult erythroid precursors. Essential for fetal hemoglobin (HbF) control is a non-redundant subcomplex of NuRD protein family paralogs, whose composition we corroborated by affinity chromatography and proximity labeling mass spectrometry proteomics. Mapping top functional guide RNAs identified key protein interfaces where in-frame alleles resulted in loss-of-function due to destabilization or altered function of subunits. We ascertained mutations of CHD4 that dissociate its requirement for cell fitness from HbF repression in both primary human erythroid progenitors and transgenic mice. Finally we demonstrated that sequestering CHD4 from NuRD phenocopied these mutations. These results indicate a generalizable approach to discover protein complex features amenable to rational biochemical targeting.

S evere hemoglobinopathies resulting from mutations of the adult β-globin gene (HBB) including sickle cell disease (SCD) and β-thalassemia affect millions worldwide1–3. Derepression of the fetal γ-globin genes (HBG1/HBG2) resulting in HbF (α2γ2) induction holds great potential to ameliorate the pathophysiology of SCD and β-thalassemia1–3. γ-Globin silencing is an active process dependent on interactions between DNA binding factors and chromatin readers, writers and erasers acting within multiprotein nuclear complexes. The NuRD complex is a chromatin modifier composed of six different protein family subunits4–6. Each subunit family is comprised of multiple paralog proteins that can variably combine to form NuRD multiprotein complexes. The catalytic subunits are the ATP-dependent nucleosome remodelers CHD3 and CHD4 and the histone deacetylases HDAC1 and HDAC2. Other members include the methyl-CpG binding proteins MBD2 and MBD3 and structural subunits MTA1, MTA2 and MTA3, GATAD2A and GATAD2B, and RBBP4 and RBBP7. Various NuRD members have been implicated in developmental silencing of HbF. NuRD members occupy embryonic and fetal globin gene5,10. Knockdown of MBD2 induces the expression of γ-globin in human β-globin locus transgenic mice14 and in human CD34+ hematopoietic stem and progenitor cells (HSPCs) derived adult erythroid cells15,16. Conditional knockout of Chd4 results in derepression of γ-globin in β-YAC transgenic mice and cultured murine chemical inducer of dimerization hematopoietic cells17. Knockdown of CHD4 in primary human erythroid cells results in robust increase of γ-globin expression18,19. A coiled–coil protein interaction between MBD2 and GATAD2A is necessary for γ-globin gene repression and could be a potential target for molecular intervention20. Genetic knockdown or chemical inhibition of HDAC1 and HDAC2 induces HbF in adult erythroid progenitors21,22. Initially discovered by genome-wide association studies (GWAS) as a locus associated with HbF levels23,24, the transcriptional repressor BCL11A has been validated as a critical negative regulator of γ-globin expression25–27. Biochemical studies have revealed that BCL11A physically interacts with NuRD complex subunits including CHD3/4, HDAC1/2, MTA1/2/3, RBBP4/7 and MBD3 (ref. 16). More recently ZBTB7A has been reported as a γ-globin repressor28. ZBTB7A confers its repressive activity non-redundantly with BCL11A, yet also physically interacts with NuRD subunits including MTA2, HDAC1/2, GATAD2B. Together these data provide the impetus to define the mechanisms through which NuRD represses HbF and to identify possible molecular targets for pharmacotherapy (also see Supplementary Note).

Here we investigated the coding sequences within the NuRD complex associated with HbF repression by using CRISPR-Cas9 dense mutagenesis in human umbilical cord blood-derived...
Results

CRISPR dense in situ mutagenesis reveals NuRD complex members essential for HbF repression. We hypothesized that CRISPR-Cas9 dense in situ mutagenesis could reveal critical NuRD sequences at which in-frame alleles result in loss-of-function. We compared HbF enrichment scores among the different NuRD subunits (Fig. 1a; also see Supplementary Note). As expected, single guide RNA (sgRNAs) targeting positive control genes BCL11A and ZBTB7A showed robust HbF enrichment as compared to non-targeting sgRNAs (Fig. 1b). We defined hit genes, that is those with biological phenotype, as those at which at least 75% of the sgRNAs exceeded the median non-targeting sgRNA score34. We discovered that among the 13 NuRD subunit genes, only five genes, CHD2, GATAD2A, HDAC2, MBD2 and MTA2, were required for HbF repression, whereas perturbations of the coding sequences of the other eight NuRD genes, CHD3, GATAD2B, HDAC1, MBD3, MTA1, MTA3, RBBP4 and RBBP7, did not substantially affect HbF repression (Fig. 1b). Notably within five paralogous gene families (CHD, GATAD2, HDAC, MBD and MTA), only one member was required for HbF repression, whereas the other gene members were dispensable. For example, HDAC2 was required for HbF repression, and while HDAC1 was not, MTA2 was required for HbF repression while MTA1 and MTA3 were not, and so forth. This observation suggested that a subcomplex of NuRD defined by constituent paralogous family members was required for γ-globin repression.

erythroid progenitor (HUDEP-2) adult-stage erythroid cells. Taking into account cellular fitness as a counter-screen, we nominated potential NuRD target regions for therapeutic derepression of HbF that escape cellular toxicity, validated their effects in primary human cells and transgenic mice, and developed a rational therapeutic strategy for HbF induction to phenocopy potent mutations.

Fig. 1 | Dense mutagenesis of NuRD genes by CRISPR-Cas9 pooled screening. a, The upper panel shows an overview of the CRISPR-Cas9 pooled lentiviral screen. All NGG protoscaler adjacent motif (PAM) restricted sgRNAs targeting 13 NuRD genes plus BCL11A and ZBTB7A targeting and non-targeting sgRNAs were synthesized by oligonucleotide array and cloned as pool to lentivirus. HUDEP-2 cells were transduced (n = 3), selected and subjected to sgRNA library distribution. As an example, HbF enrichment scores from MTA2 and HDAC2 were mapped to the dimer of PDB ID 5ICN. The lower panel shows the sgRNA score was calculated by comparing sgRNA abundance in unsorted cells to library. Enrichment scores were mapped to linear and three-dimensional protein maps. As an example, HbF enrichment scores from MTA2 and HDAC2 were mapped to the dimer of PDB ID 5ICN. The lower panel shows the sgRNA library distribution. b, c. HbF enrichment (b) and fitness (c) scores for each sgRNA (sample size per lower panel a) targeting NuRD genes and controls. The box plot shows median, 25th and 75th percentiles with whiskers and outliers per the Tukey method.

**Table 1**

| Negative Control | Positive | Total |
|------------------|----------|-------|
| MTA1             | 833      | 2796  |
| MTA2             | 337      | 2796  |
| MTA3             | 412      | 2796  |
| HDAC1            | 200      | 2796  |
| HDAC2            | 112      | 2796  |
| BCL11A           | 137      | 2796  |
| ZBTB7A           | 145      | 2796  |
| CHD2             | 1037     | 2796  |
| CHD3             | 1037     | 2796  |
| CHD4             | 833      | 2796  |
| GATAD2A          | 376      | 2796  |
| GATAD2B          | 245      | 2796  |
| MBD2             | 225      | 2796  |
| MBD3             | 173      | 2796  |
| RBBP4            | 112      | 2796  |
| RBBP7            | 134      | 2796  |
| NT: 228 sgRNA    | (4.5%)   |       |
| Positive: 429 sgRNA | (8.5%) |       |
| CHD2: 1037 sgRNA | (20.5%)  |       |
| CHD4: 833 sgRNA  | (16.5%)  |       |
| GATAD2A: 376 sgRNA | (7.4%) |       |
| GATAD2B: 245 sgRNA | (4.8%) |       |
| HDAC1: 200 sgRNA | (3.9%)   |       |
| HDAC2: 112 sgRNA | (2.2%)   |       |
| MBD2: 225 sgRNA  | (4.4%)   |       |
| MBD3: 173 sgRNA  | (3.4%)   |       |
| MTA1: 412 sgRNA  | (8.1%)   |       |
| MTA2: 337 sgRNA  | (6.6%)   |       |
| MTA3: 152 sgRNA  | (3.0%)   |       |
| RBBP4: 112 sgRNA | (2.2%)   |       |
| RBBP7: 134 sgRNA | (2.7%)   |       |
We observed that almost all gene targeting sgRNAs had a small negative effect on fitness score as compared to non-targeting sgRNAs (Fig. 1c). We speculated that this was related to the known modest cellular fitness impact of Cas9-mediated DNA damage response\textsuperscript{35–37}. Previous pooled CRISPR screens have shown modest negative fitness impact of neutral gene targeting sgRNAs as compared to non-targeting sgRNAs, consistent with these results\textsuperscript{38}. We observed only two NuRD members with obvious negative fitness scores beyond this modest shared effect, namely CHD4 and RBBP4 (Fig. 1c, also see Supplementary Note). By comparing HbF enrichment scores and fitness scores, we distinguished four functional classes of NuRD genes, GATA2A, HDAC2, MBD2 and MTA2, as essential for HbF repression but not cell fitness, CHD4 as essential both for HbF repression and cell fitness, RBBP4 as essential for cell fitness, and CHD3, GATA2B, HDAC1, MBD3, MTA1, MTA3 and RBBP7 as neither essential for HbF repression nor for cell fitness. For each of the six NuRD constituent paralogous families, just one member of each family was required for HbF repression or cellular fitness.

Proteomic analysis corroborates the HbF repressive NuRD subcomplex. We analyzed RNA-seq data generated from HUDEP-2 cells for all the NuRD genes to compare the expression of each paralog to its functional requirement (Fig. 2a). We found that messenger RNA expression levels of CHD4, GATA2A and MTA2 were significantly higher than their respective paralogs (CHD3, GATA2B, MTA1 and MTA3) suggesting that higher gene expression could account for the paralog-specific dependence. However, HDAC2 and HDAC1 shared similar mRNA expression levels, MBD3 showed higher mRNA expression as compared to MBD2, and RBBP7 showed higher expression as compared to RBBP4.

Protein abundance is not necessarily predicted by mRNA abundance\textsuperscript{39–41}. To further evaluate the paralog-specific composition of the NuRD protein complex, we performed affinity purification of proteins physically associated with MTA2 and CHD4 using immunoprecipitation and label-free quantification by high-resolution mass spectrometry (IP-MS, Supplementary Fig. 2a). To test the identity of the erythroid NuRD subcomplex members orthogonally, we evaluated the proteins physically neighboring MTA2 in its intracellular milieu by proximity labeling (Supplementary Fig. 2b,c). Overall we found strong agreement of all three methods in detecting interacting NuRD subunit members (Fig. 2b–c, Supplementary Fig. 2d–f, Supplementary Table 1, Supplementary Note). We found 31 proteins at the intersection of all three MS experiments (Fig. 2e). With respect to NuRD, the shared proteins included seven NuRD subunits, preferentially identifying the NuRD paralogs found to be functional by CRISPR screening. Specifically, six of the proteomics identified NuRD members (MTA2, RBBP4, CHD4, GATA2A, HDAC2 and MBD2) were found to be critical by CRISPR screening, whereas only MBD3 was dispensable for HbF repression and cellular fitness (also see Supplementary Note).

Functional maps of NuRD subunits essential for HbF repression. We hypothesized that sgRNAs targeting critical regions of functional NuRD subunits would show heightened enrichment scores due to an increased likelihood of loss-of-function in-frame alleles\textsuperscript{42–44}. To test this, we compared the functional scores against various protein-level sequence annotations, including evolutionary conservation, protein disorder, domain identity and secondary structure (Supplementary Table 2). We observed an inverse relationship between the Protein Variation Effect Analyzer (PROVEAN) conservation score and HbF enrichment (Spearman $r = -0.329, P < 0.0001$) indicating that targeting more conserved positions within NuRD hit genes is more likely to result in HbF derepression (Fig. 3a). Similarly, a correlation between PROVEAN score and cellular fitness score (Fig. 3b) showed that the sgRNAs targeting more conserved amino acid residues exerted greater fitness cost (Spearman $r = 0.235, P < 0.0001$). These results support the hypothesis that CRISPR–Cas9-mediated comprehensive in situ mutagenesis can identify critical NuRD complex protein coding sequences (also see Supplementary Fig. 3 and Supplementary Note).

To visualize amino acid residues that result in maximal HbF derepression and minimal cellular toxicity upon disruption, we generated linear maps of HbF enrichment and fitness scores for all sgRNAs at the NuRD hit genes (Fig. 3c). We plotted individual sgRNA scores as well as a smoothed score based on local polynomial regression. In addition, we mapped the PROVEAN conservation scores, disorder scores, secondary structure predictions and domain annotations. We observed visually apparent examples of functional scores correlating with each of these annotations. For example, at CHD4 we found heightened HbF enrichment scores at conserved, ordered sequences with secondary structure predictions, as compared to neighboring sequences. This pattern was evident around characterized domains including the catalytic ATP-dependent helicase domain, chromatin-interacting regions including two PHD fingers and two chromodomains, an N-terminal HMG-like domain, two domains of unknown function (DUF1087 and DUF1086) and a C-terminal domain HDCT2 (Fig. 3c). Similarly, we found heightened scores at the N-terminal regions of MTA2 comprising various domains critical for NuRD complex formation\textsuperscript{45}, at the catalytic histone deacetylase domain of HDAC2 and at the MBD2 domain that binds methyl-CpG DNA\textsuperscript{46–48}. The NuRD regions identified as most critical for HbF repression by dense mutagenesis included those involved in intra-subunit interactions. For example, a heterodimeric interaction of a coiled–coil region of GATA2A (AA137–179) with a coiled–coil region of MBD2 (AA360–393) is required for $\gamma$-globin repression\textsuperscript{41}. Elevated HbF enrichment scores were observed for sgRNAs targeting the GATA2A coiled–coil region as compared to flanking sequences (Fig. 3c). The sgRNA HbF enrichment score maps did not identify functional sequences within the NuRD subunits CHD3, MTA1, MTA3, MBD3, GATA2B, HDAC1 and RBBP7 (Supplementary Fig. 4).

We compared the high-resolution functional scores based on Cas9-mediated comprehensive in situ mutagenesis to available protein structures. We colored these structures by overlaying the regression HbF enrichment scores onto each amino acid. We found that the functional scores marked binding interfaces as particularly vulnerable to disruption (also see Supplementary Note). For example, recoloring with HbF enrichment scores the structure of HDAC2 in complex with the HDAC inhibitor vorinostat (PDB ID 4LXZ)\textsuperscript{46} suggested that residues adjacent to the vorinostat binding site had heightened HbF enrichment scores as compared to neighboring residues (Fig. 3d). The interface of MBD2 abutting meCpG DNA (PDB ID 6CNO)\textsuperscript{49} demonstrated elevated scores as compared to the non-DNA facing region of MBD2, suggesting that sgRNAs that disrupt meCpG binding are especially potent in terms of HbF induction (Fig. 3e).

Identification of critical residues of MTA2 required for HbF repression. MTA subunits are considered to be the scaffolds onto which the NuRD complex assembles\textsuperscript{50}. Dense mutagenesis identified multiple regions of heightened HbF enrichment scores at MTA2 overlapping its conserved, ordered domains (Fig. 3c), including the BAH domain implicated in chromatin interaction\textsuperscript{46}, the ELM2 and SANT domains involved in MTA homodimerization and HDAC recruitment, a GATA-like zinc finger (ZF) domain and RBBP4 interacting region\textsuperscript{51}. To test the hypothesis that the residues with heightened functional CRISPR scores reveal positions at which in-frame deletions are associated with loss-of-function, we characterized clones with defined mutations of the MTA2 NuRD scaffold subunit (see Supplementary Note). Clones with frameshift mutations showed derepression of HbF in a manner that was independent of whether the targeted amino acid resided in a domain.
or at NC sequences (Fig. 4a,b). In-frame deletions did not result in elevated HbF levels when targeted to the NC position with modest HbF enrichment scores but, in contrast, clones with in-frame deletions at positions of heightened HbF enrichment score within the BAH, ELM2 and SANT domains demonstrated derepression of HbF, indicating loss-of-function of MTA2. These results support the hypothesis that heightened functional CRISPR scores mark regions where in-frame deletions result in loss-of-function alleles.

We evaluated individual in-frame deletion clones (Supplementary Fig. 5c) to examine the biochemical basis for HbF induction following mutation of MTA2. We observed two classes of loss-of-function clones, one in which MTA2 protein was lost and another in which MTA2 protein level was preserved but function was impaired (Fig. 4c–e and Supplementary Fig. 5d). Immunoprecipitation by anti-MTA2 or anti-CHD4 followed by immunoblot with NuRD subunits showed that the MTA2 in-frame deletion in clone M4 resulted in reduced capacity for NuRD complex formation (Fig. 4f,g and Supplementary Fig. 5e,f). Reduced interaction of MTA2 with CHD4, MBD2, HDAC2 and GATAD2A was observed in this clone. To assess the composition of the NuRD complex, we performed glycerol gradient density sedimentation analysis of nuclear extracts\(^6\), comparing control and MTA2 M4 and M7 mutant clones (Fig. 4h). In control cells we observed cosedimentation of CHD4, MTA2, GATAD2A, MBD2, HDAC2 and RBBP4 as the NuRD subcomplex as well as in various lower molecular weight fractions, suggesting the presence of free proteins, intermediate subcomplexes (such as the so-called NuDe complex lacking CHD3/CHD4\(^{51}\) and alternative complexes. In clone M4 we found MTA2 was expressed yet relatively depleted from higher molecular weight fractions, consistent with reduced incorporation of MTA2 to the NuRD complex. Concomitantly we observed decreased incorporation of CHD4, GATAD2A, MBD2, HDAC2 and RBBP4 to the NuRD complex. In contrast, in clone M7 we observed the absence of MTA2, consistent with its destabilization, along with reduced incorporation of CHD4 and GATAD2A into the NuRD complex. These results suggested that impairment of the recruitment of CHD4 to NuRD could be a final common pathway of NuRD functional disruption. Together these results showed two modes of action of in-frame loss-of-function deletions of MTA2. One mode was loss of MTA2, associated with formation of aberrant complexes. A second mode was mutations that preserve the levels of MTA2 but disrupt its function in supporting the assembly of NuRD.
Fig. 3 | Maps of functional NuRD subcomplex. a, b, Scatter plots of HbF enrichment (a) and fitness (b) scores compared to PROVEAN conservation scores for individual sgRNAs for hit genes (n = 3765 sgRNAs for a and 1955 for b). Spearman r and P values are shown. Each dot represents sgRNA average score from three independent experiments. c, Linear maps of hit NuRD genes. HbF enrichment scores and fitness scores are shown for each sgRNA as dots, with the LOESS regression line in red. Disorder scores are shown as a heatmap from white to blue from maximal disorder to maximal order. Evolutionary conservation PROVEAN scores are shown as a heatmap from white to brown from minimal to maximal conservation. Secondary structure predictions are shown with the helix in red and sheet in blue. d, e, Structures are colored based on LOESS regression HbF enrichment scores for (d) human HDAC2 (PDB ID 4LXZ) in complex with competitive inhibitor vorinostat (in green) and (e) human MBD2 in complex with methylated DNA (in yellow) (PDB ID 6CNQ). Left panels depict cartoon models, and right panels depict surface models.
Decoupling CHD4 roles in HbF repression and cellular fitness. CHD4 possessed the highest median HbF enrichment score of the NuRD subunits (Fig. 1b). The HbF enrichment scores of sgRNAs targeting CHD4 were of similar magnitude as those targeting the critical HbF repressors BCL11A and ZBTB7A. However, unlike the other HbF repressing NuRD hit genes, CHD4 also demonstrated negative fitness scores (Fig. 1c). We observed a strong inverse correlation between HbF enrichment scores and fitness scores (Spearman r = −0.784, P < 0.0001), indicating that targeting most positions of CHD4 was associated with a similar magnitude of HbF induction and fitness cost. However, we observed a small group of sgRNAs with high HbF enrichment, yet relatively modest fitness scores within the CHDCT2 domain of CHD4 (Fig. 5a,b). We found the interval from AA1872 to 1883 to be highly significant (P = 2.93 × 10−11) as including outliers from the dataset. Using a sliding window to test all contiguous 12 AA segments within CHD4, we did not find any other significant non-overlapping outlier intervals. We introduced individual sgRNAs within CHD4 CHDCT2 AA1872–AA1883 to test if the resultant mutations could uncouple HbF derepression from negative cellular fitness. We tested two C-terminal targeting sgRNAs around A1873 and P1880 as well as sgRNAs targeting the helicase domain around A742 and a relatively NC, non-domain region of CHD4 around S221. Indeed, the CHDCT2 targeting sgRNAs showed high HbF induction approaching that of the helicase targeting sgRNA and substantially greater than the sgRNA targeting the NC region (Fig. 5c–e). The sgRNAs targeting the CHD4 helicase domain showed substantial cellular toxicity, with almost no viable cells observed after 5 d of culture. In contrast, the CHDCT2 targeting sgRNAs had minimal impact on cellular expansion throughout 14 d in culture, similar to the non-targeting on CHD4 NC targeting sgRNAs (Fig. 5d).

We evaluated the effect of CHD4 CHDCT2 mutation on HbF repression in primary erythroid cells. We electroporated CD34+ HSPCs with SpCas9 protein and sgRNA ribonucleoprotein complex (RNP). We performed 18-d erythroid differentiation culture and observed similar terminal erythroid maturation by cell expansion and immunophenotye (Supplementary Fig. 6a). We compared gene editing efficiency at three time points, at the beginning, middle and end of erythroid differentiation culture (days 4, 10 and 18). We evaluated the indels by deep sequencing of amplicons. We found that frameshift mutations were lost during the course of erythropoiesis (CHD4 frameshift allele frequency from 72.5% to 26.2% from day 4 to day 18) while in-frame deletions were gained (from 19.0% to 63.5%) (Fig. 5g). These results indicated that in-frame mutations at CHD4 CHDCT2 around A1873 selectively escaped negative fitness in primary erythroid precursors, similar to findings in HUDEP-2 immortalized erythroid precursors. Next we evaluated globin gene expression at the end of erythroid differentiation culture. In comparison with mock edited cells, erythroid cells edited at CHD4 CHDCT2 demonstrated elevation of γ-globin from 2.1% to 44.8% of total β-like globin, HbF+ cells from 28.4% to 67.7% and HbF from 4.9% to 16.0% by HPLC (Fig. 5b and Supplementary Fig. 6e,f). These results validate the comprehensive mutagenesis screen and demonstrate that in-frame mutations at AA1872–1883 of CHD4 induce HbF with minimal impact on cell fitness in human primary erythroid cells. In addition, we generated mice homozygous for in-frame deletion at Chd4 A1876 (orthologous to human CHD4 A1873) and observed impaired developmental silencing of transgenic human γ-globin (Fig. 5i, Supplementary Fig. 7 and Supplementary Note).

Rational targeting of CHD4. We conducted biochemical studies to investigate the mechanism whereby in-frame deletions at CHD4 CHDCT2 impact the NuRD complex (see Supplementary Note). In each of the four CHD4 CHDCT2 in-frame deletion clones tested, we observed reduced immunoprecipitation of CHD4 despite similar pulldown of MTA2 itself and other NuRD members, including GATAD2A and MBD2 (Fig. 5j). These results suggested a reduced ability of mutant CHD4 to interact with the NuRD complex. To further investigate the impact of CHDCT2 mutant CHD4 on NuRD complex assembly, we performed fractionation of nuclear extracts by glycerol gradient density sedimentation. We observed decreased incorporation of mutant CHD4 to the NuRD complex along with modest reduction of MTA2, GATAD2A, MBD2, HDAC2 and RBBP4 incorporation (Fig. 5k). Taken together, these results suggest that in-frame deletions at CHD4 CHDCT2 domain around AA1872–1883 impair the interaction of CHD4 with NuRD.

A recent study demonstrated that the previously poorly characterized CHDCT2 domain acts to recruit CHD4 to NuRD by binding to GATAD2 factors GATAD2A and GATAD2B. A C-terminal segment of GATAD2B was sufficient to pull down the CHD4 CHDCT2 domain in a rabbit reticulocyte lysate transcription/translation system. Since bridging interactions by endogenous NuRD subunits are minimal under these conditions, the results suggest a direct interaction between CHDCT2 and GATAD2 factors. In addition to the heightened HbF enrichment scores around the N-terminal coiled-coil domain of GATAD2A implicated in binding to MBD2, we observed another cluster of sgRNAs with heightened HbF enrichment scores at GATAD2A between AA335–486, a C-terminal region encompassing a C2C2-type GATA ZF (Fig. 5c). These GATAD2A sequences, including the GATA ZF, are homologous to sequences within the GATAD2B segment shown to bind to CHD4 CHDCT2 (ref. 43).

We hypothesized that this C-terminal region of GATAD2A could contribute to HbF repression by binding to CHD4 CHDCT2. Therefore, we reasoned that ectopic expression of this segment might competitively bind to CHD4 CHDCT2, displace CHD4 from NuRD and mimic CHDCT2 mutations that result in HbF derepression. We expressed several FLAG epitope tagged constructs of GATAD2A in HUDEP-2 cells, including the entire C-terminal segment (AA335–633), a C-terminal segment that includes a truncation of the GATA ZF domain (AA408–633) or a segment encompassing the entire GATA ZF corresponding to the region with heightened HbF enrichment scores (AA335–486) (Fig. 6a).

**Fig. 4 | In-frame deletions disrupt MTA2 function.** a, Structures colored based on LOESS regression HbF enrichment scores for MTA2 and HDAC2. MTA2 BAH domain (PDB ID 1W45, left) and MTA2 ELM2-SANT domain with HDAC2 (PDB ID 5ICN, right) with aligned HbF enrichment scores mapped. Top panels depict surface models for MTA2, and bottom panels depict cartoon models. Dotted lines indicate the positions of residues deleted in the hemizygous in-frame deletion clones at BAH (red), ELM2 (orange) and SANT (green) domains. b, Each dot indicates the fraction of HbF+ cells from hemizygous MTA2 frameshift or in-frame deletion, non-deletion (unedited) or non-targeted (NT) HUDEP-2 clones. Error bars represent mean of three independent experiments and s.e.m. c, Immunoblot (representative of three independent experiments) of MTA2 from nuclear lysates of hemizygous in-frame deletion clones shows some clones lose and others preserve MTA2 levels. d, MTA2 mRNA expression in hemizygous in-frame deletion clones by RT-qPCR. Error bars represent mean and s.d. of three replicates. e, MTA2 level by immunoblot following exposure of HUDEP-2 MTA2±hemizygous in-frame deletion clones to 4 h of MG132 proteasome inhibitor at indicated concentrations. f, Immunoprecipitation of MTA2 followed by immunoblot shows interaction with NuRD subunits (In, input). g, Immunoprecipitation of CHD4 followed by immunoblot shows interaction with NuRD subunits. h, Glycerol gradient density sedimentation analyses on nuclear extracts from control and M4 SANT domain and M7 ELM2 domain hemizygous in-frame deletion HUDEP-2 clones. Each experiment (e-h) was repeated at least three times with similar results.
With both GATAD2A and FLAG antibodies, analysis suggested that protein levels of the truncated GATAD2A were no greater than that of endogenous GATAD2A (Fig. 6b). Overexpression of these constructs in HUDEP-2 cells led to increased expression of γ-globin, with the greatest effect observed in cells expressing the GATAD2A AA335–486 segment encompassing the GATA ZF domain (Fig. 6d). To test CHD4-GATAD2A disruption in primary human erythroid precursors, we transduced and subjected to erythroid differentiation.
Fig. 5 | Targeting CHD4 CHDCT2 uncouples HbF induction from cytotoxicity. a, CHD4 domains. b, Dots indicate HbF enrichment and fitness scores of all 833 sgRNAs targeting CHD4. The 95% confidence and prediction bands after non-linear least squares curve fitting are shown. Purple dots show sgRNAs targeting AA1872–1883, the only non-overlapping contiguous 12 AA segment of CHD4 significantly enriched in outliers from prediction band (p = 2.93 × 10⁻¹¹, Fisher’s exact test, two-sided). sgRNA scores are averages of three experiments. c, Fraction HbF+ cells (%) following sgRNA targeting CHD4 NC, helicase, CHDCT2 domain or non-targeting (NT) (n = 2, error bars represent mean and s.d.). d, HUDEP-2 expansion following sgRNA targeting CHD4 S22, A742, A1873, P1880 and non-targeting (average from three experiments). e, Immunoblot from CHD4 hemizygous in-frame deletion clones. Representative of three experiments. f, β- and γ-globin RT–qPCR in CHD4 hemizygous in-frame deletion clones (mean and s.d. from n = 2). g, CD34+ HSPCs electroporated with SpCas9:sgRNA targeting CHDCT2 A1873 with indels quantified by amplicon sequencing after 4-, 10- or 18-d erythroid culture. h, β- and γ-globin expression in human erythroid precursors (mean and s.d. from n = 2). i, β-, ε- and γ-globin expression in fetal livers from E14.5, E15.5 and E16.5 Chd4−/−;β-YAC+/+;β−YAC−/−Chd4−/− Tg-Tg+ MBD2−/− embryos (mean and s.d., sample size per Supplementary Fig. 7d). j, Immunoprecipitation MTA2 followed by NuRD subunit immunoblot. Quantification of CHD4 by immunoprecipitation/input ratio. k, Density sedimentation analyses of CHDCT2 domain hemizygous in-frame deletion HUDEP-2 clone. Each experiment (j, k) was repeated at least three times with similar results.
CD34+ HSPCs with the GATAD2A AA335–486 vector. Enforced GATAD2A expression was lower as compared to HUDEP-2 cells, although relative expression increased under the control of the more potent SFFV enhancer/promoter52 (Fig. 6c).

We observed dose-dependent increases in γ-globin expression with higher expression of GATAD2A 335–486 segment in both HUDEP-2 and CD34+ HSPC derived primary erythroid precursors (Fig. 6d,e). Although we observed impaired cell expansion with extremely high level expression of GATAD2A 335–486 in HUDEP-2 cells as driven by SFFV promoter, we observed only slight impact on expansion of primary human erythroid precursors, consistent with relatively more modest expression of the GATAD2A 335–486 segment (Supplementary Fig. 9a,b). The maturation of primary human erythroid precursors by immunophenotype, enucleation frequency or morphology was unchanged with the GATAD2A 335–486 expression, while γ-globin increased from 10.8% to 33.6% of total β-like globin, F-cells from 22.1% to 37.4% and HbF level from 6.5% to 17.1% (Fig. 6e,f and Supplementary Fig. 9b,c). These data suggest that expression of a GATAD2A segment could selectively counteract HbF silencing with minimal impact on the proliferation...
or maturation of erythroid cells. We observed similar γ-globin and HbF induction upon expression of the GATAD2A 335–486 segment in primary human erythroid precursors from a patient with SCD (Fig. 6g,h).

To assess potential interaction of truncated GATAD2A with CHD4, we performed immunoprecipitation. Pulldown of GATAD2A 335–486 with anti-FLAG enriched CHD4 but not endogenous GATAD2A or MB2D, suggested that this GATAD2A segment sequesters CHD4 from NuRD. Reciprocal pulldown with CHD4 antibody enriched the FLAG-tagged GATAD2A AA335–486 segment (Fig. 6i and Supplementary Fig. 9d). To evaluate the NuRD complex, we performed glycerol gradient sedimentation in cells expressing GATAD2A AA335–486. The truncated GATAD2A segment co-sedimented with low molecular weight forms of CHD4 but not the NuRD complex. We observed increased abundance of low molecular weight forms of CHD4 in the presence of the GATAD2A AA335–486 segment (Fig. 6) and Supplementary Fig. 9e). In contrast the distribution of other endogenous members of NuRD, such as GATAD2A and MB2D, did not shift, indicating that GATAD2A 335–486 sequesters CHD4 from NuRD, leaving NuRD otherwise intact. Together, these data suggest that rational targeting of the CHD4 CHDCT2–GATAD2A interaction may disrupt CHD4 recruitment to NuRD and elevate HbF levels while sparing cytotoxicity (Fig. 6k).

Discussion

Using comprehensive CRISPR mutagenesis to reveal the NuRD coding sequences required for HbF repression, we found that a NuRD subcomplex including just one functional paralog per protein family is required. These results suggest that the HbF repressive functional NuRD complex is non-redundant and that targeting key subunits might provide biological specificity. We performed both affinity purification and proximity labeling proteomics of the MTA2 and CHD4 NuRD subcomplexes. The results were highly congruent with the CRISPR mutagenesis, identifying GATAD2A, HDAC2, MB2D and RBBP4 to be members of the functional subcomplex along with MTA2 and CHD4. Future experiments will determine how much of the HbF repressive effect depends on direct recruitment of the NuRD subcomplex to the γ-globin gene. Over the past decade specific DNA binding transcriptional repres-
sors such as BCL11A and ZBTB7A have been identified as crucial for HbF silencing. These discoveries have suggested novel gene ther-
apies based on viral transduction or gene editing of mobilized HSCs, manipulated ex vivo and then autologously re-engrafted80. However, given their complexity and cost, these gene therapies are unlikely to scale to the massive global unmet need for β-hemoglobinopathy therapeutics. Small molecule approaches are imperative, yet rational targeting of gene regulation remains challenging. By performing dense mutagenesis in situ we aimed to nominate specific NuRD positions as possible therapeutic targets while avoiding excess cellular toxicity. The results of this screen suggest numerous vulnerable regions within the HbF repressive NuRD subcomplex. For example, at MTA2 we find that disturbing protein structure at some positions destabilizes the protein and at other positions leads to impaired function. Either approach (destabilizing or competitive binding) might be phenocopied by appropriate rational design.

At CHD4, we identified in-frame deletions at the CHDCT2 domain uncoupling HbF repression from cellular toxicity. The CHDCT2 domain of CHD4 has recently been shown to recruit CHD4 to NuRD by binding to GATAD2A. We speculate that the NuRD independent effects of CHD4 may be less reliant on these sequences. We demonstrate that a fragment of GATAD2A able to bind to CHD4 sequesters CHD4 from NuRD, and leads to HbF derepression with minimal toxicity. Improved biochemical and structural understanding of the GATAD2A–CHD4 interaction may afford opportunities for small molecule targeting of this complex. More broadly this study demonstrates how dense mutagenesis correl-
ating protein sequences with function offers a means to achieve specificity when targeting ubiquitous chromatin regulatory complexes (see Supplementary Note). As the resolution and thoroughness of genome editing perturbation range advance, comprehensive mutagenesis of protein complexes in situ could be readily adapted to the study of many disease-relevant cellular processes.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of code and data availability and associated accession codes are available at https://doi.org/10.1038/s41588-019-0453-4.

Received: 4 November 2018; Accepted: 21 May 2019; Published online: 28 June 2019

References

1. Modell, B. & Darlison, M. Global epidemiology of haemoglobin disorders and derived service indicators. Bull. World Health Organ 86, 480–487 (2008).
2. Piel, F. B. et al. Global epidemiology of sickle haemoglobin in neonates: a contemporary geostatistical model-based map and population estimates. Lancet 381, 142–151 (2013).
3. Darling, R. C., Smith, C. A., Aasmussen, E. & Cohen, F. M. Some properties of human fetal and maternal blood. J. Clin. Invest. 20, 739–747 (1941).
4. Schroeder, W. A., Shelton, J. R., Shelton, J. B. & Cormick, J. The amino acid sequence of the alpha chain of human fetal hemoglobin. Biochemistry 2, 1353–1357 (1963).
5. Blit, C. A. & Stamatoyannopoulos, G. Hemoglobin switching and its clinical implications. Curr. Opin. Hematol. 1, 136–142 (1994).
6. Vinjamur, D. S., Bauer, D. E. & Orkin, S. H. Recent progress in understanding and manipulating haemoglobin switching for the haemoglobinopathies. Br. J. Haematol. 180, 630–643 (2017).
7. Bauer, D. E., Brendel, C. & FitzHugh, C. D. Curative approaches for sickle cell disease: a review of allogeneic and autologous strategies. Blood Cells Mol. Dis. 67, 155–168 (2017).
8. Nuinoon, M. et al. A genome-wide association identified the common genetic variants influence disease severity in beta-thalassemia/hemoglobin E. Hum. Genet. 127, 303–314 (2010).
9. Lettre, G. et al. DNA polymorphisms at the BCL11A, HBS1L-MYB, and p66Alpha-MBD2 coiled-coil interaction and implications into the NuRD chromatin remodeling complex. Cell Mol. Life Sci. 72, 2391–2395 (2015).
10. Millard, C. J. et al. The structure of the core NuRD repression complex provides insights into its interaction with chromatin. Cell 5, e13941 (2016).
11. Kransdorf, E. P. et al. MB2D is a critical component of a methyl cytosine-binding protein complex isolated from primary erythroid cells. Blood 108, 2241–2245 (2006).
12. Harju-Baker, S., Costa, F. C., Fedosyuk, H., Neades, R. & Peterson, K. R. Silencing of a gamma-globin gene expression during adult definitive erythropoiesis mediated by GATA-1-FOG-1-Mi2 complex binding at the ~566 GATA site. Mol. Cell Biol. 28, 3101–3113 (2008).
13. Rupon, J. W., Wang, S. Z., Gaensler, K., Lloyd, J. & Ginder, D. Methyl binding domain protein 2 mediates gamma-globin gene silencing in adult human betaYAC transgenic mice. Proc. Natl Acad. Sci. USA 103, 6617–6622 (2006).
14. Gnanapragasam, M. N. et al. p66Alpha–MBD2 coiled-coil interaction and recruitment of Mi-2 are critical for globin gene silencing by the MBD2–NuRD complex. Proc. Natl Acad. Sci. USA 108, 7487–7492 (2011).
15. Xu, J. et al. Corepressor-dependent silencing of fetal hemoglobin expression by BCL11A. Proc. Natl Acad. Sci. USA 110, 6518–6523 (2013).
16. Costa, F. C., Fedosyuk, H., Chazelle, A. M., Neades, R. Y. & Peterson, K. R. Mib2beta is required for gamma-globin gene silencing: temporal assembly of a GATA-1-FOG-1-Mi2 repressor complex in beta-YAC transgenic mice. PLoS Gen. 8, e1003155 (2012).
17. Amaya, M. et al. Mib2beta-mediated silencing of the fetal gamma-globin gene in adult erythroid cells. Blood 121, 3493–3501 (2013).
18. Bradner, J. E. et al. Chemical genetic strategy identifies histone deacetylase 1 (HDAC1) and HDAC2 as therapeutic targets in sickle cell disease. Proc. Natl Acad. Sci. USA 107, 26167–26172 (2010).
19. Esrick, E. B., McConkey, M., Lin, K., Frisbee, A. & Ebert, B. L. Inactivation of HDAC1 or HDAC2 induces gamma globin expression without altering cell cycle or proliferation. Am. J. Hematol. 90, 624–628 (2015).
21. Shearstone, J. R. et al. Chemical inhibition of histone deacetlylases 1 and 2 induces fetal hemoglobin through activation of GATA2. PLoS ONE 11, e0153767 (2016).
22. Uda, M. et al. Genome-wide association study shows BCL11A associated with persistent fetal hemoglobin and amelioration of the phenotype of beta-thalassemia. Proc. Natl Acad. Sci. USA 105, 1620–1625 (2008).
23. Menzel, S. et al. A QTL influencing F cell production maps to a gene encoding a zinc-finger protein on chromosome 2p13. Nat. Genet. 39, 1197–1199 (2007).
24. Bauer, D. E. et al. An erythroid enhancer of BCL11A subject to genetic variation determines fetal hemoglobin level. Science 342, 253–257 (2013).
25. Liu, N. et al. Direct promoter repression by BCL11A controls the fetal to adult hemoglobin switch. Cell 173, 430–442 e17 (2018).
26. Sankaran, V. G. et al. Human fetal hemoglobin expression is regulated by the developmental stage-specific repressor BCL11A. Science 322, 1839–1842 (2008).
27. Thein, S. L. et al. Intergenic variants of HBS1L-MYB are responsible for a major quantitative trait locus on chromosome 6q23 influencing fetal hemoglobin levels in adults. Proc. Natl Acad. Sci. USA 104, 11346–11351 (2007).
28. Sankaran, V. G. et al. Developmental and species-divergent globin switching are driven by BCL11A. Nature 460, 1093–1097 (2009).
29. Sankaran, V. G. et al. A functional element necessary for fetal hemoglobin silencing. N. Engl. J. Med. 365, 807–814 (2011).
30. Xu, J. et al. Transcriptional silencing of [gamma]-globin by BCL11A involves long-range interactions and cooperation with SOX6. Genes Dev. 24, 783–798 (2010).
31. Yi, Z. et al. Sox6 directly silences epsilon globin expression in definitive erythropoiesis. PLoS Genet. 2, e14 (2006).
32. Xu, J. et al. Correction of sickle cell disease in adult mice by interference with fetal hemoglobin silencing. Science 334, 993–996 (2011).
33. Masuda, T. et al. Transcription factors LRF and BCL11A independently repress expression of fetal hemoglobin. Science 351, 285–289 (2016).
34. Schoonenberg, V. A. C. et al. CRISPRo: identification of functional protein coding sequences based on genome editing dense mutagenesis. Genome Biol. 19, 169 (2018).
35. Aguirre, A. J. et al. Genomic copy number dictates a gene-independent cell response to CRISPR/Cas9 targeting. Cancer Discov. 6, 914–929 (2016).
36. Munoz, D. M. et al. CRISPR screens provide a comprehensive assessment of cancer vulnerabilities but generate false-positive hits for highly amplified genomic regions. Cancer Discov. 6, 900–913 (2016).
37. Morgens, D. W., Deans, R. M., Li, A. & Bassik, M. C. Systematic comparison of CRISPR/Cas9 and RNAi screens for essential genes. Cell 169, 169–180 (2017).
38. Esrick, E. B. & Bauer, D. E. Genetic therapies for sickle cell disease. Nat. Rev. Genet. 18, 473–487 (2017).
39. Ding, Z., Gillespie, L. L. & Paterno, G. D. Human M4-ER1 alpha and beta function as transcriptional repressors by recruitment of histone deacetylase 1 to their conserved ELM2 domain. Mol. Cell Biol. 23, 250–258 (2003).
40. Wang, X. et al. SMARCB1-mediated SWI/SNF complex function is essential for enhancer regulation. Nat. Genet. 49, 289–295 (2017).
41. Low, J. K. et al. CHD4 is a peripheral component of the nucleosome remodeling and deacetylase complex. J. Biol. Chem. 291, 15853–15866 (2016).
42. Demaison, C. et al. High-level transduction and gene expression in hematopoietic repopulating cells using a human immunodeficiency [correction of immunodeficiency] virus type 1-based lentiviral vector containing an internal spleen focus forming virus promoter. Hum. Gene Ther. 13, 803–813 (2002).
43. Erick, E. B. & Bauer, D. E. Genetic therapies for sickle cell disease. Semin. Hematol. 55, 76–86 (2018).
Methods
HUDEP-2 cell culture. HUDEP-2 cells were cultured as described previously (also see Supplementary Note).

Human primary erythroid cell culture. Human CD34+ HSPCs from mobilized peripheral blood of deidentified healthy donors were obtained from Fred Hutchinson Cancer Research Center, Seattle, Washington. CD34+ HSPCs were expanded and differentiated as described previously. SpCas9 and sgRNA were delivered by electroporation of ribonuclease complex (RNPs) (see Supplementary Note, Supplementary Methods).

Design and synthesis of human NuRD lentiviral sgRNA libraries. Every 20-mer sequence upstream of NGG PAM on sense and antisense strands was identified for human NuRD gene coding regions (Fig. 1a). The sgRNA oligos were synthesized as previously described, and were amplified using PCR. A second PCR was performed to remove barcodes and insert vector homology. Subsequently, oligos were cloned into LentiGuide-Puro (Addgene plasmid 52963) using a Gibson assembly. Sufficient colonies were isolated for about 1,800 colonies per individual sgRNA within the library. The plasmid library was deep sequenced to confirm the representation of sgRNAs. To produce lentiviruses, HEK293T cells were cultured with DMEM supplemented with 10% fetal bovine serum (FBS) and 2% penicillin-streptomycin in 15 cm tissue culture dishes. HEK293T cells were transfected at 70–80% confluence in 14 ml of media using 10 μg ml−1 of lipofectamin (Invitrogen). Cells were centrifuged for 5 min at 4 °C and then resuspended in ice cold lysis buffer (5 million cells/ml buffer) and passed through 26 g syringe five times. Lysates were centrifuged at 4 °C and 800 g for 5 min. Subsequently, pellet (nuclei) was resuspended in 1 ml of immunoprecipitation buffer (150 mM NaCl, 1 mM EDTA, 1 mM EGTA, 1% Triton X-100, 2.5 mM sodium pyrophosphate, 1 mM β-glycerophosphate, 1 mM Na3VO4, and 1 mM Pefabloc (Roche Applied Science)) and placed on ice for 30 min, sonicated using Bioruptor (Diagenode). Lysates were centrifuged at 16,000 g for 5 min. Supernatants were incubated with 5–10 μg of the antibodies overnight. Dynabeads protein G were used to collect target protein-antibody immune complexes. Bound proteins were washed and then eluted by Laemmli Sample Buffer freshly supplemented with 1 mM dithiothreitol (DTT; BioRad).

Immunoblot analysis. Protein expression, immunoprecipitation efficiency of NuRD interacting partners were analyzed using SDS–PAGE immunoblotting using 4–20% TGX Prestack Protein Gels (BioRad) with standard procedures. Details of the antibodies used are provided in Supplementary Table 4.

Liquid chromatography nanoelectrospray ionization mass spectrometry. Immunoprecipitation products eluted by Laemmli Sample Buffer were resolved by SDS–PAGE and visualized by silver staining. Gel slices were excised and destained using 30 mM potassium hexa-cyanoferrate (III) /100 mM sodium thiostiluate. Cysteine residues were reduced (10 mM DTT; 56 °C, 60 min) and alkylated (55 mM iodoacetamide, 25 °C, 45 min) before in-gel overnight proteolysis at 37 °C using 575 ng sequencing grade modified porcine trypsin (Promega) per gel slice, in 30 mM ammonium bicarbonate pH 8.4. Peptides were eluted from gel slabs by 70% acidified acetonitrile and lyophilized using a vacuum centrifuge. Peptide pellets were resuspended in 0.1% aqueous formic acid and 15% of the solution was analyzed. The liquid chromatography system (Ekspect NanoLC 425, Eksigent) consisted of a vented trap-elute architecture coupled to an Orbitrap Fusion mass spectrometer (Thermo) via a nanoelectrospray ion source (New Objective). Peptides were resolved using a 75 μm i.d. × 750 μm internal diameter column, packed with ReproSil-Pur C18-AQ 1.9 μm particles (Dr. Maisch) over 90 min using a 5–40% acetonitrile gradient in water at 250 nL min−1. The protein abundance was quantified based on LFQ intensity (sum of the integrated area of the extracted ion chromatogram for each peptide assignment) (see Supplementary Note, Supplementary Methods).

HbF expression and globin gene expression. For HbF expression 1 × 106 HUDEP-2 cells were stained and analyzed by intracellular flow cytometry as described previously. HbF expression was also measured by HPLC using a D-10 hemoglobin analyzer (BioRad). Globin gene expression was assessed using qPCR with reverse transcription (RT–qPCR; Supplementary Note and Supplementary Table 3).

Cellular expansion. HUDEP-2 cells were seeded in triplicate at 100,000 cells per well in 6-well plates and cultured in the presence of blastcandin and puromycin as appropriate. Viable cells were counted using Trypan Blue exclusion by Countess II FL Automated Cell Counter (ThermoFisher).
Immunocytochemistry and quantification of DNA damage biomarkers hYHAX and 53BP1. Cells were selected with puromycin 24h after transfection. After an additional 48h of culture, cells were attached on poly-L-lysine coated glass cover slides for 20 min. Subsequently, cells were fixed with 4% paraformaldehyde and immunostained for DNA damage repair markers hYHAX and 53BP1. Antibody labeling was visualized using Alexa Fluor secondary antibodies. To visualize nuclei, cells were counterstained with DAPI. Quantitative analysis of hYHAX foci was performed by using Zeiss (AxioSkope2) fluorescent microscope equipped with a Leica DFC300FX camera and the Leica microscopy LAS program. Pictures at ×20 and ×40 magnification were made of five areas in the culture dish. Mean fluorescent intensity of foci in each cell in the image were measured using ImageJ software. Similarly, pyknotic nuclei were counted using DAPI stain and particle counter plugin of ImageJ software.

Etoposide sensitivity assay. HUDEP-2 cell hemizygous CHD4 clones with mutations in CHDCTC2 region and control were seeded at 0.5 × 10⁴ in 24-well plates and treated with either DMSO or etoposide (100, 300, 1,000 nM) for 24h. At the end of the treatment period, cells were stained with Annexin V and DAPI, and the cell viability was measured using flow cytometry.

Cloning and expression of BioID2 in HUDEP-2 cells. Humanized BioID2 sequence together with GGGSG13 linker was amplified from Addgene plasmid ID no. 74224 (ref. 35) by introducing the BsgI and BamHI cloning sites. The fragment was cloned in lentiviral vector backbone (Addgene plasmid no. 52962) carrying dCas9 (dCas9 sequence from Addgene plasmid ID no. 74284) sequence between BsgI and BamHI cloning sites. Together plasmid was initially reconstructed as lentiviral vector backbone (Addgene plasmid ID no. 74284) sequenced by Addgene; BsgI and BamHI cloning sites were then ligated into BamHI/BsiWI digested pLVX-puro (Addgene plasmid ID no. 74988) sequenced by Addgene. To produce lentivirus HEK293T cells were transfected with 13.3 μg VSV-G and 200 ng pCMV-BG-gene expression plasmid (Invivogen, Sollentuna, Sweden) per 10 cm dish for 3 days. Supernatant was collected after 72 h post-transfection and subsequently concentrated by ultracentrifugation. 

RNA sequencing. SpCas9 expressing HUDEP-2 cells in expansion media were transduced in parallel with sgRNAs targeting CHD4 and MTAT2 and controls. Then 24h following transduction, media was replaced with fresh expansion media containing 1 μg/ml puromycin. After 48h of transduction, a small fraction of puromycin selected cells was used to extract genomic DNA. PCR was performed using extracted genomic DNA around the sgRNA cut site, PCR amplified product was purified using DNA purification kit and used for Sanger sequencing and subsequent TIDE analysis. Each sgRNAs indel formation efficiency was measured more than 80% by TIDE. Cells were lysed 4d after transduction and RNA was extracted using RNeasy Mini Kit (Qiagen). Libraries were synthesized using Illumina TruSeq Stranded mRNA sample preparation kits from 500 ng of purified total RNA according to the manufacturer's protocol. The final dsDNA libraries were sequenced on Illumina NextSeq500 with single-end 75 base pair (bp) reads by the Dana-Farber Cancer Institute Molecular Biology Core Facilities (also see Supplementary Note, Supplementary Methods).

Generation of Chd4−/−/αα mice. Animal experiments were performed under protocols approved by the Boston Children’s Hospital Animal Care and Use Committee (15–12–3065). To generate Chd4−/−/αα mice, a sgRNA targeting the terminal domain on Chd4 was designed and synthesized. Two overlapping oligonucleotides carrying the T7 RNA polymerase promoter, the target sequence and the sgRNA scaffold were annealed and used as template for in vitro transcription (MEGAscript T7 Transcription Kit, Ambion). In vitro transcribed RNA was purified using NucAway Spin Columns (Thermo). One single strand donor template DNA including a 3′-deletion and a silent mutation to prevent re-cleavage of edited alleles was obtained from IDT (PAGE Ultragene, 60bp).

Protein complex gradient sedimentation. Glycerol sedimentation assay was performed according to Wang et al. with modifications. Briefly, pellets of 40 million HUDEP-2 cells were lysed on ice for 10 min in Buffer A (10 mM Tris-Cl pH 7.4, 10 mM NaCl, 3 mM MgCl₂, 0.1 mM EDTA, 0.5% NP-40, 1 mM DTT and 1× complete protease inhibitor cocktail (Roche)). Nuclei were sedimented by centrifugation at 1,000g for 5 min at 4 °C. Nuclei were then resuspended in Buffer B (10 mM HEPS pH 7.6, 3 mM MgCl₂, 100 mM KCl, 0.1 mM EDTA, 10% glycerol, 1 mM DTT and 1× complete protease inhibitor cocktail) and were lysed by addition of ammonium sulfate to a final concentration of 0.3 M for 30 min on ice. Soluble nuclear proteins were collected from supernatant after ultracentrifugation at 40,000 × g for 10 min at 4 °C (SW 41-T, swinging bucket) and precipitated with 0.3 M ammonium sulfate for 30 min on ice. Protein precipitate was pelleted by centrifugation at 17,000g for 20 min at 4 °C and resuspended in HEMEG buffer (25 mM HEPS pH 7.9, 0.1 mM EDTA, 12.5 mM MgCl₂, 100 mM KCl, 1 mM DTT and 1× complete protease inhibitor cocktail). Resuspended nuclear extract (1 mg in 500 µl) were mixed with appropriate amounts of HEMEG buffer and overlaid onto a 12-ml linear MGE gradient (prepared in buffer containing 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1% Igepal CA-630, 0.5% Na-deoxycholate, 0.1% SDS, 1 mM DTT) prepared in 14 ml 14 × 95 mm polycarbonate centrifuge tube (Beckman Coulter, 331374). Tubes were centrifuged at 40,000 × g, 4 °C for 16h in a SW 40 Ti swinging bucket rotor. Fractions (0.5 ml) were collected from top to bottom of the tube and were analyzed using 26-well 4–20% Criterion TGX gels (BioRad) and subsequent immunoblotting.

Expression of GATA2A constructs. Human GATA2A ORF was purchased from Gentools (clone ID no. OHu17601D, RefSeq no. NM_017660.3) and was PCRamplified and cloned between EcoRI and BamHI restriction sites of a pLVX-Puro vector (Clontech) using Gibson assembly with primers provided in Supplementary Table 3. To generate SFFV promoter driven GATA2A 335–486 construct, CMV promoter of pLVX-puro was replaced with SFFV promoter (amplified from pH8-SFFV-dCas9-BFP-KRAB, Addgene no. 46911). Briefly, pLVX-puro GATA2A 335–486 was digested with Clal and EcoRI, and the large vector fragment was gel purified for further Gibson assembly with PCR amplified SFFV promoter. The sequence integrity of all the cloned products were verified by Sanger sequencing. All the cloned GATA2A constructs contain a SV40 NLS (PKKKRKV) and a FLAG-tag (DYKDDDDK) at the N terminus. Wild-type HUDEP-2 cells were transduced with lentivirus carrying GATA2A constructs for 24h, followed by selection with puromycin (1 μg/ml).

References
54. Kurita, R. et al. Establishment of immortalized human erythroid progenitor cell lines able to produce enucleated red blood cells. PLoS One 8, e59890 (2013).
55. Canver, M. C. et al. BCL11A enhancer dissection by Cas9-mediated in vivo screen, we used a custom computational pipeline available at https://gitlab.com/bauerlab/crispro.
56. Wu, Y. et al. Highly efficient therapeutic gene editing of human hematopoietic stem cells. Nat. Med. 25, 776–783 (2019).
57. Shalem, O. et al. Genome-scale CRISPR-Cas9 knockout screening in human cells. Science 343, 84–87 (2014).
58. Sanjana, N. E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. Nat. Methods 11, 783–784 (2014).
59. Chen, S. et al. Genome-wide CRISPR screen in a mouse model of tumor growth and metastasis. *Cell* **160**, 1246–1260 (2015).
60. Doench, J. G. et al. Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat. Biotechnol.* **32**, 1262–1267 (2014).
61. Love, M. L., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
62. Cock, P. J. et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**, 1422–1423 (2009).
63. Cifani, P. & Kentsis, A. High sensitivity quantitative proteomics using automated multidimensional nano-flow chromatography and accumulated ion monitoring on quadrupole-orbitrap-linear ion trap mass spectrometer. *Mol. Cell Proteom.* **16**, 2006–2016 (2017).
64. Cox, J. et al. Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell Proteom.* **13**, 2513–2526 (2014).
65. Kim, D. I. et al. An improved smaller biotin ligase for BioID proximity labeling. *Mol. Biol. Cell* **27**, 1188–1196 (2016).
66. Seruggia, D., Fernandez, A., Cantero, M., Pelczar, P. & Montoliu, L. Functional validation of mouse tyrosinase non-coding regulatory DNA elements by CRISPR-Cas9-mediated mutagenesis. *Nucleic Acids Res.* **43**, 4855–4867 (2015).
67. Harms, D. W. et al. Mouse genome editing using the CRISPR/Cas system. *Curr. Protoc. Hum. Genet.* **83**, 1–27 (2014).
68. Pinello, L. et al. Analyzing CRISPR genome-editing experiments with CRISPResso. *Nat. Biotechnol.* **34**, 695–697 (2016).
Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
- Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted. Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen’s d, Pearson’s r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

- Data collection: MS/MS data were processed using MaxQuant Version 1.6.0.16. Flow cytometer (cell sorting) data was collected using FACSDiva Version 6.1.3. The sgRNA abundance in pooled CRISPR screen data was measured using DESeq2.

- Data analysis: For analysis and visualization of functional readout from tiled pooled CRISPR Cas9 screen, we used a custom computational pipeline available at https://gitlab.com/bauerlab/crispro. CRISPRO is based on R packages, Biopython pairwise2 module and PyMOL. The editing outcomes of individual sgRNAs were analyzed using open source algorithms TIDE, ICE and CRISPResso. RNA-seq data were analyzed using open source software Tuxedo suite: Bowtie, TopHat, Cufflinks and Cuffdiff. For FACS analysis FlowJo software package was used.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Pooled CRISPR Cas9 screen next generation sequencing (NGS) data, and RNA-Seq data have been deposited in NCBI sequence read archive (SRA) under accession number PRJNA496556. Mass spectrometer data sets have been deposited in ProteomeXchange repository under the accession number PXD009793.
Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences  ✔ Behavioural & social sciences  ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

| Sample size | For pooled CRISPR Cas9 screen, three independent biological replicates were performed. For cellular and molecular assays, the sample sizes were chosen based on expected phenotypes and previous experience with assay variability and were completed at least in triplicate biological replicates unless stated otherwise in figure legends. |
| Data exclusions | No data exclusion |
| Replication | Yes, reliably reproduced |
| Randomization | No randomization |
| Blinding | Assessment of DNA damage (e.g supplementary figure 7c-d, f) was done by blinded observer. |

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

| Materials & experimental systems | Methods |
| n/a | Involved in the study |
| ☒ | Antibodies |
| ☒ | Eukaryotic cell lines |
| ☒ | Palaeontology |
| ☒ | Animals and other organisms |
| × | Human research participants |
| × | Clinical data |
| ☒ | ChIP-seq |
| ☒ | Flow cytometry |
| ☒ | MRI-based neuroimaging |

Antibodies

Antibodies used

- Anti BRG-1, Santa Cruz cat no. sc-17796, (WB: 1:800)
- Anti CHD4 (D8B12) Rabbit mAb, Cell Signaling Technology cat no. #11912, (WB: 1:2000, IP: 1:200-500)
- Anti GAPDH (D16H11) XP® Rabbit mAb, Cell Signaling Technology cat no. 5174S, (WB: 1:5000)
- Anti GATA2A , Bethyl Laboratories cat no. A302-358A, (WB: 1:2000)
- Anti HbF-APC, Life Technologies cat no. MHFH05, (FACS; 2ul: 1million cells)
- Anti HDAC2 , Abcam cat no. ab32117-100 (WB: 1:2000)
- Anti Histone H3 , Abcam cat no. ab1791, (WB: 1:2000)
- Anti MB2 , Bethyl Laboratories cat no. A301-632A (WB: 1:2000)
- Anti MTA1 (D17G10) Rabbit mAb, Cell Signaling Technology cat no. s6465S, (WB: 1:2000)
- Anti ZBTB7A , eBioscience cat no. 14-3309-80 (WB: 1:3000)
- Goat anti-Armenian hamster IgG-HRP, Santa Cruz cat no. sc-2443 (WB: 1:3000)
- Goat anti-Mouse Secondary Antibody, Alexa Fluor Plus 488, Thermofisher cat no. A32723 (ICC: 1:500)

- Anti-mouse IgG, HRP-linked Antibody, Cell Signaling Technology cat no. 7076s, (WB: 1:3000)
- Anti-MTA2 antibody, Abcam cat no. ab8106 (WB: 1:2000)
- Anti-phospho-Histone H2A.X (Ser139), clone JBW301, Millipore/Upstatecat no. 05-636 (ICC: 1:500)
- Anti-rabbit IgG, HRP-linked Antibody, Cell Signaling Technology cat no. 7076s, (WB: 1:3000)
- Anti Hemoglobin β (37-8) FITC, Santa Cruz cat. no. sc-21757 FITC (FACS: 5ul:1million cells)
- Anti Fetal Hemoglobin FITC, Thermo Fisher cat no. MHFH01-4 (FACS; 1ul: 1million cells)
- Anti Histone H3, Cell Signaling Technology cat no. #9713, (WB: 1:2000)
- Anti-Histone H4, Cell Signaling Technology cat no. #9714, (WB: 1:2000)
- Anti-53BP1 Antibody, clone BP13, Millipore cat no. MA3802, (ICC: 1:5000)
- Anti-ChD3 antibody [EPNCIR110A], Abcam cat no. ab109195, (WB: 1:2000)
- Anti-ChD4 antibody [3F2/4] - ChIP Grade, Abcam cat no. ab70469, (WB: 1:2000, IP: 1:200-500)
- Anti-CTIP1 (BCL11A) antibody [14B5], Abcam cat no. ab19487 (WB: 1:2000)
- Anti-Histone H3, Cell Signaling Technology cat no. #9713, (WB: 1:2000)
- Anti-MT2A Antibody, Abcam cat no. ab18057 (WB: 1:2000)
- Anti-phospho-Histone H2A.X (Ser139), clone JBW301, Millipore cat no. 05-636 (ICC: 1:500)
- Anti-rabbit IgG, HRP-linked Antibody, Cell Signaling Technology cat no. 7076s, (WB: 1:2000)
Eukaryotic cell lines

Cell line source(s)

- 1) Human Umbilical Cord Blood-Derived Erythroid Progenitor 2 (HUDEP-2) cells were obtained from R. Kurita and Y. Nakamura (Kurita et al., 2013) RIKEN Cell Engineering Division, RIKEN BioResource Center, Comprehensive Human Sciences, University of Tsukuba, Tsukuba, Ibaraki 305-0074, Japan.
- 2) Deidentified CD34+ hematopoietic stem and progenitor cells were obtained from the Center for Excellence in Molecular Hematology at the Fred Hutchinson Cancer Research Center (Seattle, Washington). These studies with anonymous, deidentified samples were conducted with IRB exemption by the Boston Children’s Hospital IRB.

Authentication

HUDEP-2 cells were authenticated by the analysis of expression of human globin genes, karyotyping and STR (small tandem repeat) profiling. Purity of the human deidentified CD34+ hematopoietic stem and progenitor cells was guaranteed by the supplier, and was validated in-house through FACS analysis.

Mycoplasma contamination

Cell lines were routinely tested (negative) for mycoplasma contamination.

Commonly misidentified lines

None

Animals and other organisms

Laboratory animals Mus musculus, C57BL/6NCrl x YACTg. The human beta globin locus introduced by YAC transfer exhibits a specific and reproducible pattern of developmental regulation in transgenic mice. Age: we used 14.5, 15.5 and 16.5 dpc (days post coitum) embryos.

Wild animals No animals from wild

Field-collected samples No field sampling

Ethics oversight Boston Children’s Hospital

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Flow Cytometry

Plots

Confirm that:
- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a ‘group’ is an analysis of identical markers).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation

HUDEP-2 cells, human deidentified CD34+ hematopoietic stem and progenitor cells and mouse material. The detailed sample preparation is provided in the text. Briefly, cells were fixed with 0.05% glutaraldehyde in PBS, incubated at room temperature for 10 min and centrifuged at 600 x g for 5 min. Cells were then permeabilized with 0.1% Triton, 0.1% BSA in PBS, for 5 min at room temperature, and centrifuged at 600 x g for 15 min. Cells were resuspended in 0.1% BSA in PBS and incubated with antibody (e.g. 1-2 μl of anti-Human Fetal Hemoglobin antibody;ThermoFisher) in the dark at room temperature for 30 min. Subsequently, cells washed with 0.1% BAS in PBS using centrifuge at 600 x g for 10 min and resuspended in 0.1% BSA in PBS before analysis.

Instrument

BD LSRII flow cytometer

Software

FACSDiva Version 6.1.3 and Flowjo

Cell population abundance

Fetal Hemoglobin (HbF) positive cells, top 10% population was sorted.
In total four gates were drawn in the following order. First gate: FSC-A (x-axis) vs. SSC-A (y-axis) for viability of cells. Second gate (remove doublets): FSC-A (x-axis) vs. FSC-H (y-axis). Third gate (remove doublets): SSC-W (x-axis) vs. SSC-H (y-axis). Fourth gate (population of interest): FSC-A (x-axis) vs. FITC/APC (y-axis). Appropriate negative and positive controls for each antibody were used. We have included exemplifying supplemental data.

Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.