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

Falak Sher[1,2,11], Mir Hossain[1,11], Davide Seruggia[1,11], Vivien A. C. Schoonenberg[1,3], Qiuming Yao[1,4], Paolo Cifani[5], Laura M. K. Dassama[1], Mitchel A. Cole[1], Chunyan Ren[1], Divya S. Vinjamur[1], Claudio Macias-Trevino[1], Kevin Luk[8], Connor McGuckin[1], Patrick G. Schupp[1], Matthew C. Canver[1], Ryo Kurita[6], Yukio Nakamura[7], Yuko Fujiwara[1], Scot A. Wolfe[8], Luca Pinello[4], Takahiro Maeda[10], Alex Kentsis[5], Stuart H. Orkin[1,9], Daniel E. Bauer[1,12]

Supplementary Note

Motivation for dense mutagenesis screen of coding sequences

In general, how widely expressed chromatin regulatory complexes serve highly specific biological functions remains incompletely understood. As a corollary, selective targeting of these complexes to achieve specific biological phenotypes is challenging. We hypothesized that dense mutagenesis by gene editing coupled with phenotypic selection could identify critical endogenous sequences required for specific complex functions. CRISPR-Cas9 is a versatile tool for genome editing and interrogation of gene regulation. Recently, we investigated the functional importance of an intronic erythroid-specific enhancer of *BCL11A* and trait-associated sequences intergenic to *HBS1L-MYB* by using CRISPR-Cas9 mediated dense mutagenesis[1,2]. Here we adapted the tiling pooled screen approach to all the coding sequences of human NuRD complex members. CRISPR-Cas9 induced double strand DNA breaks in absence of homology donors are repaired by non-homologous end joining (NHEJ) which results in a combination of frameshift and in-frame short insertion and deletion (indel) alleles. Frameshift indels typically trigger nonsense mediated decay (NMD) resulting in loss of gene expression or C-terminal truncations or extensions that often impair function. In contrast, if the repaired allele is an in-frame short deletion of non-critical amino acid residues, the resulting mutant protein may retain much of its original function. Accordingly, targeting non-critical protein coding sequences is expected to achieve a blunted phenotype since in a mixed population, some cells will have biallelic frameshifts whereas others have at least one function-preserving in-frame allele. Targeting critical protein coding sequences is expected to achieve a heightened phenotype since cells with either frameshift or in-frame alleles are expected to result in loss of function[3-5]. Therefore this approach may systematically identify the amino acid residues that are critical for protein function.

Pooled CRISPR screen design and performance

Consistent with previous results, nontargeting or neutral locus targeting by SpCas9 had no effect on fetal globin expression during in vitro erythroid maturation (Supplementary Fig. 1a). We constructed a lentiviral CRISPR library saturating all NGG protospacer adjacent motif (PAM) single guide RNAs (sgRNAs) at human NuRD coding sequences, with median distance of 9.3 bp (3.1 AA) between adjacent cleavage sites (Fig. 1a and Supplementary Fig. 1b). We included nontargeting sgRNAs as negative controls and sgRNAs targeting *BCL11A* and *ZBTB7A* coding sequences as positive controls. We transduced HUDEP-2 erythroid precursor cells constitutively expressing SpCas9 with the lentiviral sgRNA library at low multiplicity so that most selected cells received only one sgRNA integrant (Supplementary Fig. 1c). Following 12 days of differentiation, cells were intracellularly stained with HbF antibody and physically sorted (by fluorescence activated cell sorting, FACS) for high HbF expression (Supplementary Fig. 1d). In order to compare the abundance of integrated sgRNAs we deep sequenced the sgRNA cassette amplified from three sources: 1) plasmid library; 2) genomic DNA (gDNA) from total unsorted cells at end of differentiation; and 3) sorted high HbF expressing cells. We calculated two scores for each sgRNA. First, an HbF enrichment score was based on the abundance of

sgRNA in the HbF high as compared to unsorted cells. Second, a fitness score was based on the abundance of the sgRNA in the unsorted population of cells as compared to the library.

HbF screen results

RBBP4 was required for cellular fitness but not for HbF enrichment. Since RBBP4 is known to participate in additional chromatin associated complexes besides NuRD, such as PRC2 and SIN3A complexes[6,7], the *RBBP4* fitness effect might not be limited to its roles within NuRD. Since CHD4 has been reported to have NuRD-dependent and independent effects and to constitute a peripheral member of NuRD[8,9], the fitness impact of CHD4 might also reflect NuRD independent roles. In contrast, we observed positive fitness scores for *ZBTB7A*, consistent with the previously observed requirement of ZBTB7A for terminal erythroid maturation[10,11]

Proteomic results

We identified 243 specific MTA2-interacting proteins by comparing MTA2 pulldown to IgG control (Fig. 2e and Supplementary Fig. 2d). The NuRD paralogs that were found to be functional by CRISPR screening also demonstrated higher abundance following MTA2 IP-MS. For instance, MTA2 was more abundant than MTA1 or MTA3, as evidenced by the comparison of their extracted ion chromatogram peaks. Similarly, CHD4 was more abundant than CHD3, GATAD2A than GATAD2B, HDAC2 than HDAC1, and RBBP4 than RBBP7. The only exception was the apparent abundance of MBD2, which was similar to MBD3, although MBD2 but not MBD3 was found to be required for HbF repression (Fig. 2b and Supplementary Fig. 2d). To orthogonally test the identity of the erythroid NuRD subcomplex members, we evaluated the proteins physically neighboring MTA2 in its intracellular milieu by proximity labeling. We fused a promiscuous prokaryotic biotin lysine protein ligase (BioID2) to MTA2 by means of a flexible linker to perform proximity-dependent biotinylation of neighboring proteins[12,13]. We expressed the MTA2-BioID2 fusion protein in HUDEP-2 cells in which we had first knocked out endogenous *MTA2.* We observed that expression of MTA2-BioID2 resulted in both auto-biotinylation and partial HbF repression, indicating that this fusion gene product was functional (Supplementary Fig. 2b-c). We isolated biotinylated proteins by streptavidin affinity capture and identified them by MS. In total we identified 60 proteins in proximity to MTA2 (Fig. 2e). Remarkably, we found that for 5 of the 6 NuRD subunit paralog families, only the paralog identified as functional by CRISPR screening was retrieved by proximity labeling (MTA2, RBBP4, CHD4, GATAD2A, HDAC2) (Fig. 2c and Supplementary Fig. 2e). Again the only exception was the MBD family, in which we retrieved both MBD2 and MBD3. Biotinylation of subcomplex subunits essential for HbF repression, such as HDAC2, was also validated by immunoblot analysis (Supplementary Fig. 2b). As additional evaluation of the NuRD subcomplex relevant for HbF repression, we performed CHD4 affinity purification IP-MS. These proteomics results were highly similar to those of MTA2 IP and MTA2 proximity labeling (Fig. 2e). Within NuRD, we found preferential interactions with functional members, including GATAD2A, HDAC2, MTA2, and RBBP4. Similar to results from MTA2 IP and proximity labeling, we detected similar interaction with both MBD2 and MBD3 (Fig. 2d and Supplementary Fig. 2f). HDAC1 and RBBP7, despite relatively high expression at the RNA level, were not identified as comprising the MTA2/CHD4 NuRD subcomplex by the combined MS experiments and were found to be functionally dispensable for HbF repression (although prior shRNA studies had implicated HDAC1 in HbF repression)[14,15]. With respect to the MBD family, MBD2 and MBD3 have previously been shown to assemble into mutually exclusive NuRD subcomplexes[16], have distinct phenotypes in knockout mice[17], and MBD3 lacks meCpG binding activity[18]. The intersection of proteins identified by MS included 10 histone proteins, especially members of the H2B family, consistent with NuRD as a histone-binding protein complex. The similarity of the genome editing and proteomics results enhances confidence in the role of this NuRD subcomplex in HbF silencing.

Functional scores and protein annotations
We computed the Protein Variation Effect Analyzer (PROVEAN) score, a measure of interspecies conservation[19], for single amino acid deletion of each amino acid within the NuRD genes. Lower PROVEAN score indicates more amino acid conservation. Analysis of protein disorder scores showed that sgRNAs targeting ordered regions of NuRD genes showed enhanced phenotype (HbF enrichment score vs. disorder score; Spearman r -0.358, p <0.0001; fitness score vs. disorder score; Spearman r 0.384, p < 0.0001; Supplementary Fig. 3a-b). We compared functional scores to domain annotation and found that sgRNAs targeting sequences within domains showed more robust HbF enrichment (p<0.0001) and cellular fitness phenotypes (p<0.0001; Supplementary Fig. 3c-d). We predicted secondary structures for all NuRD proteins, and found that sgRNAs targeting helix or sheet secondary structures showed enhanced effect on HbF enrichment (Kruskal-Wallis, p<0.0001) and fitness scores (Kruskal-Wallis, p<0.0001) (Supplementary Fig. 3e-f).

Mapping functional scores to protein structures
We observed elevated HbF scores throughout the coiled-coil complex between MBD2 and GATAD2A (PDB ID 2L2L) (Supplementary Fig. 3g)[20]. We recolored a structure showing the interaction of MTA1 and RBBP4 (PDB ID 5FXY)[21], coloring MTA2 functional scores at positions of confident alignment to MTA1. We observed the interacting regions showed especially reduced fitness scores for RBBP4 and elevated HbF enrichment scores for MTA2 (Supplementary Fig. 3h). We mapped CHD4 HbF enrichment scores at the PHD fingers and chromodomains (PDB ID 2L5U, 2L75 and 4O9I)[22,23] with especially potent scores found at the interface between PHD finger 2 and H3K9me3 (Supplementary Fig. 3i-k).

Generation of hemizygous in-frame and frameshift *MTA2* mutant clones
First, we identified individual sgRNAs with heightened HbF enrichment scores targeting structured, ordered, conserved regions of MTA2. We colored a structure of a BAH domain (PDB ID 1W4S)[24] based on HbF enrichment scores of MTA2 aligning residues. In addition we colored a structure of MTA1:HDAC1 (PDB ID 5ICN)[25], with HbF enrichment scores of aligning residues of the paralogs MTA2 and HDAC2 (Fig. 4a). We validated sgRNAs targeting around N20 in the BAH domain, D228 in the ELM2 domain, and T313 in the SANT domain, as well as A653 at relatively nonconserved (NC), disordered sequences. We used a two-step approach to produce in-frame and frameshift clones. First, we deleted one copy of *MTA2* by paired Cas9-mediated cleavages (Supplementary Fig. 5a). These *MTA2* heterozygous cells showed no increase in γ-globin or HbF level as compared to unedited cells (Supplementary Fig. 5b). Then we introduced individual sgRNAs targeting each of these amino acid residues and genotyped clones to identify hemizygous in-frame and frameshift mutations.

Two classes of *MTA2* loss-of-function clones
In-frame deletions around the nonconserved (NC) region of MTA2 A653 (including deletion lengths ranging from 5-13 amino acids) and around T313 at the SANT domain did not show an effect on MTA2 protein level (Fig. 4c). In contrast, in-frame deletions around N20 at the BAH domain (deletion length 2 amino acids) and D228 at the ELM2 domain (deletion lengths 7 and 8 amino acids) were associated with nearly undetectable MTA2 protein by immunoblot. The MTA2 mRNA level was intact in these clones, suggesting an impact on protein stability rather than production (Fig. 4d). Treatment with the proteasome inhibitor MG132 partially rescued expression of MTA2 in clone M7, consistent with elevated protein instability associated with in-frame deletion (Fig. 4e). Reciprocal immunoprecipitation of CHD4 from the MTA2 SANT deletion clone M4 validated this reduced interaction between CHD4 and mutant MTA2 (Fig. 4g

and Supplementary Fig. 5f). IP of CHD4 showed increased interaction with MTA1, suggesting that MTA1 was able to partially replace MTA2 in the complex, although not able to repress HbF.

Generation of hemizygous in-frame *CHD4* mutant clones

We took a hemizygous clone approach, similar to that employed at *MTA2* (Supplementary Fig. 6a) to characterize cells of defined *CHD4* genotype. First one copy of *CHD4* was deleted and then mutations introduced to the remaining allele to obtain hemizygous clones by limiting dilution (Supplementary Fig. 6b). In contrast to results at *MTA2,* we did not isolate any frameshift *CHD4* mutant clones, despite identifying eight clones with in-frame deletions. These results were consistent with the hypothesis that complete loss-of-function of *CHD4* was not tolerated by the erythroid precursors. We further characterized five clones with in-frame mutations at CHDCT2. These included four clones with single AA deletions (of either A1873 or T1874) or a three amino acid deletion (A1873_I1875del). Each of these clones demonstrated similar CHD4 protein level as control cells (Fig. 5e). In addition these clones showed a similar rate of expansion as control cells (Supplementary Fig. 6c). We observed elevated γ-globin and HbF level in all of these clones with in-frame mutations at CHDCT2 (Fig. 5f and Supplementary Fig. 6d).

Mouse model of *Chd4* CHDCT2 in-frame deletion

In the mouse, *Chd4* is required for lineage specification in the early embryo and its loss results in preimplantation embryonic lethality[26]. In addition, *Chd4* plays essential roles in the establishment and maintenance of cell identity in a variety of lineages, including hematopoietic stem cells, T-lymphocytes, epidermal cells, Schwann cells, cardiomyocytes, skeletal muscle cells, kidney progenitors, neural progenitors, and glial cells[27-35]. We tested the requirement of *Chd4* in erythropoiesis by generating *Chd4*$^{fl/fl}$; *EpoR-Cre*+ mice. Embryos with erythroid *Chd4* deficiency were pale at E12.5 and not viable at E14.5 (data not shown). These results suggest that *Chd4* plays an essential role in erythropoiesis. The sequences at *CHD4* around AA1872-1883 are highly conserved between human and mouse (Supplementary Fig. 7a). We hypothesized that in-frame deletions surrounding mouse *Chd4* CHDCT2 A1876, the orthologous position to human *CHD4* A1873, might avoid early embryonic lethality while still interfering with developmental γ-globin repression.

We designed an sgRNA and DNA donor template encoding deletion of *Chd4* A1876 with silent mutation at the PAM sequence to facilitate generation of homology-directed repair (HDR) alleles (Supplementary Fig. 7a). We delivered SpCas9:sgRNA ribonucleoprotein and single strand oligonucleotide donor to mouse fertilized oocytes and obtained 46 live pups, of which four founders possessed the *Chd4*^{A1876del} allele (Supplementary Fig. 7b). Targeted amplicon sequencing from tail biopsies of these founder mice identified reads with the 3 bp deletion and silent mutation indicating HDR at frequencies exceeding those of NHEJ-derived indels (17.14 to 49% of mapped reads for HDR and 0.1 to 9.6% for NHEJ). Three of these founders transmitted *Chd4*^{A1876del} to F1 progeny (Supplementary Fig. 7b). To assess the role of *Chd4* A1876 in fetal hemoglobin regulation, *Chd4*^{A1876del} mice were bred to mice transgenic for the human β-globin gene cluster[36]. Unlike *Chd4*$^{-/-}$ embryos with preimplantation lethality and *Chd4*$^{fl/fl}$; *EpoR-Cre*+ embryos for which viable embryos could not be isolated after E12.5, *Chd4*^{A1876del} homozygous embryos at E14.5 and E15.5 were indistinguishable from wild type controls although E16.5 embryos appeared pale suggesting delayed onset anemia (Supplementary Fig. 7c). Expression of γ-globin as fraction of total human β-like globin was increased in fetal livers of E14.5 embryos from 16.1% in wt mice to 37.8% in *Chd4*^{A1876del} homozygous mice, $p < 0.0001$ (Fig. 5i). Similarly at E15.5, *Chd4*^{A1876del} homozygous embryos demonstrated impaired γ-globin silencing, with 11.8% in wt mice and 23.1% in homozygous mutants, $p = 0.0003$, and at E16.5, γ-globin was increased from 5.2% to 17.9% (Fig. 5i). We observed modestly impaired silencing of mouse

endogenous β-like globins, most prominent at E16.5 where there was a 10-fold increase in εy expression in *Chd4*^{A1876del} homozygous embryos (Supplementary Fig. 7d). Of note *BCL11A* haploinsufficiency, which in humans results in hereditary persistence of fetal hemoglobin[37-39], in mice results in impaired silencing during embryogenesis of transgenic human γ-globin and endogenous mouse embryonic globins of similar magnitude as *Chd4*^{A1876del} homozygotes[40]. We have yet to obtain live births from the *Chd4*$^{+/A1876del}$ breeding pairs so cannot determine potential impact of *Chd4*^{A1876del} on survival to birth.

DNA damage response and cell death following *CHD4* editing

To explore the cellular impact of *CHD4* mutagenesis we performed transcriptional profiling by RNA sequencing. We compared the gene expression signatures of Cas9 expressing HUDEP-2 cells with sgRNA targeting *CHD4* CHDCT2 (A1873) and helicase domain (A742) as well as nontargeting control sgRNA and sgRNA targeting *MTA2* SANT domain (T313) (Supplementary Fig. 8a). We compared the differentially expressed genes between each of these gene targeting sgRNAs and the nontargeting control. We found 112 shared upregulated differentially expressed genes between these three perturbations. By GO term analysis, we found four significant terms, each hierarchically nested (response to stimulus, response to drug, gas transport, oxygen transport) of which three of the four genes in the shared node were the embryonic or fetal globins, *HBG1*, *HBG2*, *HBZ* (Supplementary Fig. 8b). This suggests that the major shared effect of these perturbations is HbF derepression. In contrast, Reactome pathway analysis of the 270 unique differentially expressed upregulated genes following *CHD4* helicase targeting identified 30 terms, including several associated with DNA damage response (Supplementary Fig. 8b).

CHD4 has been previously implicated in the maintenance of genomic integrity following DNA damage[9,41-44]. We hypothesized that the reduced fitness observed after *CHD4* helicase domain targeting might be associated with an elevated DNA damage response. We transduced sgRNAs targeting *CHD4* coding sequences around S221 (NC), A742 (helicase) and A1873 (CHDCT2) and a nontargeting sgRNA in Cas9 expressing HUDEP-2 cells. Three days after transduction we assessed the DNA damage response by γH2AX immunostaining. We found a large increase in γH2AX foci by immunocytochemistry in helicase domain targeted cells. We observed only a modest increase in γH2AX foci in CHDCT2 targeted cells as compared to nontargeting control, similar to that observed for cells targeted at the NC sequences of *CHD4* (Supplementary Fig. 8c-d). Immunoblotting cells targeted at *CHD4* helicase domain showed much greater induction of γH2AX as compared to cells targeted at CHDCT2 domain or NC sequences (Supplementary Fig. 8e). We observed similarly prominent foci induction by *CHD4* helicase but not CHDCT2 domain targeting by immunostaining 53BP1, another protein recruited to sites of DNA damage (Supplementary Fig. 8c). Moreover, Hoechst staining of cells from all four conditions showed a significantly higher number of pyknotic nuclei in helicase domain targeted condition compared to control, NC and CHDCT2 domain targeted conditions (Supplementary Fig. 8c, f). We compared the tolerance to etoposide, a known DNA strand break inducer, of CHDCT2 in-frame deletion compared to control cells. We did not observe hypersensitivity of these cells to etoposide (Supplementary Fig. 8g). Overall these data suggest that both in-frame and frameshift mutations at the *CHD4* helicase domain are associated with an elevated DNA damage response and induction of cell death whereas in-frame deletions at CHDCT2 elude the DNA damage response and cellular toxicity.

Mechanism whereby in-frame deletions at *CHD4* CHDCT2 impact the NuRD complex

We first performed immunoprecipitation of CHD4 from the *CHD4* CHDCT2 in-frame deletion mutant hemizygous clones as compared to control cells. However we found that the ability to pull-down CHD4 was reduced as compared to control clones even though the total level of CHD4 as measured by immunoblotting was unchanged (Supplementary Fig. 8h). Therefore it

was difficult to interpret if the CHD4 mutation impacted immunoprecipitation efficiency directly or through indirect effects on NuRD complex assembly. As an alternative approach, we used an antibody recognizing MTA2 to perform immunoprecipitation.

Future directions for dense mutagenesis in situ
The resolution of the CRISPR mutagenesis in situ in these experiments was limited by availability of NGG PAM sites restricting SpCas9 cleavage. As genome editing tools advance, we anticipate that future dense mutagenesis screens may achieve even higher resolution. Direct measurement of indels in phenotypically selected cell populations following pooled screening could enable direct genotype-phenotype assignment, but could come at the cost of sensitivity, throughput, and cost-effectiveness. Dense indel screening might improve by using nucleases with less restrictive PAM sequences, such as recently described engineered Cas9 variants[45,46]. Alternative approaches could be dense base editing in absence of double strand breaks[8,47-50] or high-throughput homologous repair[51]. As the resolution and throughput of genome editing perturbation range continue to progress, we anticipate acceleration of the transition from genetic results to rational therapeutic design. Comprehensive mutagenesis of protein complexes in situ could be readily adapted to the study of many disease-relevant cellular processes.

Methods (Supplementary Note)

HUDEP-2 cell culture. HUDEP-2 cells[52] were cultured as described previously[2], in StemSpan SFEM (Stem Cell Technologies) supplemented with 1 μM dexamethasone (Sigma), 100 ng ml^{-1} human stem cell factor (SCF) (R&D), 3 international units (IU) ml^{-1} erythropoietin (Amgen), 1% L-glutamine (Life Technologies), and 1% penicillin/streptomycin. 1ug ml^{-1} doxycycline (Sigma) was included in the culture to induce the expression of the human papilloma virus type 16 E6/E7 genes. HUDEP-2 cells were differentiated in Iscove's Modified Dulbecco's Medium (IMDM) (Life Technologies) supplemented with 330 μg ml^{-1} holo-transferrin (Sigma),10 μg ml^{-1} recombinant human insulin (Sigma), 2 IU ml^{-1} heparin (Sigma), 5% human solvent detergent pooled plasma (AB) (Rhode Island Blood Center), 3 IU ml^{-1} erythropoietin, 100 ng ml^{-1} human SCF, 1 μg ml^{-1} doxycycline, 1% L-glutamine, and 1% penicillin/streptomycin.

Human primary erythroid cell culture CRISPR/Cas9 delivery.
Human $CD34^+$ HSPCs were cultured in X-Vivo 10 medium (Lonza) supplemented with 100 ng/mL SCF, 100 ng/mL TPO, and 50 ng/mL FLT3 for 24 hours before RNP electroporation. RNP complexes containing 100 pmol SpCas9 protein[53] and 100 pmol modified sgRNAs (Synthego) targeting *CHD4* were electroporated into 50,000 $CD34^+$ HSPCs using a 4D-Nucleofector™ System (Lonza) according to manufacturer's instructions. Immediately after electroporation, cells were transferred to erythroid differentiation medium 1 (EDM1) consisting of base EDM (IMDM supplemented with 1% P/S, 330 μg/mL human holo-transferrin, 10 μg/mL recombinant human insulin, 2 IU/mL heparin, 5% inactivated human plasma, and 3 IU/mL EPO) supplemented with 1μM hydrocortisone, 5 ng/mL IL-3, and 100 ng/mL SCF, and cultured for 7 days, followed by culture in EDM2 (base EDM supplemented with 100 ng/mL SCF) for 4 days. Cell concentrations were maintained below 1 million cells/mL during EDM1 and EDM2 culture. After EDM2, cells were seeded in EDM3 (base EDM with no additional supplement) at a concentration of 1 million cells/mL and cultured for 7 days without further media change. Genomic DNA was collected at day 2 in EDM1, day 4 in EDM2, and day 7 in EDM3 to perform PCR amplification and subsequent deep sequencing. Cells were collected for RNA analysis, HbF expression, HPLC analysis, enucleation assay, CD71 and CD235a staining at day 7 in EDM3.

Generation of hemizygous deletion clones.
Hemizygous cellular clones carrying in-frame or frameshift mutations were generated following a two step strategy. In the first step, one allele was deleted using a pair of sgRNA targeting outside of the coding region while in the second step, sgRNA was introduced targeting the region of interest in the second allele of gene (Supplementary Fig. 5a, 6a). sgRNA sequences were cloned into LentiGuide-Puro (digested with BsmBI). 3-4 million HUDEP-2 cells with 3 µg of pX330-Cas9 and 1.5 µg LentiGuide-Puro-sgRNA were electroporated using VCA-1003 kit and L-29 electroporation program of Nucleofector I device (Lonza). Quickly after electroporation, cells were recovered in 7-10 mL pre-warmed recovery media (20% FBS in IMDM) in 15 ml centrifugation tube and shifted to standard cell culture incubator for 15 minutes. Following 15 minutes incubation, cells were centrifuged at 200 x *g* for three minutes. Most of the supernatant was taken out by leaving behind 0.7 ml, in which the cell pellet was re-suspended and shifted back in the standard incubator for another 40-60 minutes. Then cells were centrifuged again at 200 x *g* for 3 minutes, and resuspended in 1 ml pre-warmed HUDEP-2 media in 12 well plates. 48 hours post-nucleofection, media was replaced with fresh HUDEP-2 media containing 1µg ml^{-1} puromycin, following another 48 hours media was refreshed and puromycin was removed in the fresh media (puromycin pulse) and using Quick Extract, DNA sample was collected for PCR. All the PCR reactions used for generation and characterization of in-frame and frameshift clones were performed using Accuprime Supermix II (Life Technologies) with following conditions: 94°C for 2 minutes; 35 cycles of 94°C for 20s, 60°C for 20S, 68°C for 1 minute kb^{-1} of PCR product; 68°C for 5 minutes. If the deletion band was detected, then the bulk cultures were plated clonally at limiting dilution. After 14 days of clonal expansion, genomic DNA was extracted using QuickExtract (50 µl/well) and PCR was performed. For the identification of heterozygous clones having only one copy of the gene, clones were screened for deletion (one copy) by conventional PCR with the PCR reaction internal to segment to be deleted (non-deletion band) and one gap-PCR reaction across the deletion junction (deletion band)[54] that would only amplify in the presence of deletion. Heterozygous clones will have both deletion and non-deletion bands. Heterozygous clones were used to generate hemizygous indels using sgRNA targeting the region of interest. Following electroporation and pulse (48 hours) of puromycin treatment, PCR was conducted using primers around sgRNA targeting region. PCR product was purified using Qiagen PCR purification kit and used for Sanger sequence. Sequence traces were analyzed using TIDE. If the editing efficiency was higher than 80% then the bulk cultures were plated clonally at limiting dilution. 96 well plates with greater than 30 clones per plate were excluded to avoid mixed clones. Following expansion period (14 days) DNA samples were collected from the clones and PCR was performed around the sgRNA target site. Purified PCR product was used for Sanger sequencing. Sequence traces were aligned against the DNA sequence of gene of interest to read the genotype of a specific clone. Biallelic deletion of *MTA2* clone was identified as the absence of the non-deletion PCR band and the presence of deletion PCR band. sgRNA and primer sequences used can found in Supplementary Table 3.

HbF expression and globin gene expression.
$1x10^6$ HUDEP-2 cells were prepared for flow cytometric analysis. 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 1-2 µl of purified FITC/APC-conjugated IgG_1 anti-Human Fetal Hemoglobin antibody (ThermoFisher) in the dark at room temperature for 30 min. For surface marker staining, 0.5 x 10^6 cells were incubated with 5 µl each of anti-human CD71 (PE-Cy7) and anti-human CD235a (APC) in 0.1% BSA in PBS at 4°C for 30 min. For enucleation assay, 0.5 x 10^6 cells were incubated with Hoechst 33342 (1:5000) in 0.1% BSA in PBS at 4°C for 10

min. Cells were analyzed using an LSRII flow cytometer, recording 10,000 events per condition. For HPLC, 5 x 10^6 cells were lysed in hemolysate reagent (Helena laboratories) on ice for 20 min. The supernatants were clarified by centrifugation at 14,000 x *g* for 15 min at 4°C, and were analyzed using a D-10 hemoglobin analyzer (Bio-Rad) according to manufacturer's instructions. Total RNA was obtained from either single-cell clones or bulk cells using RNeasy mini kit (Qiagen) and subjected for quantitative real-time PCR (RT-qPCR) using SYBR® Select Master Mix (Life Technologies). Primers used are found in Supplementary Table 3. Gene expression was normalized to that of GAPDH. All gene expression data reported from these clones represents the mean of at least three replicates.

RNA sequencing.
Total 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 quantified by Qubit fluorometer, Agilent TapeStation 2200, and RT-qPCR using the Kappa Biosystems library quantification kit according to manufacturer's protocols. Uniquely indexed libraries were pooled in equimolar ratios and sequenced on an Illumina NextSeq500 with single-end 75 bp reads by the Dana-Farber Cancer Institute Molecular Biology Core Facilities. RNA-seq data were analyzed using open source software Tuxedo suite: Bowtie, TopHat, Cufflinks and Cuffdiff[55-55]. The single end sequencing reads were first quality assessed by fastqc[56] and trimmed at Q20. Then TopHat2 was used to align RNA-seq reads to human genome GRCh38, which was indexed by Bowtie2. Cufflinks2 assembled mapped reads into transcripts and then generates quantification for each transcript. Cuffdiff2 was applied to do differential analyses on CHD4-CHDCT2, CHD4-helicase, MTA2-SANT perturbation samples compared with the control samples treated with nontargeting guide. The up- and down- regulated genes were characterized by significant q-value<0.05 in the Cuffdiff report and the overlapping were displayed by Venn diagram. The gene ontology enrichment test on functional annotations was performed using the web service from GO consortium[57].

Liquid chromatography nanoelectrospray ionization mass spectrometry.
The mass spectrometer was operated in data-dependent acquisition mode, with MS1 recorded in the orbitrap detector (400–1800 m/z range, 240,000 resolution), and MS2 collected using the linear ion trap (top speed mode, dynamic precursor exclusion enabled for 90 sec). Raw files are available via the ProteomeXchange repository[58] with accession number PXD009793. Raw files were analyzed using MaxQuant Version 1.6.0.16[59] against the human SwissProt reference database (version 01/27-16, containing isoforms) supplemented with contaminant protein sequences. FDR was set at <0.01 for both peptide and protein level. Mass tolerance was set at 5 ppm for orbitrap spectra and 0.5 Da for linear ion trap spectra. C-carbamidomethylation was set as fixed modification, while acetylation of protein N-terminus and M-oxidation were allowed as variable. Chromatographic feature alignment for label-free quantification (LFQ) was enabled.

The protein abundance was quantified based on LFQ intensity (sum of the integrated area of the extracted ion chromatogram (XIC) for each peptide assigned to the protein)[60]. The missing values were imputed with the minimum LFQ across all samples[61]. The control samples (IgG) were normalized to their total LFQ values while MTA2 samples were normalized to the MTA2 level and CHD4 were normalized to the CHD4 level in order to assess the relative level of co-immunoprecipitation to the bait proteins. Specific interactions were identified by comparing test to control using simple t-test and p-values were corrected by Benjamini-Hochberg procedure.

Supplementary Table 1: Quantitative analysis of mass spectrometery data. In MTA2_CoIP and CHD4_CoIP sheets, MTA2_CoIP (*n* = 2) and CHD4_CoIP (*n* = 4) sample and IgG control

(*n* = 4) columns display the normalized log2 (LFQ intensity), p-value shows the statistical significance of the differential t-test comparing IP samples with IgG controls, and corrected-p-values are based on Benjamini–Hochberg procedure. In MTA2_BioID2 sheet, p-value shows the statistical significance of the target protein enrichment comparing MTA2-BioID2 (*n* = 1) with NLS-BioID2 (*n* = 1) and streptavidin (*n* = 1) and corrected-p-values are based on Benjamini–Hochberg procedure.

Supplementary Table 2: Functional scores of CRISPR library (all sgRNAs) against various protein-level sequence annotations.

Supplementary Table 3: List of oligonucleotides and PCR primer sequences used in the study.

Supplementary Table 4: Key resources.

References (Supplementary Note)

1. Canver, M.C. et al. Variant-aware saturating mutagenesis using multiple Cas9 nucleases identifies regulatory elements at trait-associated loci. *Nat Genet* **49**, 625-634 (2017).
2. Canver, M.C. et al. BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. *Nature* **527**, 192-7 (2015).
3. 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**, 900-13 (2016).
4. Shi, J. et al. Discovery of cancer drug targets by CRISPR-Cas9 screening of protein domains. *Nat Biotechnol* **33**, 661-7 (2015).
5. Donovan, K.F. et al. Creation of Novel Protein Variants with CRISPR/Cas9-Mediated Mutagenesis: Turning a Screening By-Product into a Discovery Tool. *PLoS One* **12**, e0170445 (2017).
6. Wolffe, A.P., Urnov, F.D. & Guschin, D. Co-repressor complexes and remodelling chromatin for repression. *Biochem Soc Trans* **28**, 379-86 (2000).
7. Kuzmichev, A., Nishioka, K., Erdjument-Bromage, H., Tempst, P. & Reinberg, D. Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. *Genes Dev* **16**, 2893-905 (2002).
8. Arribere, J.A. et al. Translation readthrough mitigation. *Nature* **534**, 719-23 (2016).
9. O'Shaughnessy, A. & Hendrich, B. CHD4 in the DNA-damage response and cell cycle progression: not so NuRDy now. *Biochem Soc Trans* **41**, 777-82 (2013).
10. Maeda, T. et al. LRF is an essential downstream target of GATA1 in erythroid development and regulates BIM-dependent apoptosis. *Dev Cell* **17**, 527-40 (2009).
11. Masuda, T. et al. Transcription factors LRF and BCL11A independently repress expression of fetal hemoglobin. *Science* **351**, 285-9 (2016).
12. Roux, K.J., Kim, D.I., Raida, M. & Burke, B. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J Cell Biol* **196**, 801-10 (2012).
13. Kim, D.I. et al. An improved smaller biotin ligase for BioID proximity labeling. *Mol Biol Cell* **27**, 1188-96 (2016).
14. 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 U S A* **107**, 12617-22 (2010).

15. 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**, 624-8 (2015).
16. Le Guezennec, X. et al. MBD2/NuRD and MBD3/NuRD, two distinct complexes with different biochemical and functional properties. *Mol Cell Biol* **26**, 843-51 (2006).
17. Hendrich, B., Guy, J., Ramsahoye, B., Wilson, V.A. & Bird, A. Closely related proteins MBD2 and MBD3 play distinctive but interacting roles in mouse development. *Genes Dev* **15**, 710-23 (2001).
18. Saito, M. & Ishikawa, F. The mCpG-binding domain of human MBD3 does not bind to mCpG but interacts with NuRD/Mi2 components HDAC1 and MTA2. *J Biol Chem* **277**, 35434-9 (2002).
19. Choi, Y. & Chan, A.P. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics* **31**, 2745-7 (2015).
20. 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 U S A* **108**, 7487-92 (2011).
21. Lauffer, B.E. et al. Histone deacetylase (HDAC) inhibitor kinetic rate constants correlate with cellular histone acetylation but not transcription and cell viability. *J Biol Chem* **288**, 26926-43 (2013).
22. Mansfield, R.E. et al. Plant homeodomain (PHD) fingers of CHD4 are histone H3-binding modules with preference for unmodified H3K4 and methylated H3K9. *J Biol Chem* **286**, 11779-91 (2011).
23. Wiggs, K.R., Chruszcz, M., Su, X., Minor, W., Khorasanizadeh, S. Structure of CHD4 double chromodomains depicts cooperative folding for DNA binding. in

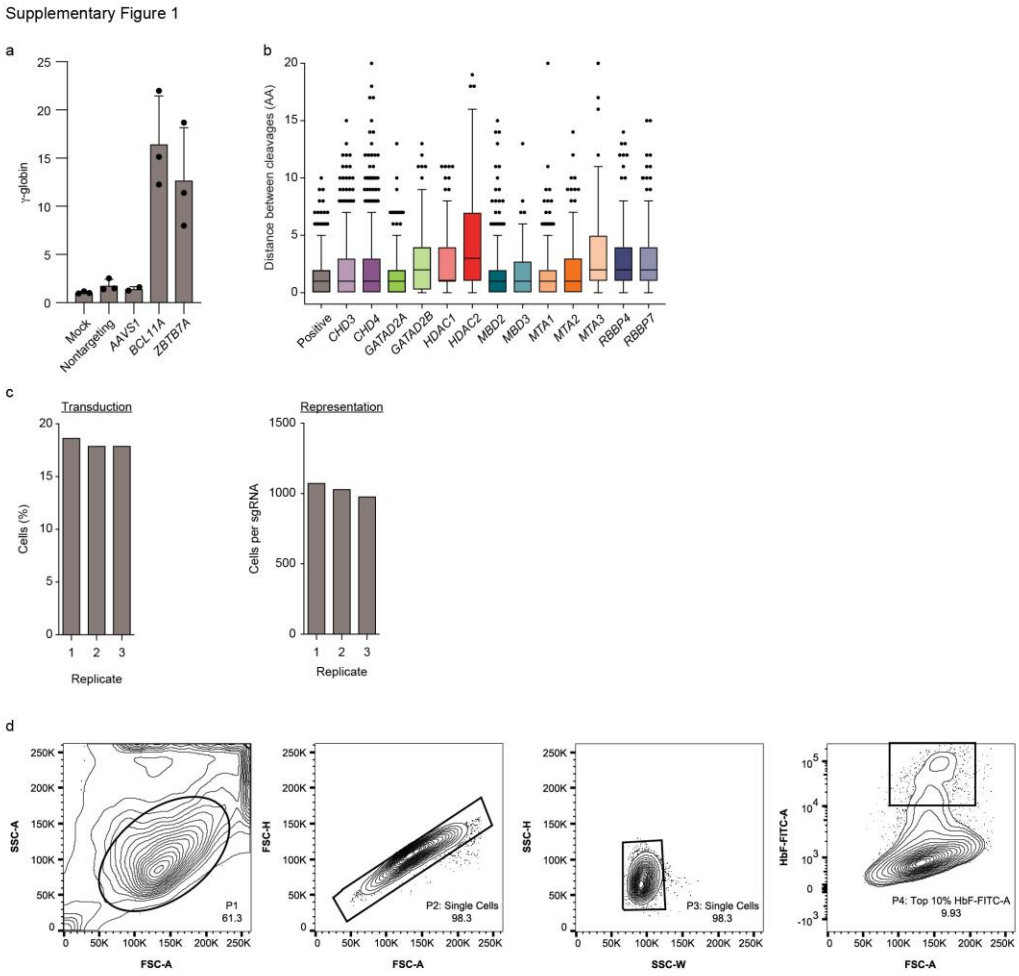
(Protein Data Bank (4O9I)

2017).
24. Oliver, A.W. et al. Crystal structure of the proximal BAH domain of the polybromo protein. *Biochem J* **389**, 657-64 (2005).
25. Millard, C.J. et al. The structure of the core NuRD repression complex provides insights into its interaction with chromatin. *Elife* **5**, e13941 (2016).
26. O'Shaughnessy-Kirwan, A., Signolet, J., Costello, I., Gharbi, S. & Hendrich, B. Constraint of gene expression by the chromatin remodelling protein CHD4 facilitates lineage specification. *Development* **142**, 2586-97 (2015).
27. Hung, H., Kohnken, R. & Svaren, J. The nucleosome remodeling and deacetylase chromatin remodeling (NuRD) complex is required for peripheral nerve myelination. *J Neurosci* **32**, 1517-27 (2012).
28. Wilczewski, C.M. et al. CHD4 and the NuRD complex directly control cardiac sarcomere formation. *Proc Natl Acad Sci U S A* **115**, 6727-6732 (2018).
29. Kashiwagi, M., Morgan, B.A. & Georgopoulos, K. The chromatin remodeler Mi-2beta is required for establishment of the basal epidermis and normal differentiation of its progeny. *Development* **134**, 1571-82 (2007).
30. Gomez-Del Arco, P. et al. The Chromatin Remodeling Complex Chd4/NuRD Controls Striated Muscle Identity and Metabolic Homeostasis. *Cell Metab* **23**, 881-92 (2016).
31. Williams, C.J. et al. The chromatin remodeler Mi-2beta is required for CD4 expression and T cell development. *Immunity* **20**, 719-33 (2004).
32. Yoshida, T. et al. The role of the chromatin remodeler Mi-2beta in hematopoietic stem cell self-renewal and multilineage differentiation. *Genes Dev* **22**, 1174-89 (2008).

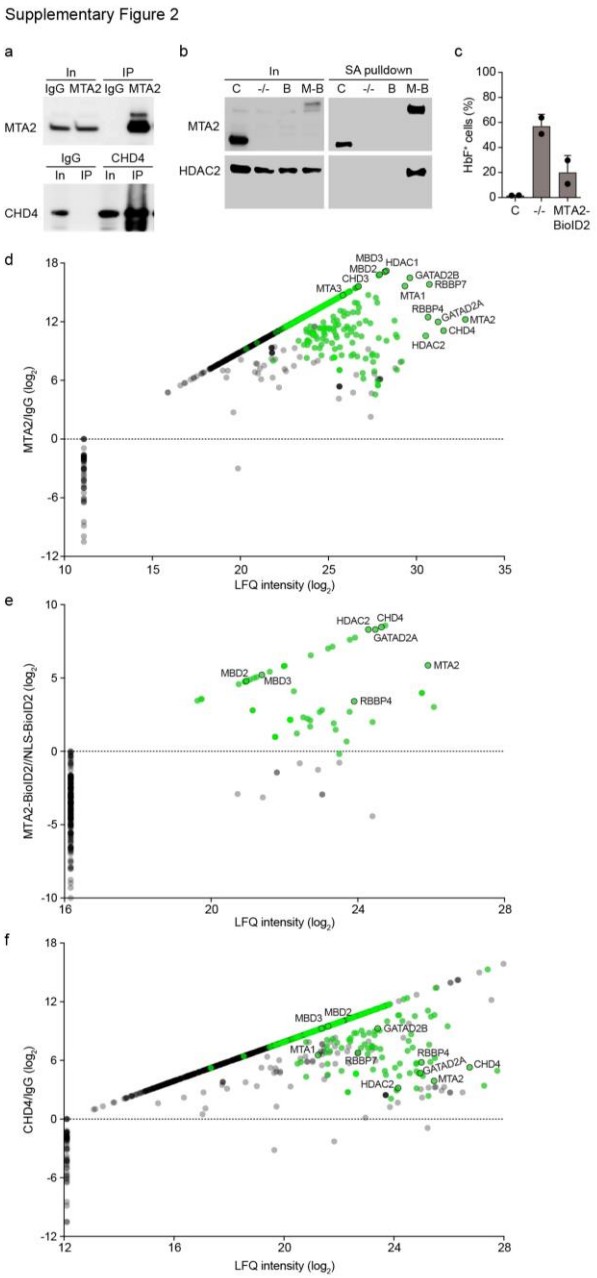
33. Denner, D.R. & Rauchman, M. Mi-2/NuRD is required in renal progenitor cells during embryonic kidney development. *Dev Biol* **375**, 105-16 (2013).
34. Nitarska, J. et al. A Functional Switch of NuRD Chromatin Remodeling Complex Subunits Regulates Mouse Cortical Development. *Cell Rep* **17**, 1683-1698 (2016).
35. Sparmann, A. et al. The chromodomain helicase Chd4 is required for Polycomb-mediated inhibition of astroglial differentiation. *EMBO J* **32**, 1598-612 (2013).
36. Porcu, S. et al. The human beta globin locus introduced by YAC transfer exhibits a specific and reproducible pattern of developmental regulation in transgenic mice. *Blood* **90**, 4602-9 (1997).
37. Basak, A. et al. BCL11A deletions result in fetal hemoglobin persistence and neurodevelopmental alterations. *J Clin Invest* **125**, 2363-8 (2015).
38. Funnell, A.P. et al. 2p15-p16.1 microdeletions encompassing and proximal to BCL11A are associated with elevated HbF in addition to neurologic impairment. *Blood* **126**, 89-93 (2015).
39. Dias, C. et al. BCL11A Haploinsufficiency Causes an Intellectual Disability Syndrome and Dysregulates Transcription. *Am J Hum Genet* **99**, 253-74 (2016).
40. Sankaran, V.G. et al. Developmental and species-divergent globin switching are driven by BCL11A. *Nature* **460**, 1093-7 (2009).
41. Schmidt, D.R. & Schreiber, S.L. Molecular association between ATR and two components of the nucleosome remodeling and deacetylating complex, HDAC2 and CHD4. *Biochemistry* **38**, 14711-7 (1999).
42. Matsuoka, S. et al. ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. *Science* **316**, 1160-6 (2007).
43. Stokes, M.P. et al. Profiling of UV-induced ATM/ATR signaling pathways. *Proc Natl Acad Sci U S A* **104**, 19855-60 (2007).
44. Stanley, F.K., Moore, S. & Goodarzi, A.A. CHD chromatin remodelling enzymes and the DNA damage response. *Mutat Res* **750**, 31-44 (2013).
45. Hu, J.H. et al. Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* **556**, 57-63 (2018).
46. Nishimasu, H. et al. Engineered CRISPR-Cas9 nuclease with expanded targeting space. *Science* (2018).
47. Komor, A.C., Kim, Y.B., Packer, M.S., Zuris, J.A. & Liu, D.R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420-4 (2016).
48. Gaudelli, N.M. et al. Programmable base editing of A*T to G*C in genomic DNA without DNA cleavage. *Nature* **551**, 464-471 (2017).
49. Findlay, G.M., Boyle, E.A., Hause, R.J., Klein, J.C. & Shendure, J. Saturation editing of genomic regions by multiplex homology-directed repair. *Nature* **513**, 120-3 (2014).
50. Rajagopal, N. et al. High-throughput mapping of regulatory DNA. *Nat Biotechnol* **34**, 167-74 (2016).
51. Findlay, G.M. et al. Accurate classification of BRCA1 variants with saturation genome editing. *Nature* (2018).
52. Kurita, R. et al. Establishment of immortalized human erythroid progenitor cell lines able to produce enucleated red blood cells. *PLoS One* **8**, e59890 (2013).
53. Wu, Y. et al. Highly efficient therapeutic gene editing of human hematopoietic stem cells. *Nat Med* (2019).
54. Canver, M.C. et al. Characterization of genomic deletion efficiency mediated by clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 nuclease system in mammalian cells. *J Biol Chem* **289**, 21312-24 (2014).
55. Kim, D. et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* **14**, R36 (2013).

56. Andrews, S. FastQC A Quality Control tool for High Throughput Sequence Data: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/. (2010).
57. Amaya, M. et al. Mi2beta-mediated silencing of the fetal gamma-globin gene in adult erythroid cells. *Blood* **121**, 3493-501 (2013).
58. Hermjakob, H. & Apweiler, R. The Proteomics Identifications Database (PRIDE) and the ProteomExchange Consortium: making proteomics data accessible. *Expert Rev Proteomics* **3**, 1-3 (2006).
59. Tyanova, S., Temu, T. & Cox, J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat Protoc* **11**, 2301-2319 (2016).
60. Cox, J. et al. Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol Cell Proteomics* **13**, 2513-26 (2014).
61. Lazar, C., Gatto, L., Ferro, M., Bruley, C. & Burger, T. Accounting for the Multiple Natures of Missing Values in Label-Free Quantitative Proteomics Data Sets to Compare Imputation Strategies. *J Proteome Res* **15**, 1116-25 (2016).

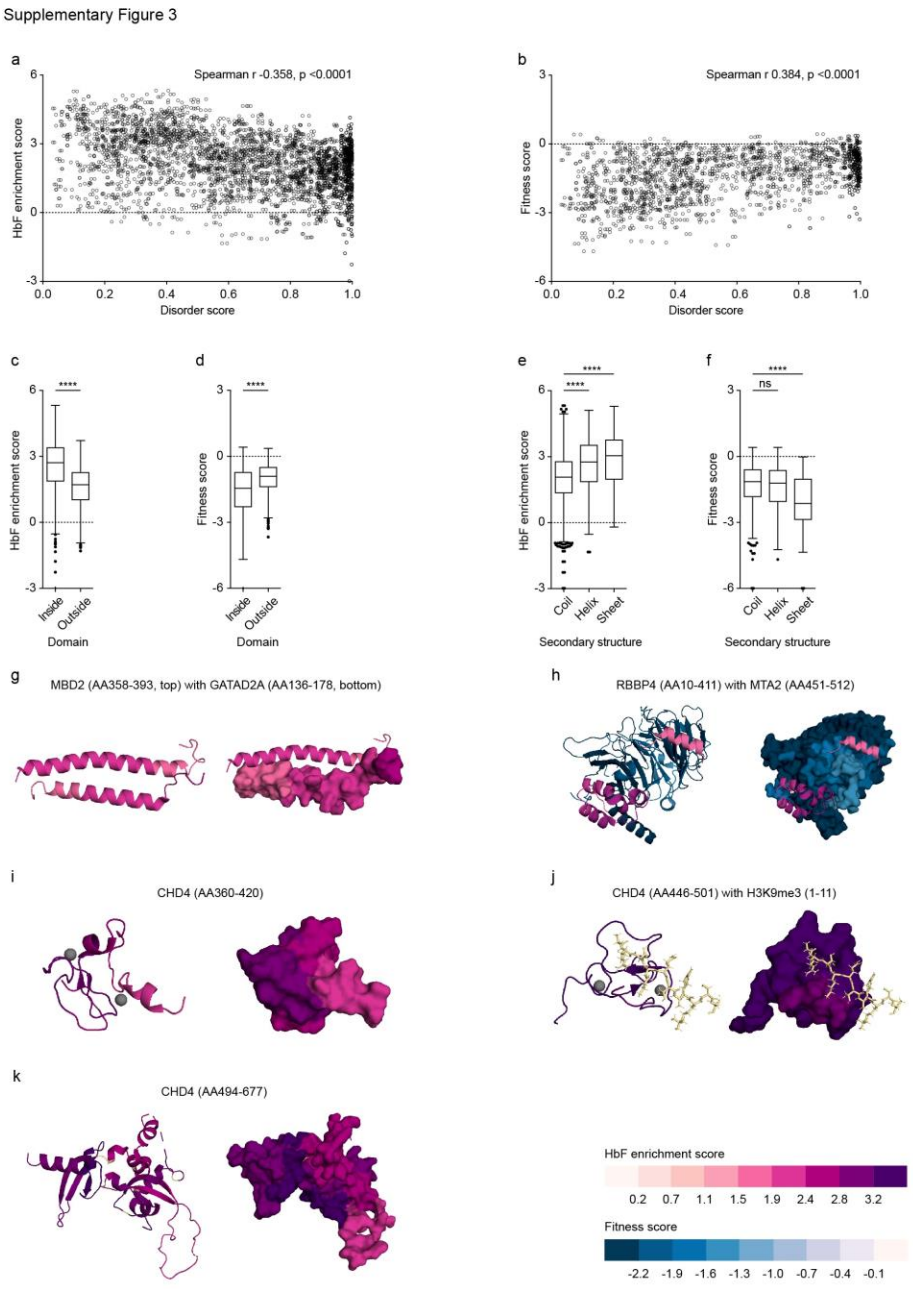
Supplementary figures



Supplementary Fig. 1: NuRD gene dense mutagenesis SpCas9 sgRNA library screen. a, HUDEP-2 cells expressing SpCas9 demonstrate increase in fetal γ-globin expression when transdueced with positive control *BCL11A* or *ZBTB7A* targeting sgRNAs but not with no sgRNA (mock), nontargeting sgRNA, or neutral locus *AAVS1* targeting sgRNA. β- and γ-globin expression in HUDEP-2 cells in erythroid differentiation culture for 11 days as measured by RT-qPCR. Error bars show mean and SD, $n = 3$. **b,** Distribution of distances between adjacent cleavages for target NuRD genes (sgRNAs per gene per Fig. 1a). Boxplot shows median, 25th and 75th percentiles with whiskers and outliers per Tukey method. **c**, Transduction of SpCas9 expressing HUDEP-2 cells by pooled sgRNA library determined by counting puromycin resistant compared to sensitive cells for three biological replicates. Representation calculated by dividing transduced cells per replicate by number of sgRNAs in pooled library. **d,** Flow cytometry plots (representative of three independent experiments) demonstrating gating strategy to sort 10% brightest HbF-FITC positive cells.

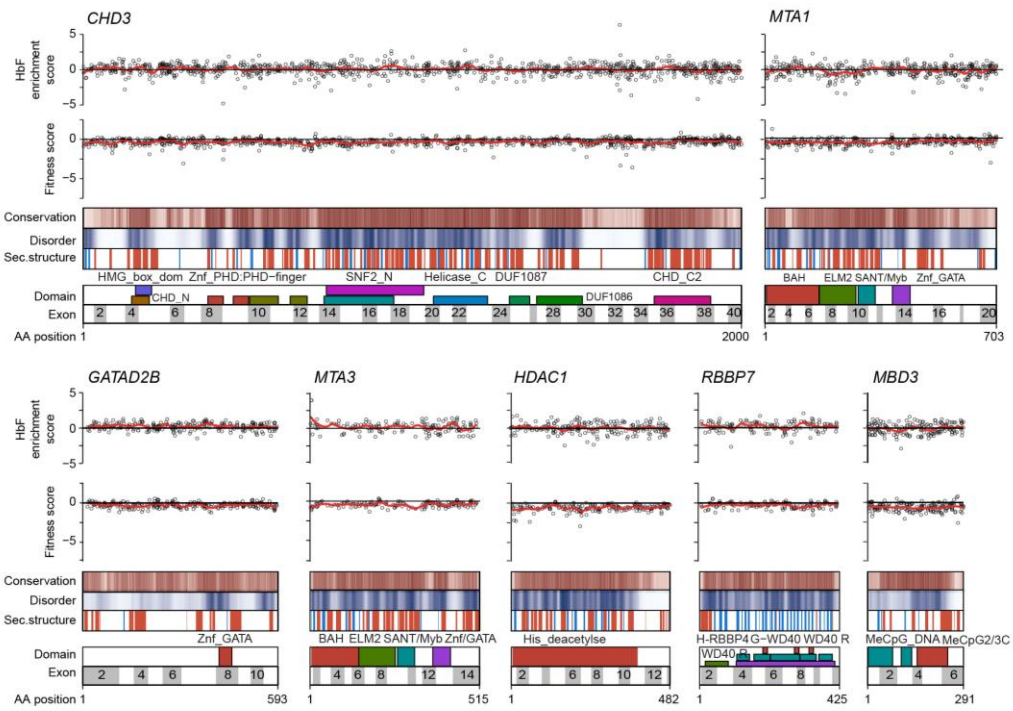


Supplementary Fig. 2: Affinity purification or proximity labeling mass spectrometry. **a,** Representative images of IP MTA2 (upper panel) and IP CHD4 (lower panel) followed by immunoblot analysis, representative of three independent experiments. **b**, Streptavidin pulldown from cellular lysates expressing MTA2-BioID2 followed by immunoblot using antibodies against MTA2 and HDAC2, representative of two independent experiments. **c,** MTA2-BioID2 expression in *MTA2* KO HUDEP-2 cells partially restores the HbF repression function of MTA2. The graph shows FACS based quantification of HbF^+ cells in control (SpCas9 expressing HUDEP-2), *MTA2* knockout and MTA2-BioID2 HUDEP-2 cells, error bars show mean and SD, *n* = 3. **d**-**f**, Each dot represent a protein detected in MS. y-axis shows $\log_2$ fold change in sample compared to control (IgG) and x-axis shows $\log_2$ of average LFQ intensity. Proteins with corrected p-value<0.2 for specific interactions based on intensity greater in test compared to control labeled in green, and those for NuRD members outlined and labeled with name. (**d**) IP MTA2 followed by MS, (**e**) SA pulldown of MTA2-BioID2 followed by MS, (**f**) IP CHD4 followed by MS (*n* = 2 in **d**, *n* = 1 in **e**, *n* = 4 in **f**).

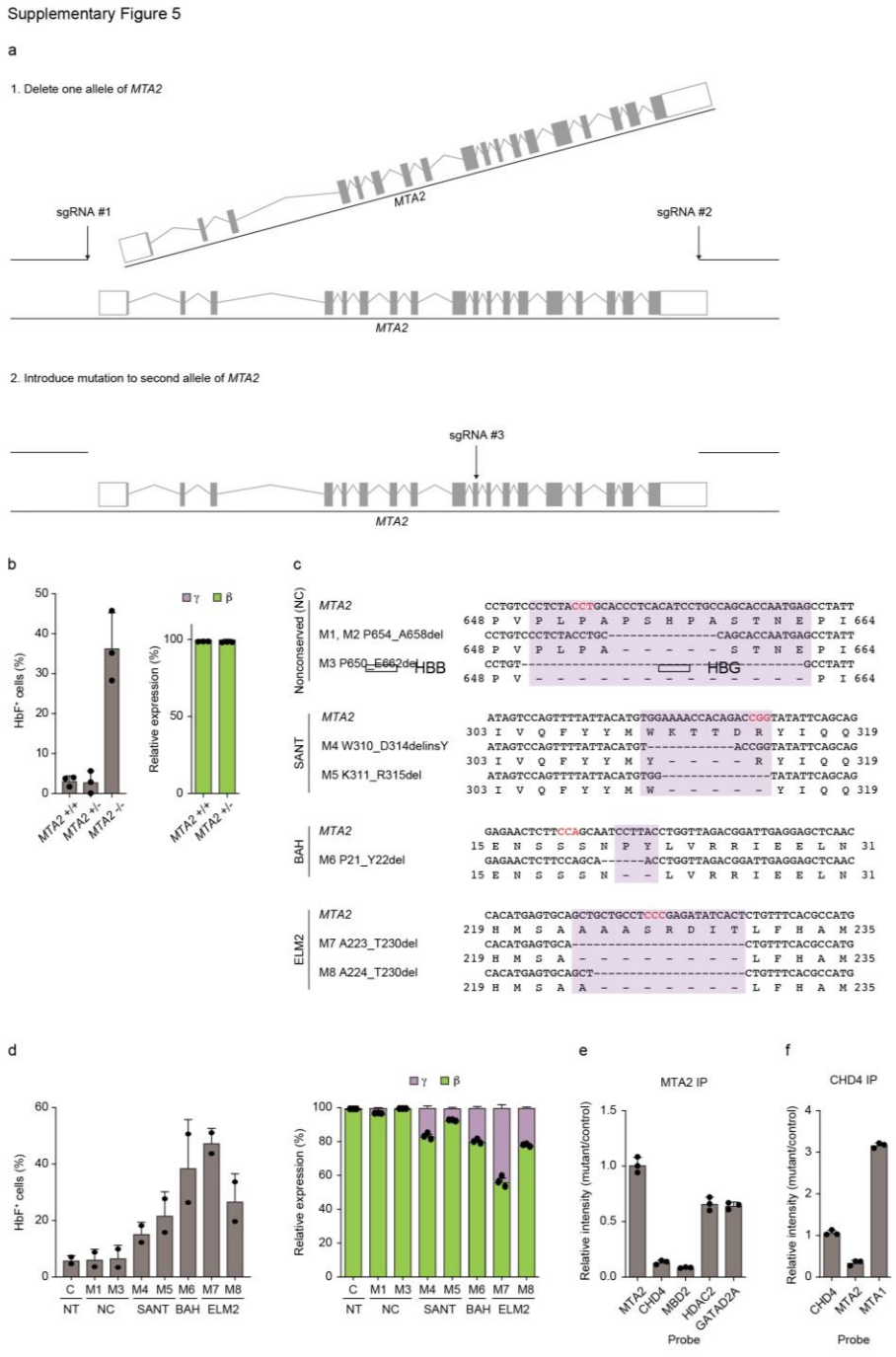


Supplementary Fig. 3: CRISPR-Cas9 mediated dense mutagenesis identifies critical residues within NuRD complex. a and **b**, Scatter plots of HbF enrichment and fitness scores compared to disorder scores for individual sgRNAs for hit genes (*n* = 3765 sgRNAs for **a**, **c** and 1955 for **b**, **d**). Spearman r and p-values shown. Each dot represents sgRNA average score from three independent experiments. **c** and **d**, Box plots showing HbF enrichment and fitness scores for sgRNAs of hit genes inside compared to outside domains (**c**, **d**) and within coiled/unstructured, helix, and sheet secondary structures (**e**, **f**). **** indicates p < 0.0001. Boxplot shows median, 25th and 75th percentiles with whiskers and outliers per Tukey method. **g-k**, Structures colored based on LOESS regression HbF enrichment scores for MBD2, GATAD2A, MTA2, and CHD4 and fitness scores for RBBP4. Left panels depict cartoon and right panels surface models. **g**, Coiled coil interaction between MBD2 and GATAD2A with aligned scores mapped to PDB ID 2L2L. **h**, MTA2 and RBBP4 interaction with aligned scores mapped to PDB ID 5FXY. **i**, CHD4 aligned scores mapped to PDB ID 2L5U (zinc ions in gray). **j**, CHD4 aligned scores mapped to PDB ID 2L75 (histone tail in yellow, zinc ions in gray). **k**, CHD4 aligned scores mapped to PDB ID 4O9I. **l**, Linear maps of nonhit NuRD genes. HbF enrichment scores and fitness scores are shown for each sgRNA as red dots, with LOESS regression line in blue. Disorder scores are shown as heatmap from white to blue from maximal disorder to maximal order. Evolutionary conservation PROVEAN scores are shown as heatmap from white to brown from minimal to maximal conservation. Secondary structure predictions are shown with helix in red and sheet in blue.

Supplementary Figure 4

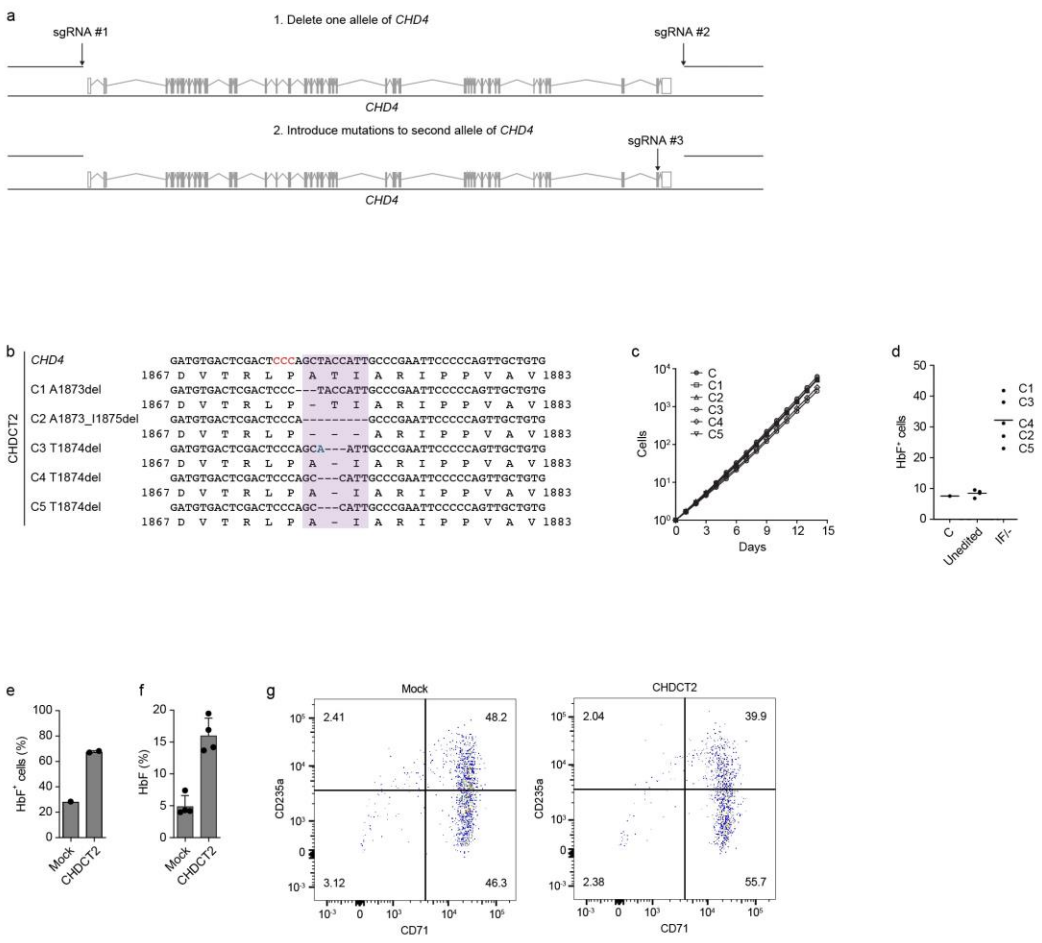


Supplementary Fig. 4: Linear maps of non-hit NuRD genes. HbF enrichment scores and fitness scores are shown for each sgRNA as dots, with LOESS regression line in red. Disorder scores are shown as heatmap from white to blue from maximal disorder to maximal order. Evolutionary conservation PROVEAN scores are shown as heatmap from white to brown from minimal to maximal conservation. Secondary structure predictions are shown with helix in red and sheet in blue.

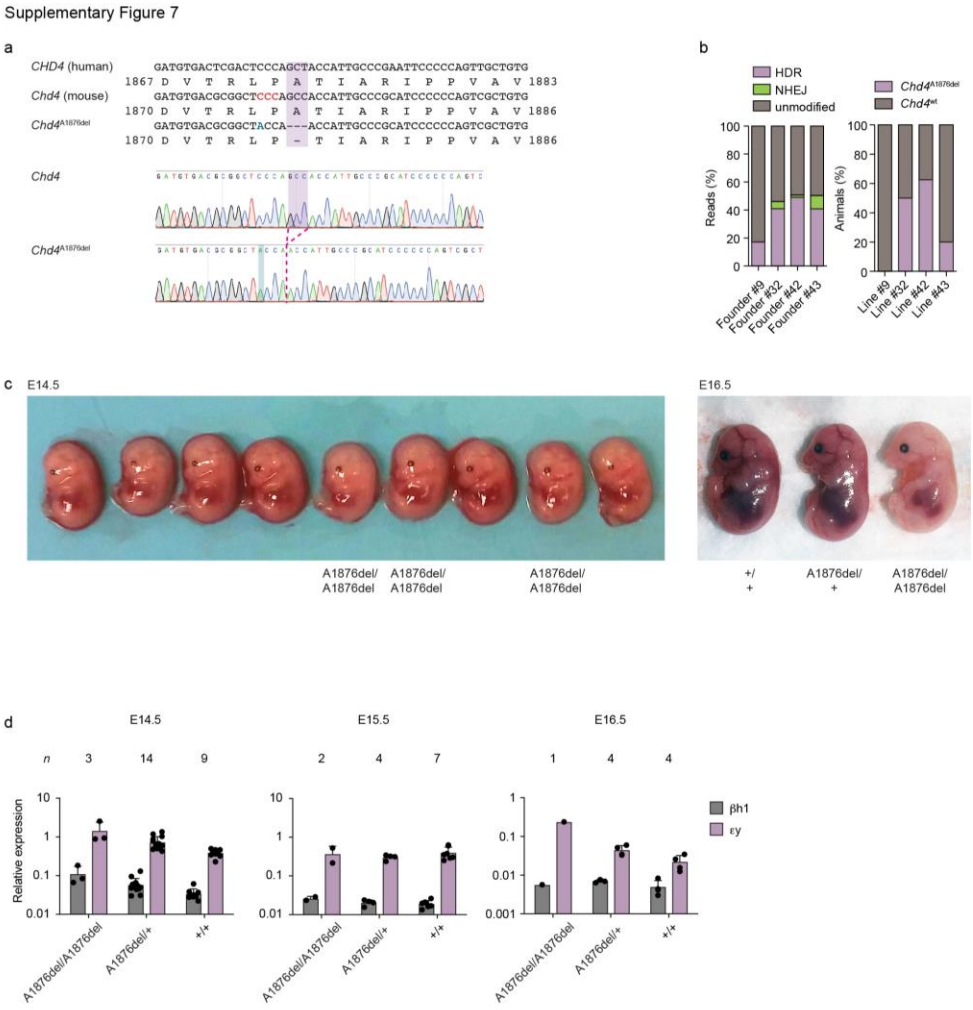


Supplementary Fig. 5: *MTA2* mutant clone characterization. a, Two step approach to generate hemizygous mutation clones of *MTA2*. In the first step, a pair of sgRNAs targeting outside the coding sequences are introduced to identify a clone in which one copy of *MTA2* is deleted (*MTA2*$^{+/-}$ cells). In the second step, individual sgRNA target specific coding sequences are introduced to identify hemizygous in-frame deletion clones. **b**, HbF^{+} cell quantification and β- and γ-globin expression measured by RT-qPCR in *MTA2*$^{+/-}$ cells, error bars show mean and SD of n = 3. **c**, Genotype of hemizygous *MTA2* in-frame deletion clones M1-M8 with mutations at indicated sequences and domains. PAM sequence restricting SpCas9 cleavage shown in red. **d**, HbF^{+} cell fraction and β- and γ-globin expression measured by RT-qPCR in control (NT) and *MTA2* hemizygous in-frame deletion clones, error bars show mean and SD of n = 2. **e-f**, Quantification of immunoprecipitation (IP) of MTA2 (**e**) and CHD4 (**f**) followed by immunoblot with indicated NuRD subunit antibody comparing relative intensity in MTA2 M4 SANT domain hemizygous in-frame deletion to control HUDEP-2 clone. Mean and SD of 3 immunoblot replicates shown. Representative immunoblots shown in **Fig. 4f-g**.

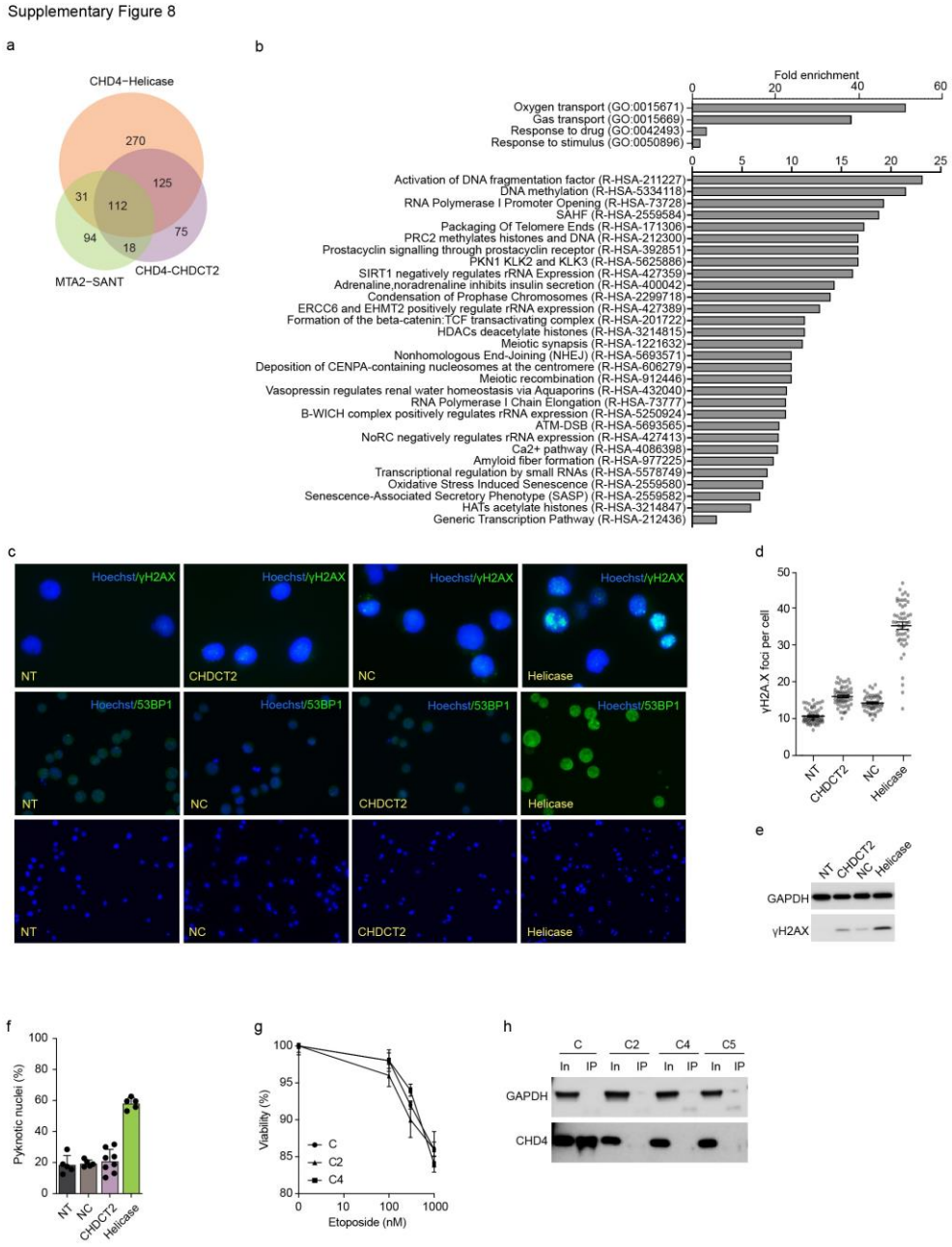
Supplementary Figure 6



Supplementary Fig. 6: *CHD4* mutation characterization. a, Two step approach to generate hemizygous mutation clones of *CHD4* in HUDEP-2 cells. In the first step, a pair of sgRNAs targeting outside the coding sequences are introduced to identify a clone in which one copy of *CHD4* is deleted (*CHD4*$^{+/-}$ cells). In the second step, individual sgRNA target specific coding sequences are introduced to identify hemizygous in-frame deletion clones. **b**, Genotype of hemizygous *CHD4* in-frame deletion clones C1-C5 with mutations at CHDCT2 domain. PAM sequence restricting SpCas9 cleavage shown in red. **c**, Cell expansion of CHDCT2 in-frame deletion clones along with parental *CHD4*$^{+/-}$ HUDEP-2 cells. Cell numbers were counted daily for 14 days. Data is mean of 3 experiments. **d**, HbF^{+} cell fraction in parental control *CHD4*$^{+/-}$ (C) and *CHD4* CHDCT2 hemizygous unedited and in-frame deletion clones (*CHD4*$^{IF/-}$). **e-g**, CD34^{+} hematopoietic stem and progenitor cells (HSPCs) electroporated with SpCas9:sgRNA RNP targeting CHDCT2 A1873 and unelectroporated (mock) cells were exposed to erythroid differentiation culture. HbF intracellular staining, hemoglobin HPLC, and CD71 and CD235a flow cytometry performed after 18 days of differentiation. Error bars show mean and SD of n = 2 for **e** and n = 4 for **f**; **g** is representative of 3 experiments**.**

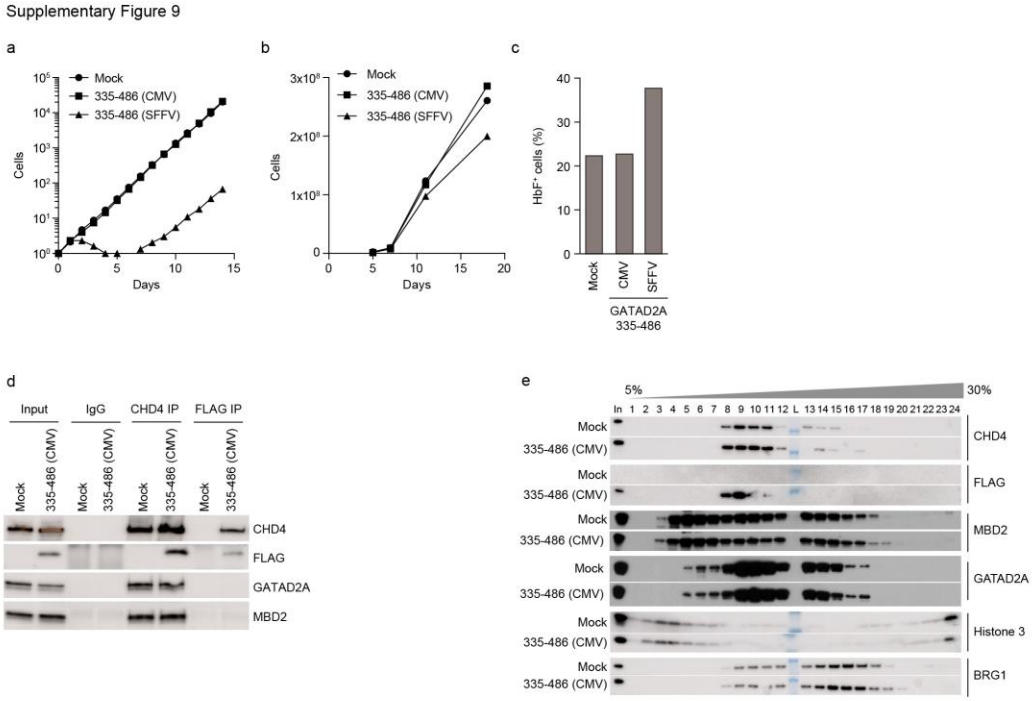


Supplementary Fig. 7: *Chd4* mutant mouse characterization. a, Sanger sequencing of wild type and *Chd4*^{A1876del} homozygous mice. The deleted nucleotides are highlighted in red and a silent C>A substitution is highlighted in blue. **b,** *Left*, distrubution of *Chd4* CHDCT2 alleles in founder mice generated by microinjection of RNPs/ssODN into oocytes and analyzed by targeted amplicon sequencing. Read percentages of unmodified, HDR and NHEJ sequences are shown. *Right*, germ line transmission of *Chd4* CHDCT2dAla founder mice. **c,** Images of *β-YAC*+;*Chd4*^{A1876del} homozygous, heterozygous and wild type, E14.5 and E16.5 embryos. *β-YAC* indicates human β-globin gene cluster transgene carried on integrated yeast artificial chromosome. **d**, Expression of mouse embryonic β-like globins βh1 and εy in fetal livers from E14.5, E15.5 and E16.5 *Chd4*$^{A1876del/+}$;*β-YAC*+ x *Chd4*$^{A1876del/+}$;*β-YAC*+ embryos. Error bars show mean and SD, *n* embryos as listed in figure.



Supplementary Fig. 8: CHD4 CHDCT2 disruption spares cytotoxicity. **a,** Gene expression impact of disruption of CHD4 helicase or CHDCT2. Venn diagram showing the overlap of differentially expressed genes by RNA-seq from SpCas9 expressing HUDEP-2 cells three days after transduction with sgRNAs targeting *CHD4* CHDCT2 A1873, *MTA2* SANT T313, and *CHD4* helicase A742 as compared to nontargeting sgRNA. **b**, Enriched GO terms (top) in 112 upregulated differentially expressed genes shared in SpCas9 expressing HUDEP-2 cells exposed to sgRNAs targeting *CHD4* CHDCT2 (A1873), *CHD4* helicase domain (A742), and *MTA2* SANT domain (T313) as compared to nontargeting control by RNA-seq. Enriched Reactome pathway terms (bottom) in the 270 uniquely upregulated differentially expressed genes in SpCas9 expressing HUDEP-2 cells exposed to sgRNAs targeting the *CHD4* helicase domain (A742) as compared to nontargeting control but not the *CHD4* CHDCT2 (A1873) or *MTA2* SANT domain (T313). SAHF: Formation of Senescence-Associated Heterochromatin Foci, PKN1 KLK2 and KLK3: A Activated PKN1 stimulates transcription of androgen receptor (androgen receptor, ATM-DBS: Recruitment and ATM-mediated phosphorylation of repair and signaling proteins at DNA double strand breaks). **c-f**. HUDEP-2 cells expressing SpCas9 transduced with lentivirus expressing sgRNA targeting *CHD4* nonconserved (NC) (S221), CHDCT2 (A1873), or helicase (A742) coding sequences or a nontargeting (NT) control and

stained with γH2AX or 53BP1 antibody and with Hoechst dye. Representative images (upper and middle panels 20X, lower panels 10X) from randomly seleced areas of a stained slide *(n=3)* (**c**), quantification of γH2AX foci formation (**d**), γH2AX immunoblot of cell lysates (**e**), and quantification of pyknotic nuclei (**f**). Each dot in **d** represents mean florescence by immunocytochemistry (ICC) of γH2AX foci in an individual cell. Pictures at 20x magnification were acquired of randomly selected areas after ICC. Fluorescent intensity in each cell was measured using ImageJ (at least 50 cells were measred in each condition). Error bars show mean flourscent intensity **g**, Viability of HUDEP-2 *CHD4* CHDCT2 hemizygous in-frame deletion clones and *CHD4*$^{+/-}$ control (C) after 24 hour exposure to indicated concentrations of etoposide. Mean ± SD from three biological replicates shown. (d,f,g) **h,** CHD4 IP followed by immunoblot with CHD4 from nuclear lysates of control and CHDCT2 hemizygous in-frame deletion clones shows the total level of CHD4 remains unchanged although the pull-down efficiency of mutant CHD4 is reduced. Representative of 2 experiments.



Supplementary Fig. 9: GATAD2A GATA ZF sequesters CHD4 from NuRD. a, Cell expansion of HUDEP-2 cells expressing GATAD2A AA335-486 driven by CMV or SFFV promoter. Cell numbers were counted daily for 14 days. **b**, Cell expansion of CD34+ HSPCs expressing GATAD2A AA335-486 driven by CMV or SFFV promoter during 18 day erythroid differentiation culture. **c**, HbF^{+} cell fraction in primary human erythroid precursors expressing GATAD2A AA335-486 driven by CMV or SFFV promoter. **d**, CHD4 IP or FLAG IP followed by immunoblot analyses in HUDEP-2 cells expressing GATAD2A AA335-486 segment driven by CMV promoter. **e**, Density sedimentation analyses using 5-30% glycerol gradients on nuclear extracts from HUDEP-2 cells expressing GATAD2A AA335-486 driven by CMV promoter. Panels **a**-**b** representative of 3 experiments and **d**-**e** of 2 experiments.

Supplementary Fig. 10: Shows full scans of westernblot images corresponding to mentioned figure number**.**

Figure 4c

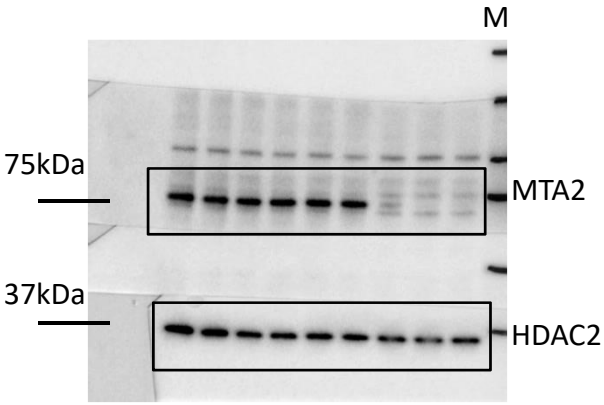


Figure 4e

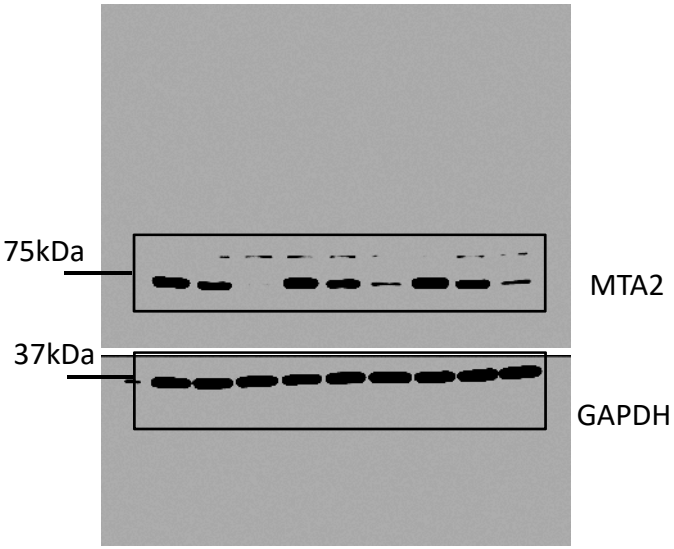


Figure 4f

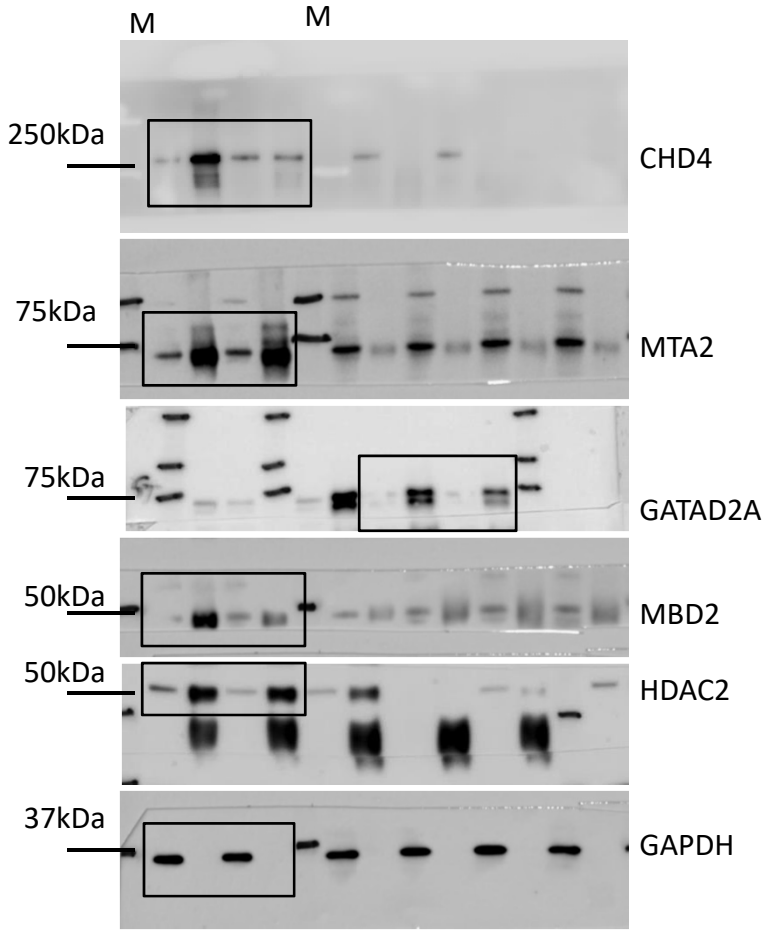


Figure 4g

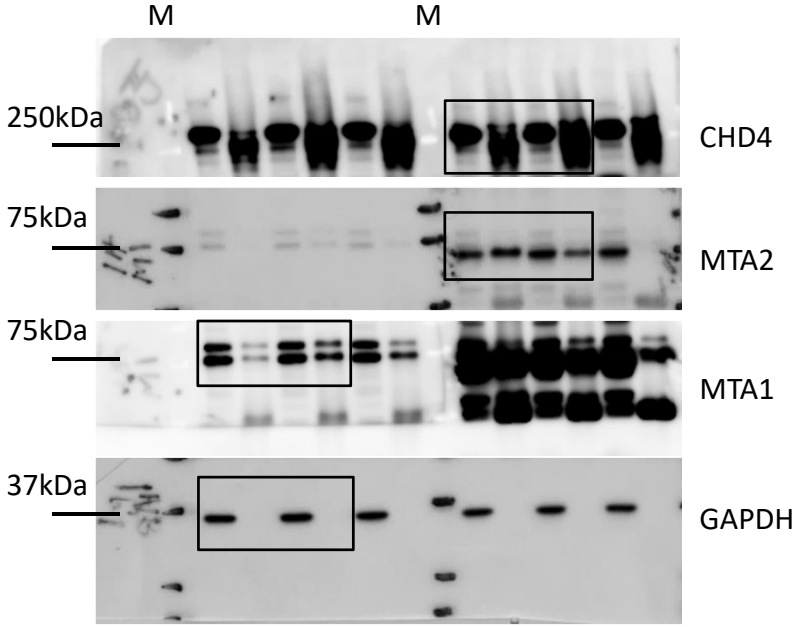


Figure 4h

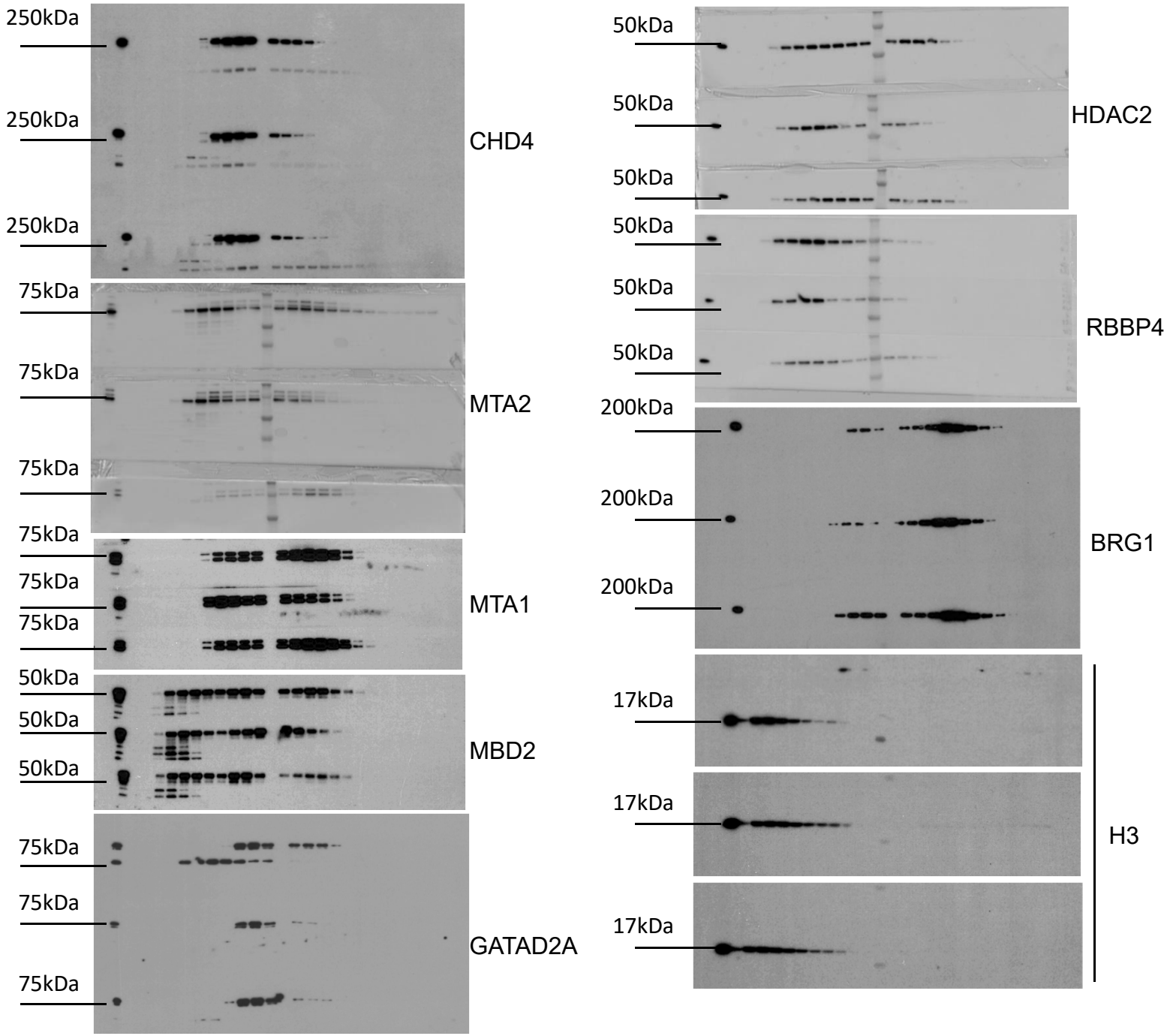

250kDa
250kDa
250kDa
CHD4
75kDa
75kDa
75kDa
MTA2
75kDa
75kDa
75kDa
MTA1
50kDa
50kDa
50kDa
MBD2
75kDa
75kDa
75kDa
GATAD2A
50kDa
50kDa
50kDa
HDAC2
50kDa
50kDa
50kDa
RBBP4
200kDa
200kDa
200kDa
BRG1
17kDa
17kDa
17kDa
H3

Figure 5e

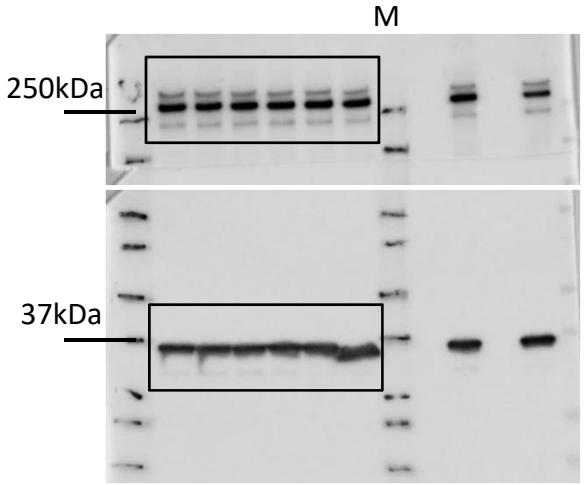


Figure 5j

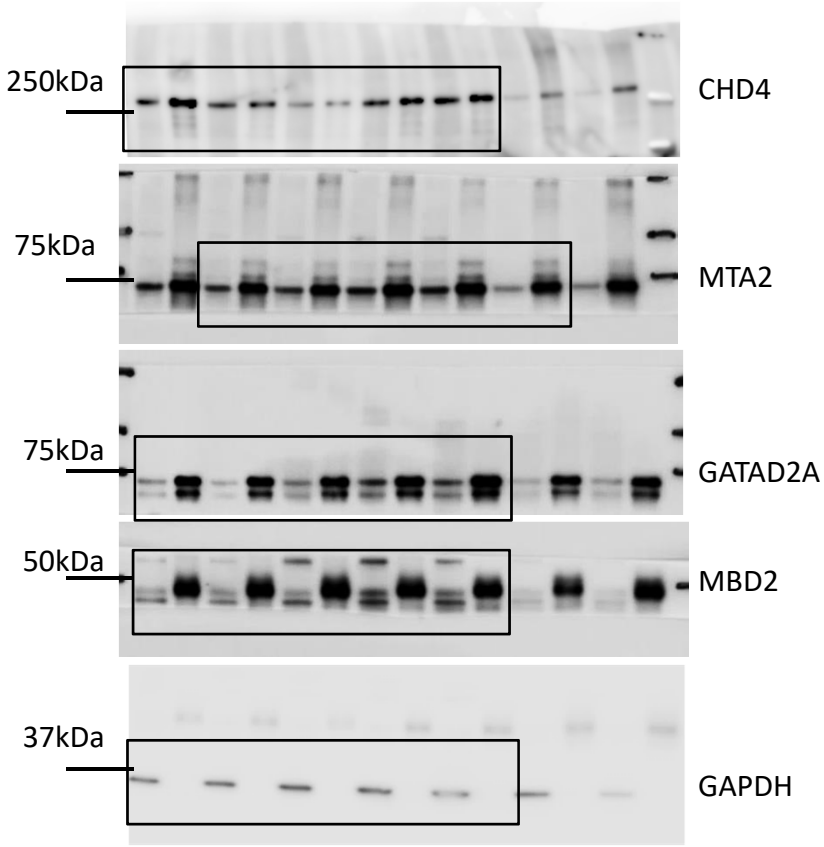


Figure 5k

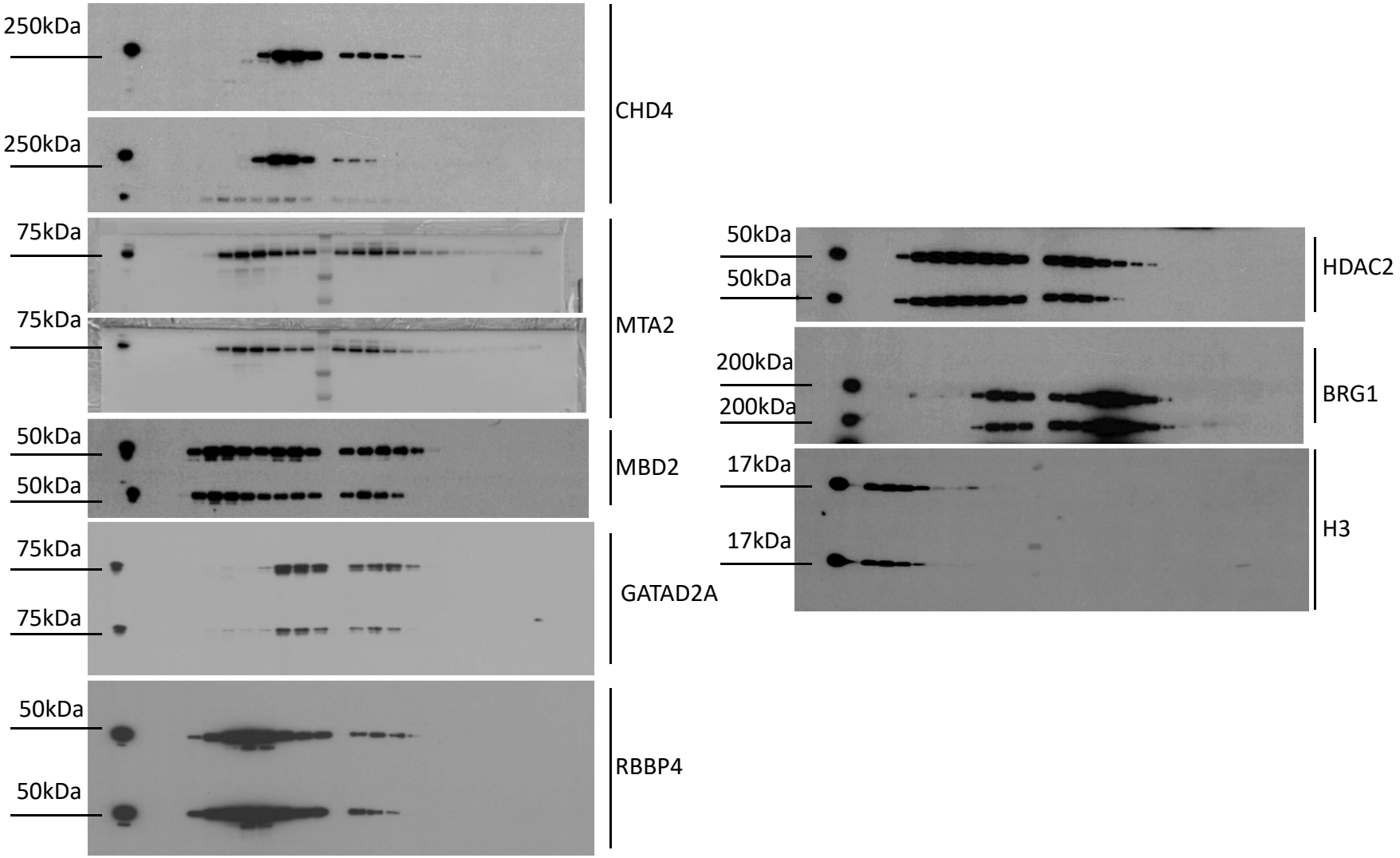

250kDa
250kDa
CHD4
75kDa
75kDa
MTA2
50kDa
50kDa
MBD2
75kDa
75kDa
GATAD2A
50kDa
50kDa
RBBP4
50kDa
50kDa
HDAC2
200kDa
200kDa
BRG1
17kDa
17kDa
H3

Figure 6b

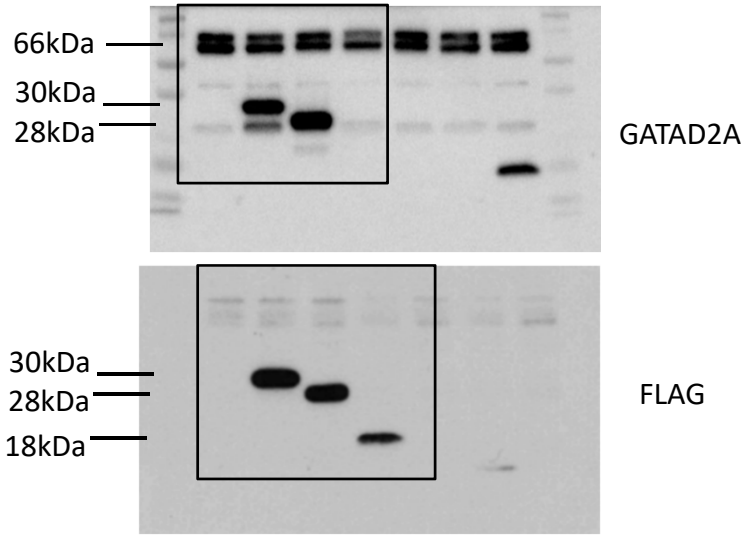


Figure 6c

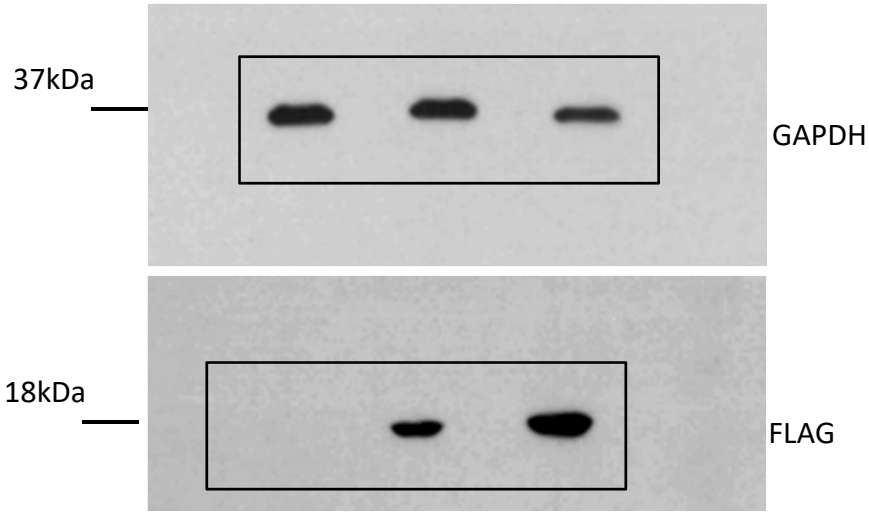


Figure 6i

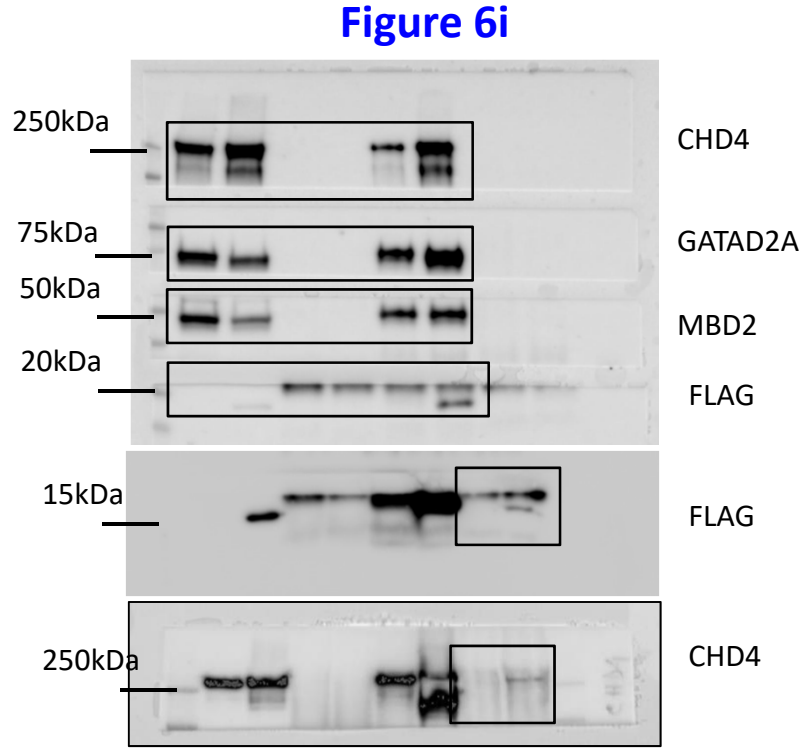


Figure 6j

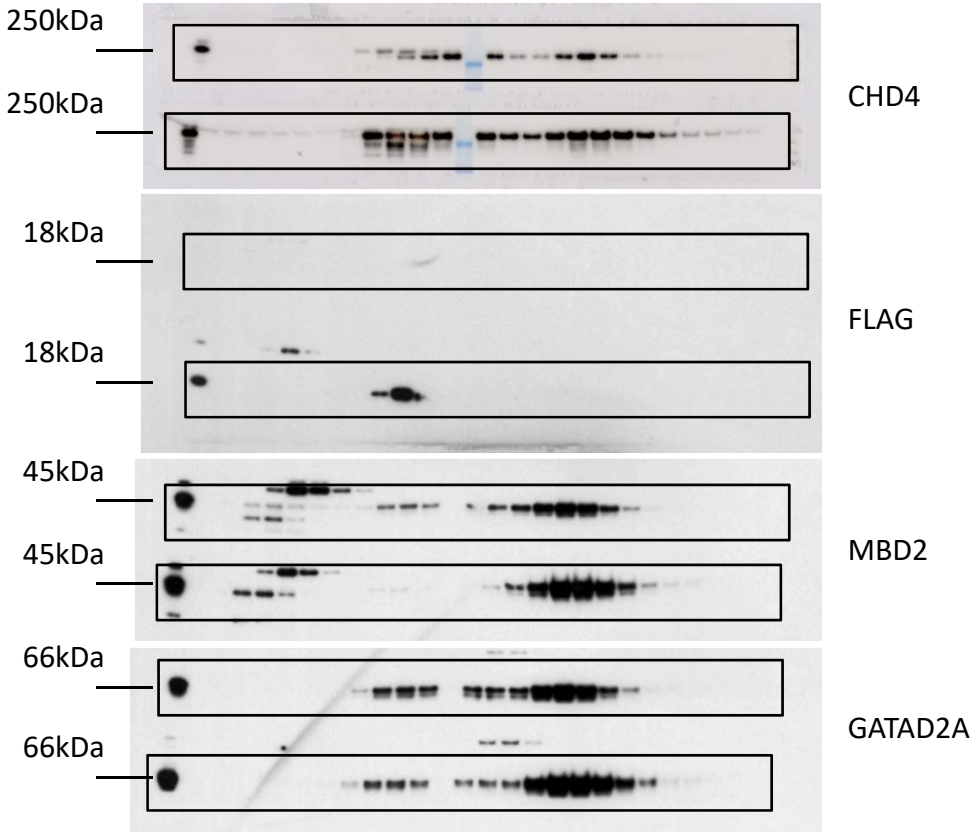

250kDa
250kDa
CHD4
18kDa
18kDa
FLAG
45kDa
45kDa
MBD2
66kDa
66kDa
GATAD2A

Figure S2 a

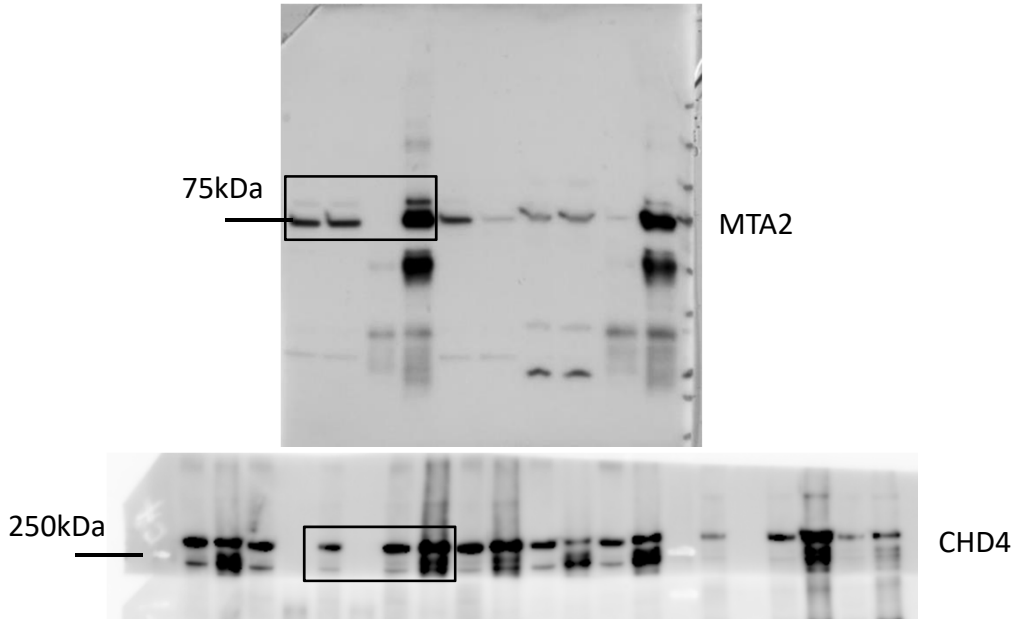


Figure S2 b

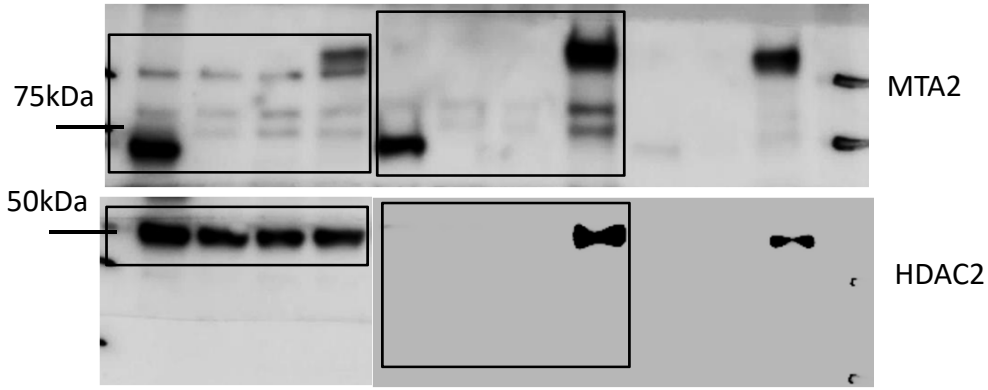


Figure S8 e

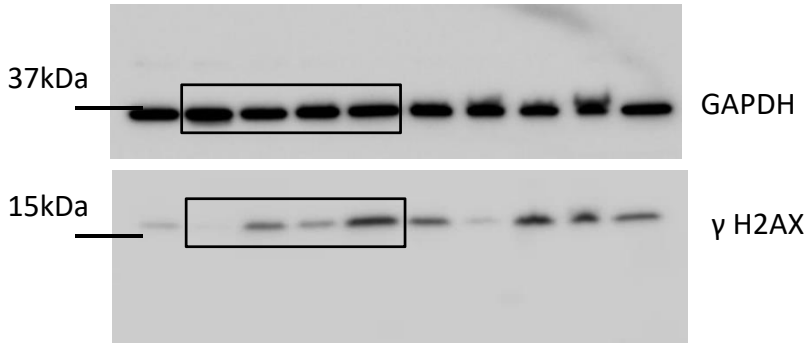


Figure S8 h

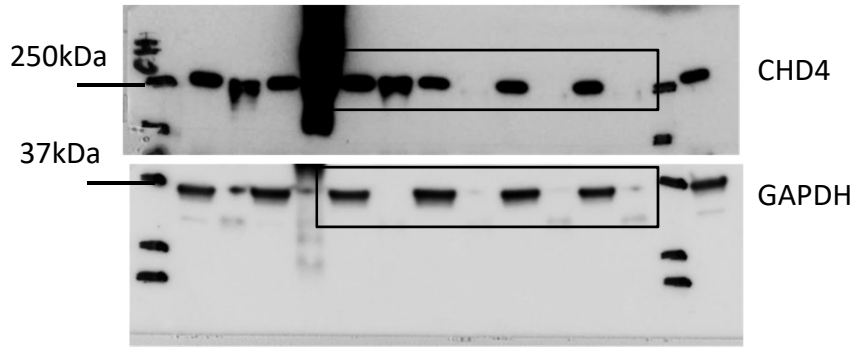


Figure S9 d

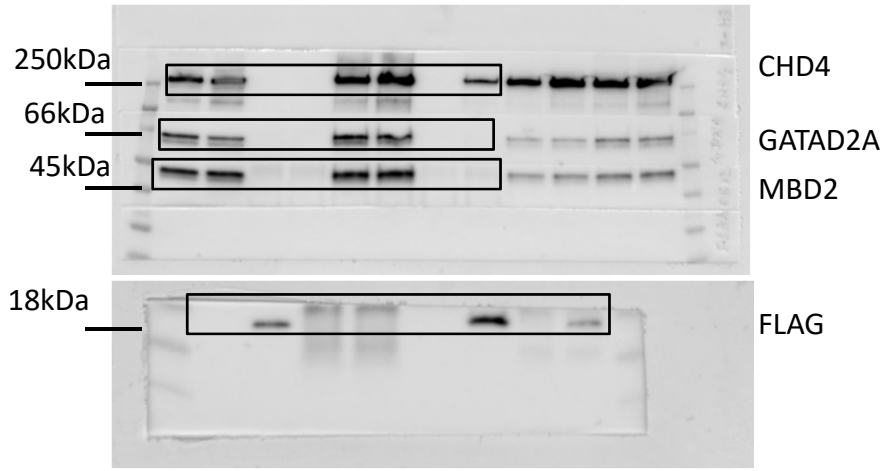
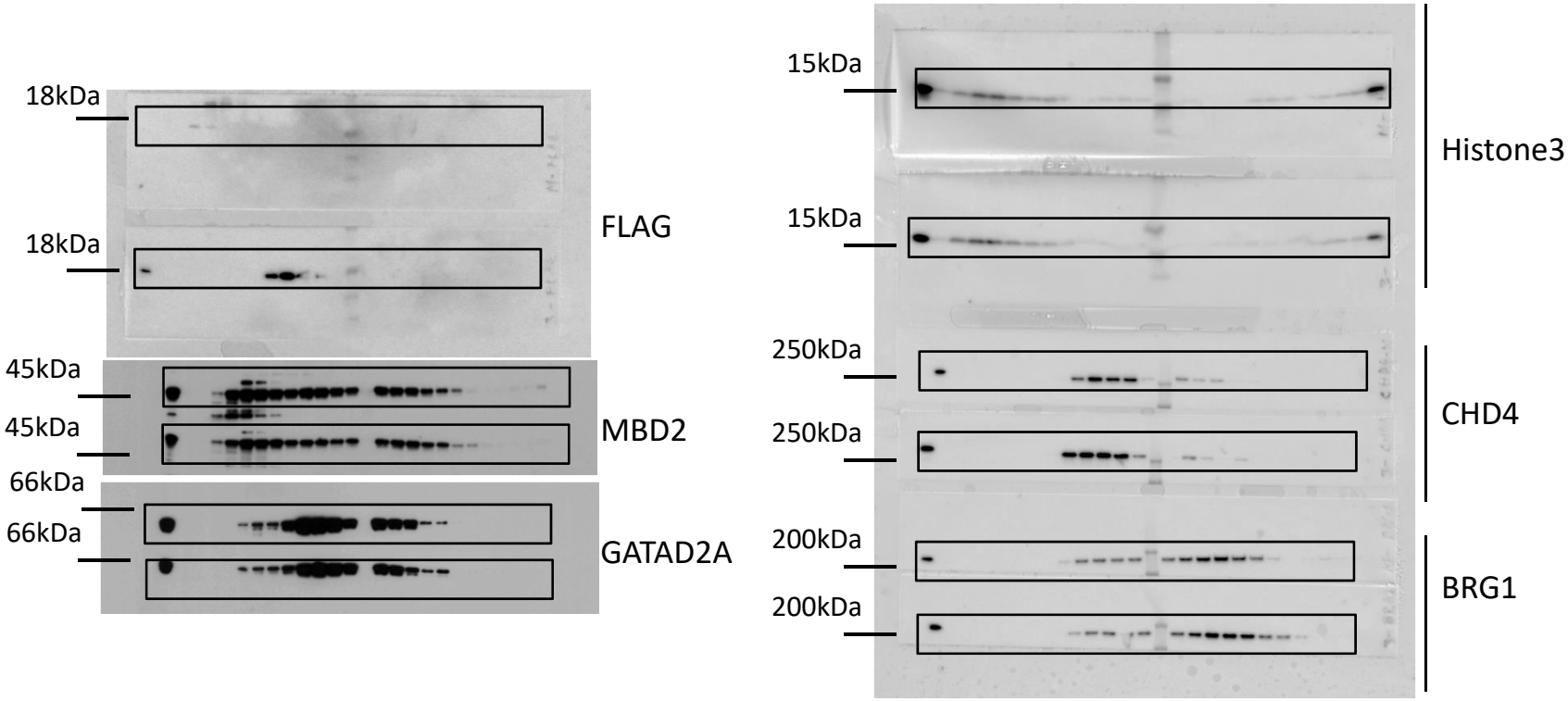


Figure S9 e


18kDa
18kDa
FLAG
45kDa
45kDa
MBD2
66kDa
66kDa
GATAD2A
15kDa
15kDa
Histone3
250kDa
250kDa
CHD4
200kDa
200kDa
BRG1