

Title

ZNF410 represses fetal globin by devoted control of CHD4/NuRD

Authors

Divya S. Vinjamur¹, Qiuming Yao^{1,2}, Mitchel A. Cole¹, Connor McGuckin¹, Chunyan Ren¹, Jing Zeng¹, Mir Hossain¹, Kevin Luk³, Scot A. Wolfe³, Luca Pinello², Daniel E. Bauer^{1,4}

¹Division of Hematology/Oncology, Boston Children's Hospital, Department of Pediatric Oncology, Dana-Farber Cancer Institute, Harvard Stem Cell Institute, Broad Institute, Department of Pediatrics, Harvard Medical School, Boston, Massachusetts 02115, USA

²Molecular Pathology Unit, Center for Cancer Research, and Center for Computational and Integrative Biology, Massachusetts General Hospital, Department of Pathology, Harvard Medical School, Boston, Massachusetts 02129, USA

³Department of Molecular, Cell and Cancer Biology, Li Weibo Institute for Rare Diseases Research, University of Massachusetts Medical School, Worcester, Massachusetts 01605, USA

⁴Correspondence: bauer@bloodgroup.tch.harvard.edu

Abstract

Major effectors of adult-stage fetal globin silencing include the transcription factors (TFs) BCL11A and ZBTB7A/LRF and the NuRD chromatin complex, although each has potential on-target liabilities for rational β -hemoglobinopathy therapeutic inhibition. Here through CRISPR screening we discover ZNF410 to be a novel fetal hemoglobin (HbF) repressing TF. ZNF410 does not bind directly to the γ -globin genes but rather its chromatin occupancy is solely concentrated at *CHD4*, encoding the NuRD nucleosome remodeler, itself required for HbF repression. *CHD4* has two ZNF410-bound regulatory elements with 27 combined ZNF410 binding motifs constituting unparalleled genomic clusters. These elements completely account for ZNF410's effects on γ -globin repression. Knockout of ZNF410 reduces CHD4 by 60%, enough to substantially de-repress HbF while avoiding the cellular toxicity of complete CHD4 loss. Mice with constitutive deficiency of the homolog Zfp410 are born at expected Mendelian ratios with unremarkable hematology. ZNF410 is dispensable for human hematopoietic engraftment potential and erythroid maturation unlike known HbF repressors. These studies identify a new rational target for HbF induction for the β -hemoglobin disorders with a wide therapeutic index. More broadly, ZNF410 represents a special class of gene regulator, a conserved transcription factor with singular devotion to regulation of a chromatin subcomplex.

1 Introduction

2

3 Despite renewed enthusiasm for novel approaches to β -hemoglobinopathies, the clinical unmet
4 need for these most common monogenic diseases remains vast¹⁻³. Induction of fetal γ -globin
5 gene expression could bypass the underlying β -globin molecular defects and ameliorate the
6 pathophysiological cascades that result in elevated morbidity and mortality. Critical regulators of
7 the switch from fetal to adult globin gene expression include the DNA-binding transcription
8 factors (TFs) BCL11A and ZBTB7A and the nucleosome remodeling and deacetylase (NuRD)
9 chromatin complex⁴⁻⁷. BCL11A and ZBTB7A each bind to unique sites at the proximal
10 promoters of the duplicated fetal γ -globin genes *HBG1* and *HBG2* and each physically interact
11 with NuRD^{5,8-10}. Although the molecular details underpinning this switch, including the precise
12 sequences bound and NuRD subcomplex members required, are increasingly understood, still
13 the feasibility to directly perturb these mechanisms through pharmacology remains uncertain.
14 One challenge is the pleiotropic molecular, cellular and organismal effects of each of the
15 aforementioned fetal hemoglobin (HbF) repressors which makes the therapeutic window
16 uncertain and the risk of undesired on-target liabilities considerable. An ideal target would have
17 a wide therapeutic index through which inhibition of function could be tolerated across a diverse
18 set of cellular contexts.

19

20 To better define additional molecular players orchestrating the developmental regulation of
21 globin gene expression, we performed a CRISPR screen focusing on putative DNA-binding TFs
22 that contribute to HbF silencing. We identified ZNF410 as a novel DNA-binding TF required for
23 HbF repression. Little was previously known about ZNF410. We show indeed this gene is
24 required for HbF silencing. Surprisingly we find that it displays a narrowly restricted pattern of
25 chromatin occupancy, not binding to the globin locus directly, but rather binding to upstream
26 elements, through an unusual set of clustered motifs, controlling the expression of the catalytic
27 NuRD subunit CHD4. We observe that ZNF410 and its mouse homolog *Zfp410* are dispensable
28 for survival to adulthood as well as normal erythropoiesis and hematopoietic repopulation.

29

30

31

32

33

34

35 Results

36

37 CRISPR screen for novel transcriptional regulators of HbF level

38

39 We performed a CRISPR screen in a primary human erythroid precursor cell line (HUDEP-2)
40 that expresses an adult-type pattern of globins to discover genes required for repression of HbF.
41 The screen targeted 1591 transcription factors and 13 genes of the NuRD complex as controls.
42 HUDEP-2 cells stably expressing SpCas9 were first generated. HUDEP-2/Cas9 cells were then
43 transduced by the sgRNA library at low multiplicity and selected for sgRNA cassette integration
44 by acquisition of puromycin resistance. Following erythroid maturation culture, cells were
45 stained for HbF expression and HbF⁺ cells (range 1.8-7%) selected by FACS (**Fig. 1a**).
46 Integrated sgRNAs were amplified from genomic DNA and counted by deep sequencing. We
47 calculated two sgRNA enrichment scores. First, sgRNA abundance was compared in HbF⁺ and
48 total cells at the end of erythroid maturation to obtain an HbF enrichment score. Second, sgRNA
49 abundance was compared in cells at the end of erythroid maturation and the starting library to
50 define a cell fitness score. Negative cell fitness scores indicate relative depletion whereas
51 positive scores indicate relative enrichment of cells bearing these sgRNAs.

52

53 As expected, we found that known HbF regulators like *BCL11A* and *ZBTB7A* showed highly
54 elevated HbF enrichment scores (**Fig. 1b**). For *BCL11A* we observed a modest negative fitness
55 score, suggesting that loss of this gene had a modest negative impact on cell accumulation in
56 vitro. For *ZBTB7A* we observed a positive fitness score, suggesting cells mutated at this gene
57 accumulated in the population, consistent with its known requirement for terminal erythroid
58 maturation¹¹. In addition, we validated prior findings that a NuRD subcomplex including *CHD4*,
59 *MTA2*, *GATAD2A*, *MBD2*, and *HDAC2* was required for HbF control⁶. Editing *CHD4* led to
60 potent HbF induction but was associated with negative cell fitness.

61

62 We observed sgRNAs targeting *ZNF410* were associated with robust HbF induction (**Fig. 1b**).
63 Unlike other regulators like *BCL11A*, *ZBTB7A*, and *CHD4*, we observed no fitness effects of
64 targeting *ZNF410*. Relatively few prior studies have investigated *ZNF410*, encoding a zinc finger
65 protein with a cluster of five C2H2 zinc fingers. It has not been previously implicated in globin
66 gene regulation or erythropoiesis. One previous study indicated that over-expression of *ZNF410*
67 in human foreskin fibroblasts led to increased expression of matrix remodeling genes *MMP1*,
68 *PAI2* and *MMP12* and *ZNF410* sumoylation extended its half-life¹². The biochemical functions

69 and biological roles of endogenous ZNF410 remain largely unexplored. Therefore we focused
70 on ZNF410 as a potentially novel regulator of HbF.

71

72 Validation in HUDEP-2 cells and primary adult erythroid precursors

73

74 We first tested the role of ZNF410 in HbF repression by targeting it in HUDEP-2 cells with
75 individual gRNAs. Upon editing in a bulk population of cells, we found induction of HbF, as
76 measured by HbF⁺ cells by intracellular flow cytometry, *HBG1/2* expression by RT-qPCR, and
77 HbF induction by HPLC (**Fig. 1c**). We generated 3 single cell derived HUDEP-2 *ZNF410*
78 biallelic KO clones. In each clone, the fraction of HbF⁺ cells was elevated. Upon re-expression
79 of ZNF410, HbF was partially silenced, consistent with a causal role of ZNF410 in repressing
80 HbF (**Fig. 1d**).

81

82 We next examined the role of ZNF410 in HbF repression in primary erythroblasts derived from
83 erythroid culture of adult mobilized peripheral blood CD34⁺ HSPCs. Using cells from 3
84 independent donors, we found that targeting *ZNF410* by 3xNLS-SpCas9:sgRNA RNP
85 electroporation produced >99% indels with a +1 insertion allele (**Fig. 1e**). *ZNF410* targeted
86 erythroblasts displayed normal erythroid maturation based on immunophenotype and
87 enucleation, yet robust induction of HbF from a median of 5.5% in mock to 21.1% in *ZNF410*
88 targeted samples (**Fig. 1e**, $p < 0.0001$).

89

90 *ZNF410 is a DNA-binding protein with highly restricted chromatin occupancy*

91

92 We performed dense mutagenesis of *ZNF410* to identify critical minimal sequences required for
93 function. In this experimental design, heightened HbF enrichment scores indicate sequences
94 where not only frameshift but also in-frame mutations may be associated with loss-of-function⁶.
95 We observed heightened HbF enrichment scores especially when targeting sequences from
96 exons 6-9 encoding the cluster of five C2H2 zinc fingers of *ZNF410* (**Fig. 2a**). This dependence
97 on its putative DNA binding domain suggested that the DNA-binding function of ZNF410 might
98 be important for its role in HbF repression.

99

100 We examined the chromatin occupancy of ZNF410 by conducting CUT&RUN, an approach to
101 studying protein-DNA interactions in situ without fragmentation or cross-linking¹³. Initially we
102 used an HA antibody to probe for epitope tagged ZNF410 in HUDEP-2 cells. Known HbF

103 repressing TFs like BCL11A and ZBTB7A act by binding to proximal promoter elements at the
104 fetal γ -globin (*HBG1* and *HBG2*) genes. We did not observe any chromatin occupancy of
105 ZNF410 at the α -globin (*HBA1* and *HBA2*) or β -globin (*HBB*) gene clusters (**Supp. Fig. 1a, b**).
106 Unlike typical DNA binding transcription factors which show thousands of binding sites genome
107 wide, ZNF410 showed highly restricted chromatin occupancy. With standard peak calling
108 parameters, we found 49 peaks, but most of these had marginal enrichment of ZNF410-HA
109 signal compared to an IgG control. The top two peaks were found at the *CHD4* locus, one at the
110 promoter (57-fold enrichment) and the other 6 kb upstream at a region of open chromatin (77-
111 fold enrichment, **Fig. 2b, 2d**). This latter element we subsequently refer to as the *CHD4* -6 kb
112 enhancer. The third most enriched peak was in *CHD4* intron 2, with ~10 fold enrichment (**Fig.**
113 **2b, 2d**). The fourth most enriched peak was in *TIMELESS* intron 1, with ~10 fold enrichment,
114 around accessible chromatin at sequences bearing an LTR element (**Fig. 2b, Supp. Fig. 2a**).
115

116 Subsequently we used a ZNF410 antibody to probe for endogenous ZNF410 with CUT&RUN,
117 both in HUDEP-2 cells and in CD34+ HSPC derived erythroid precursors (**Fig. 2d**). In both
118 cases, we found that ZNF410 chromatin occupancy was highly restricted to *CHD4*. In HUDEP-2
119 cells, only 5 total peaks were identified, the top 4 of which were at the *CHD4* promoter and
120 *CHD4* -6 kb enhancer (**Fig. 2d, Supp. Fig. 2b**). The 5th peak was at intronic sequences of
121 *DPY19L3* bearing an LTR element (5-fold enrichment) (**Supp. Fig. 2d**). In CD34+ HSPC-
122 derived erythroid precursors, only 5 total peaks were identified, all of which were at the *CHD4*
123 promoter or *CHD4* -6 kb enhancer, with no other genomic sites of ZNF410 occupancy found
124 (**Fig. 2d, Supp. Fig. 2c**).
125

126 The ZNF410 binding motif has previously been described by high-throughput SELEX using
127 expression of the DNA-binding domain in 293FT cells¹⁴. We observed a striking cluster of these
128 motifs at *CHD4*, with numerous motif instances found at both the promoter (16 motifs) and the -
129 6 kb enhancer (11 motifs, **Fig. 2c, d**). We scanned the genome for the ZNF410 binding motif,
130 dividing the genome into 3 kb windows. 4306 genomic windows had 1 motif instance and 16
131 windows had 2 motif instances (**Fig. 2c**). Only 3 windows had more than 2 motif instances, of
132 which 2 were the aforementioned *CHD4* elements. We observed 6 motif instances within a
133 window at *GALNT18* intron 1, although we observed neither ZNF410 occupancy nor chromatin
134 accessibility at this locus in erythroid precursors (**Supp. Fig. 1c**).
135

136

137 *ZNF410 regulates HbF through CHD4*

138

139 These results suggested that ZNF410 exhibits singular binding to *CHD4*. We performed RNA-
140 seq of HUDEP-2 cells edited at *ZNF410* to measure gene expression changes (**Fig. 3a**). Based
141 on \log_2 fold change >1 and adjusted p-value <0.01 , there were 63 differentially expressed
142 genes. *CHD4* was the most significantly downregulated gene upon *ZNF410* editing (L2FC -1.07,
143 $p_{\text{adj}} 2.27 \times 10^{-43}$). *HBG2* was the 4th most significantly upregulated gene (L2FC 2.35, $p_{\text{adj}} 5.93 \times 10^{-25}$).
144 Gene set enrichment analysis showed that genes differentially expressed after *ZNF410*
145 editing were enriched in those differentially expressed after *CHD4* editing (for upregulated
146 genes, NES 1.61, q 0.05; for downregulated genes, NES -1.39, q 0.09; **Fig. 3b, Supp. Fig. 3a**).
147 The expression of *ZNF410* and *CHD4* were significantly correlated across 54 human tissues
148 from the GTEx dataset¹⁵ (**Supp. Fig. 3b**, Pearson correlation, $r 0.77$, $p < 0.0001$). We evaluated
149 a repository of genome-wide CRISPR KO screen data spanning 558 cell lines to identify genes
150 with a similar pattern of cellular dependency as *ZNF410*^{16,17}. We found that *CHD4* was the most
151 similarly codependent gene across cell lines, indicating a pervasive relationship between
152 *ZNF410* and *CHD4* (**Fig. 3c**). These results suggest that a major function of ZNF410 across
153 numerous cellular contexts appears to be control of CHD4 expression.

154

155 We validated the changes in CHD4 expression after *ZNF410* editing by RT-qPCR in both
156 HUDEP-2 cells and primary erythroid precursors derived from CD34+ HSPCs. We found that
157 CHD4 mRNA expression was reduced by 57% after *ZNF410* editing (**Fig. 3d**, $p < 0.01$). To test
158 the requirement of ZNF410 binding for CHD4 expression, we generated HUDEP-2 cell clones in
159 which the two upstream ZNF410 motif clusters at *CHD4* were both deleted by paired genomic
160 cleavages (**Fig. 3e, Supp. Fig. 3c**). We isolated 4 biallelically deleted HUDEP-2 clones. We
161 found that CHD4 expression decreased by 56-79% after deletion of the upstream elements,
162 similar to the decrease observed after editing *ZNF410* itself. Consistent with reduced expression
163 of CHD4, γ -globin was induced (**Fig. 3f, Supp. Fig 3d, 3e**). No change in CHD4 expression was
164 observed upon *ZNF410* editing in the absence of the upstream elements, suggesting that the
165 control of CHD4 expression requires these elements. We did not observe further γ -globin
166 induction in *CHD4* upstream element deleted cells upon *ZNF410* editing (**Fig. 3g, Supp. Fig.**
167 **3e**). In contrast, γ -globin increased in these same cells upon *ZBTB7A* editing, indicating the
168 cells were competent for further γ -globin induction. We performed RNA-seq of *CHD4* $\Delta 6.7$ kb
169 element deleted cells after *ZNF410* editing (**Fig. 3h**). In contrast to HUDEP-2 cells, we only
170 observed 2 differentially expressed genes after *ZNF410* editing in *CHD4* $\Delta 6.7$ kb element

171 deletion cells, consistent with our prior results that nearly all gene expression changes found
172 after *ZNF410* editing are due to changes in CHD4 expression. Together these results suggest
173 that ZNF410 represses γ -globin exclusively by binding upstream elements and trans-activating
174 CHD4.

175

176 *ZNF410 is a non-essential gene*

177

178 ZNF410 and its mouse ortholog Zfp410 share 94% amino acid identity, including 98% at the
179 cluster of 5 ZnFs¹². We performed CUT&RUN to investigate the chromatin occupancy of
180 endogenous Zfp410 in a mouse erythroid cell line (MEL cells). Similar to results in human
181 erythroid precursors, we observed that genomic enrichment of Zfp410 binding was highly
182 restricted to the *Chd4* locus, with 77-fold enrichment at the promoter and 45-fold enrichment at
183 the *Chd4* -6 kb enhancer, each overlapping accessible chromatin regions and motif clusters
184 (**Fig. 4a, b**). The third most enriched site for Zfp410 occupancy was at the promoter of
185 *Hist1h2bl*, with ~14 fold enrichment, although no motifs were observed at this site (**Supp. Fig.**
186 **4a**). To evaluate the requirement of ZNF410 in normal development and homeostasis, we
187 investigated mice with a loss-of-function allele of the mouse ortholog *Zfp410*. We obtained
188 mouse embryonic stem cells that are heterozygous for a *Zfp410* gene trap allele (Gt) from the
189 European Mouse Mutant Cell Repository (EuMMCR). The targeting cassette was inserted to
190 intron 5 to disrupt expression of full-length Zfp410 (**Supp. Fig. 4b**). Of note, exons 6-9 encode
191 the five ZnFs (**Supp. Fig. 4c**). We derived heterozygous mice with germline transmission of this
192 allele. Although the sample size is currently small, we observed 6 *Zfp410*^{Gt/Gt} homozygotes out
193 of 20 live births from *Zfp410*^{+ /Gt} heterozygote intercrosses, consistent with expected Mendelian
194 transmission (**Fig. 4c**). *Zfp410* expression was reduced by >98% in *Zfp410*^{Gt/Gt} homozygous
195 mouse whole blood (**Fig. 4d**). The *Zfp410*^{Gt/Gt} homozygotes showed moderately reduced body
196 weight compared to heterozygotes or wt mice (**Fig. 4e**), but otherwise appeared healthy and
197 active. Analysis of complete blood counts showed apparently unremarkable hematologic
198 parameters in *Zfp410*^{Gt/Gt} homozygous mice, including no evidence of anemia or hemolysis (**Fig.**
199 **4f**). The absence of a severe phenotype of constitutive *Zfp410* loss-of-function is notable in
200 comparison to other HbF regulators. For example, *Bcl11a* deficient mice experience perinatal
201 lethality¹⁸, *Zbtb7a* deficient mice mid-gestation embryonic lethality due to anemia¹¹, and *Chd4*
202 deficient mice pre-implantation embryonic lethality¹⁹. Together these results suggest that
203 ZNF410 is an evolutionarily conserved HbF repressor that is not essential for vertebrate
204 survival.

205

206 *ZNF410 appears dispensable for human erythropoiesis and hematopoiesis*

207

208 To evaluate the role of *ZNF410* in human hematopoiesis, we performed gene editing of *ZNF410*
209 in primary human hematopoietic stem and progenitor cells (HSPCs). We electroporated 3xNLS-
210 SpCas9 and sgRNA as ribonucleoprotein (RNP) to CD34+ HSPCs from two healthy donors and
211 achieved >99% indels (**Fig. 5a, b**). Since all of these measured indels were +1 insertions,
212 biallelic *ZNF410* knockouts comprised nearly all cells in the population. To test the role of
213 *ZNF410* more broadly in hematopoiesis, we performed xenotransplantation of edited HSPCs to
214 immunodeficient NBSGW mice (**Fig. 5a**). NBSGW mice support multilineage (lymphoid, myeloid
215 and erythroid) human engraftment in absence of conditioning therapy²⁰. After 16 weeks we
216 analyzed bone marrow from engrafted recipients. We observed similar human hematopoietic
217 engraftment for *ZNF410* edited HSPCs compared to mock control xenografts (**Fig. 5c**). *ZNF410*
218 +1 insertion frameshift indels were observed at >99% frequency in total BM human
219 hematopoietic cells similar to the input cell product (**Fig. 5b**). A comparable distribution of
220 multilineage hematopoietic reconstitution was found in control and *ZNF410* edited recipients,
221 including B-lymphocyte, T-lymphocyte, granulocyte, monocyte, HSPC and erythroid
222 contributions (**Fig. 5d, e**). We found that CHD4 expression was decreased by ~60% in human
223 erythroid cells sorted from bone marrow, similar to *in vitro* results (**Fig. 5f**). The level of HbF as
224 measured by HPLC from engrafting human erythrocytes was ~2.5% in controls and ~17% in
225 *ZNF410* edited recipients (**Fig. 5g**).

226

227 For comparison, we also performed xenotransplant experiments with *BCL11A* and *ZBTB7A*
228 edited HSPCs. Consistent with the known role of *BCL11A* in supporting HSC self-renewal (and
229 unlike the selective erythroid impact of *BCL11A* erythroid enhancer editing²¹), we observed
230 reduced human chimerism in the bone marrow of recipients of *BCL11A* exon edited HSPCs
231 after 16 weeks, reduced *BCL11A* edits compared to input cell product, and reduced fraction of
232 frameshift alleles compared to total edits (**Supp. Fig. 5a, b**). For *ZBTB7A*, the fraction of
233 engrafting human hematopoietic cells was similar to controls but the gene edits were reduced
234 compared to input cell product (**Supp. Fig. 5a, b**). During erythroid maturation culture, *ZBTB7A*
235 edited HSPCs showed impaired terminal erythroid maturation potential based on
236 immunophenotype and enucleation frequency, in contrast to *ZNF410* edited cells (**Supp. Fig.**
237 **5c, d**). Together these results suggest HSPCs bearing *BCL11A* and *ZBTB7A* loss-of-function

238 alleles are under negative selective pressure during hematopoietic repopulation and
239 erythropoiesis unlike *ZNF410* edited cells.

240

241 **Discussion**

242

243 The advances in knowledge of the molecular details of hemoglobin switching have begun to
244 bear fruits in the form of novel autologous therapies¹. A host of HSC-based therapies that
245 reduce the expression of BCL11A in erythroid cells or prevent its binding to *HBG1/2* promoter
246 sequences are in clinical trials or late preclinical development. However the clinical unmet need
247 remains vast, with ~300,000 infants estimated to be born each year worldwide with sickle cell
248 disease and tens of thousands more with severe β -thalassemia. The feasibility in terms of cost
249 and infrastructure to scale up autologous cell-based therapies remains uncertain. Furthermore
250 the toxicity of myeloablative transplantation will likely render these therapies out of reach to
251 many patients.

252

253 The most realistic near-term hope to develop scalable therapies to address the root cause of
254 these diseases would be through pharmacotherapy. Drugs that could interrupt molecular
255 vulnerabilities required for adult erythroid cells to maintain fetal globins in a silenced state are
256 greatly needed. These could complement or even supplant existing treatments like
257 hydroxyurea²². BCL11A itself would certainly represent a preeminent target. Its roles in
258 erythropoiesis besides HbF silencing are modest. However BCL11A plays essential roles in
259 various hematopoietic lineages, including in B-lymphocytes, dendritic cells and hematopoietic
260 stem cells^{18,23-25}. In addition, it has functions beyond hematopoiesis not only in the central
261 nervous system but also in breast and pancreatic cells^{26,27}. Another exciting target would be
262 ZBTB7A given its potent role in HbF repression. However ZBTB7A is required for terminal
263 erythropoiesis and germinal center B cell maturation and plays important roles in T-
264 lymphocytes, osteoclasts and HSCs²⁸. A specific NuRD subcomplex including CHD4,
265 GATAD2A, MBD2, MTA2 and HDAC2 is required for HbF silencing^{6,29}. Targeting NuRD
266 including key protein-protein interactions appears promising but would need to navigate the
267 numerous gene expression programs that depend on this chromatin complex. For most of the
268 known HbF regulators, their pleiotropic roles could yield potential on-target liabilities with narrow
269 therapeutic index even if rational targeting approaches could be devised.

270

271 Here we identify ZNF410 as a novel HbF repressor that acts specifically to enhance the
272 expression of CHD4. Complete knockout of ZNF410 is well-tolerated, apparently since the
273 remaining level of CHD4 is sufficient to maintain cellular functions. *Zfp410* mutant mice survive
274 to adulthood and *ZNF410* knockout HSPCs demonstrate no defects in erythroid maturation or
275 hematopoietic reconstitution. Traditionally TFs have been considered undruggable targets.
276 However the example of small molecules binding and resulting in specific degradation of zinc
277 finger proteins like IKZF1 has encouraged the development of ligands to modulate DNA-binding
278 factors^{30,31}.

279
280 ZNF410 appears to represent a special form of gene regulator. Conventional DNA-binding TFs
281 bind and directly control the expression of thousands of genomic targets. In contrast, ZNF410
282 shows unique binding to *CHD4*. This exquisite specificity appears to be achieved through a
283 remarkable clustering of 27 ZNF410 binding sites at the *CHD4* promoter and -6 kb enhancer, a
284 density unlike anywhere else in the genome. Both *ZNF410* itself and its two target elements at
285 *CHD4* are highly conserved across vertebrates. Despite thousands of ZNF410 motifs across the
286 genome, we detected minimal ZNF410 occupancy at these sites. The absence of detectable
287 ZNF410 occupancy or chromatin accessibility even at *GALNT18* intron 1 with 6 clustered motif
288 instances suggests that motif clusters may be necessary but insufficient for ZNF410 binding.
289 Another example of clustered homotypic TF binding sites associated with gene control is the
290 binding of ZFP64 to the *MLL* gene promoter, activating the expression of the chromatin
291 regulator MLL, although in this case ZFP64 shows a limited set of additional direct target
292 genes³². CHD4 is an especially abundant nuclear protein in erythroid precursors³³.
293 Haploinsufficiency of *CHD4* (or *MLL*) causes impaired intellectual development and congenital
294 anomalies, suggesting that chromatin regulatory complexes must be maintained at precise
295 levels to maintain proper gene regulation, particularly during development³⁴⁻³⁶. There are more
296 than a thousand putative DNA-binding TFs, for many of which the genomic binding sites and
297 regulons remain poorly characterized³⁷. ZNF410 may be emblematic of a class of TFs relying on
298 homotypic motif clusters³⁸ with limited gene targets that are devoted to maintenance of core
299 cellular programs.

300
301 In summary, here we identify ZNF410 as a dispensable TF that represses HbF level in adult-
302 stage erythroid precursors by devoted maintenance of NuRD subcomplex levels through binding
303 a singular cluster of sequences upstream of *CHD4*.

304

305 **Methods**

306

307 Cell culture

308 HUDEP-2 cells³⁹ were cultured as previously described⁴⁰. Expansion phase medium for
309 HUDEP-2 cells consists of SFEM (Stemcell Technologies #09650) base medium supplemented
310 with 50 ng/ml recombinant human SCF (R&D systems #255-SC), 1 µg/ml doxycycline (Sigma
311 Aldrich #D9891), 0.4 µg/ml dexamethasone (Sigma Aldrich #D4902), 3 IU/ml EPO (Epoetin Alfa,
312 Epogen, Amgen) and 2% Penicillin-Streptomycin solution (10,000 U/mL stock). Erythroid
313 differentiation medium (EDM) for HUDEP-2 cells consists of Iscove's Modified Dulbecco's
314 Medium (IMDM, ThermoFisher #12440053) supplemented with 1% L-Glutamine (Gibco
315 #25030081), 330 µg/mL human holo-Transferrin (Sigma #T0665), 10 µg/mL human insulin
316 (Sigma #I9278), 2 IU/mL heparin (Sigma #H3149), 5% inactivated human plasma (Octaplas,
317 blood group AB, Octapharma), 3 IU/mL EPO (Epoetin Alfa, Epogen, Amgen) and 2% Penicillin-
318 Streptomycin solution (10,000 U/mL stock). EDM-2 medium for HUDEP-2 cells is EDM
319 supplemented with 100 ng/ml human SCF and 1 µg/ml doxycycline. CD34+ HSPCs from adult
320 mobilized peripheral blood from de-identified healthy donors were purchased from Fred
321 Hutchinson Cancer Research Center, Seattle, Washington. Upon thawing, CD34+ HSPCs were
322 resuspended in X-VIVO 15 medium (Lonza #04-380Q) containing 50 ng/ml recombinant human
323 Flt-3 ligand (Peprotech #300-19), 100 ng/ml recombinant human TPO (Peprotech #300-18) and
324 100 ng/ml recombinant human SCF (R&D systems #255-SC) (referred to as X-VIVO complete
325 medium). Erythroid differentiation was performed in 3 phases as previously described⁴¹. Mouse
326 erythroleukemia cells, MEL-745A cl. DS19, were cultured in RPMI 1640 medium supplemented
327 with 10% fetal bovine serum and 1% Penicillin-Streptomycin.

328

329 sgRNA library screening

330 For library screening, HUDEP-2 cells with stable expression of LentiCas9-Blast (Addgene
331 plasmid 52962) were transduced at a low multiplicity of infection (MOI) with virus containing
332 sgRNA library cloned in lentiGuide-Puro (Addgene plasmid 52963) to ensure that most cells
333 received only one sgRNA in expansion phase medium⁴². The sgRNA library included 18,020
334 gRNA overlapping those in GeCKOv2⁴³ and Avana⁴⁴ libraries targeting 1591 transcription
335 factors and 13 genes of the NuRD complex as controls. After 24 hours, cells were transferred to
336 and cultured in erythroid differentiation medium for 14 days. At the end of erythroid culture, cells
337 were processed for intra-cellular HbF staining using Fetal Hemoglobin Monoclonal Antibody
338 (HBF-1) conjugated to FITC (Thermo Fisher #MHFH01), and HbF+ cells were sorted by FACS

339 as previously described⁶. Genomic DNA was extracted from the total cell population and from
340 HbF+ sorted cells and deep sequenced to identify guide RNAs with enrichment in the HbF+ pool
341 as previously described⁶. Briefly, two step PCR was performed to amplify sgRNA cassette from
342 genomic DNA, using Herculase II Fusion DNA polymerase (Agilent #600677). Multiple reactions
343 of the first PCR were set up for each sample in order to maximize genomic DNA input up to
344 1000 cell equivalents per sgRNA. After the first PCR, all reactions for each sample were pooled
345 and 1 ul of this mix used as input for the second PCR reaction which was performed in
346 duplicate. Illumina adaptor and sample barcodes added in the second PCR. Primers for the
347 second PCR were of variable length to increase library complexity⁴². Sequences of PCR primers
348 can be found in the **Supplementary Table**. Amplicons obtained from the second PCR were
349 purified by gel extraction and quantified using the Qubit dsDNA HS assay kit (Invitrogen
350 #Q32851). Single-end 75 bp sequencing was performed on the NextSeq 500 platform by the
351 Molecular Biology Core Facilities at Dana-Farber Cancer Institute. Candidate HbF regulators
352 were identified by analyzing sequencing data using the model-based analysis of genome-wide
353 CRISPR-Cas9 Knockout (MAGeCK) computational tool⁴⁵.

354

355 Validation in HUDEP-2 cells

356 Candidate HbF regulators identified by the screen were validated in arrayed format in HUDEP-2
357 cells. HUDEP-2 cells with stable expression of lentiCas9-Blast were transduced with sgRNA
358 cloned in lentiGuide-Puro in expansion phase medium. 24 hours after transduction, cells were
359 cultured in EDM2 for 4 days, EDM with doxycycline for 3 days and EDM without doxycycline for
360 2 days as previously described⁴⁰. These culture conditions result in differentiation of normal
361 HUDEP-2 cells to orthochromatic erythroblasts. At the end of erythroid differentiation cells were
362 divided into aliquots and processed for intra-cellular HbF staining, RNA isolation and
363 hemoglobin HPLC. In addition to mock treated cells, non-targeting sgRNAs or sgRNAs targeting
364 either *AAVS1* or a functionally neutral locus on chr2 (so-called “safe targeting” sgRNA)⁴⁶ were
365 used as experimental controls as indicated in each figure legend. For RNA sequencing
366 experiments HUDEP-2 cells were cultured in expansion phase medium for 6 days after
367 transduction or electroporation. RNA was isolated using Trizol according to the manufacturer’s
368 protocol (Thermo Fisher #15596026). Purified RNA was treated with DNase I. mRNA libraries
369 were prepared and sequenced by the Molecular Biology Core Facilities at Dana-Farber Cancer
370 Institute.

371

372

373 Generation of *ZNF410* null HUDEP-2 cell clones

374 The entire coding sequence of *ZNF410* was deleted in HUDEP-2 cells using paired Cas9
375 cleavages. *ZNF410* null HUDEP-2 cell clones were generated in two steps. In the first step a
376 cell clone with heterozygous deletion of *ZNF410* was obtained using the gRNAs *ZNF410*-del-5'-
377 tgt1 and *ZNF410*-del-3'-tgt1. In the second step this heterozygous *ZNF410* null clone was
378 retargeted using a second pair of guide RNAs, *ZNF410*-del-5'-tgt2 and *ZNF410*-del-3'-tgt2, to
379 obtain biallelic deletion of *ZNF410*. Three individual *ZNF410* null clones were obtained by
380 limiting dilution of bulk edited cells. Mono- or biallelic deletion clones were identified by PCR
381 amplification of the genomic DNA flanking the deletion (outer PCR) and inside the targeted
382 region (inner PCR) using the following primers: *ZNF410*-outer-FP/RP and *ZNF410*-inner-FP/RP.
383 For the rescue experiment, the three *ZNF410* null clones were transduced with either an HA-
384 tagged *ZNF410* construct or an HA-tagged nuclear localization sequence (NLS) containing
385 control vector. Successfully transduced cells were obtained by selection of cells using blasticidin
386 (Invivogen #ant-bl-05).

387

388 Validation in CD34+ HSPCs

389 CD34+ cells were thawed and maintained in X-VIVO complete medium for 24 hours. 100,000
390 cells per condition were electroporated using the Lonza 4D nucleofactor with 100 pmols 3xNLS-
391 Cas9 protein and 300 pmols modified sgRNA targeting the gene of interest. In addition to mock
392 treated cells, *AAVS1* targeting or “safe-targeting” RNPs were used as experimental controls as
393 indicated in each figure legend. After electroporation cells were differentiated to erythroblasts as
394 described previously⁴¹. 4 days after electroporation, genomic DNA was isolated from an aliquot
395 of cells, the sgRNA targeted locus was amplified by PCR and processed for Sanger
396 sequencing. Sequencing results were analyzed by Synthego’s ICE algorithm to obtain editing
397 efficiency and allele contributions. At the end of erythroid culture (day 18) cells were processed
398 for surface marker / enucleation analysis by staining with anti-CD71 (PE-Cy7 conjugated,
399 eBioscience #25-0719-42), anti-CD235a (APC conjugated, eBioscience #17-9987-42) and
400 Hoechst 33342 (Invitrogen #H3570) following manufacturer’s recommendations for antibody
401 concentration and flow cytometry data acquisition on the BD LSR Fortessa. Cells were also
402 processed for hemoglobin HPLC using the Bio-Rad D-10 hemoglobin testing system.

403

404 Dense mutagenesis of *ZNF410*

405 180 guide RNAs were identified by searching for 20-mer sequences upstream of an NGG PAM
406 on the sense and antisense strands of the consensus coding sequence (CCDS) for *ZNF410*

407 obtained from the Ensembl genome browser (Transcript ID ENST00000555044.6). Lentiviral
408 sgRNA libraries were synthesized as previously described⁴⁷ and pooled screening was
409 performed as described in the sgRNA library screening section above. Sequencing results were
410 analyzed by the CRISPRO tool⁴⁸. For each gRNA an HbF enrichment score was calculated
411 comparing the abundance of the gRNA in HbF-high cells to the total cell pool at the end of
412 erythroid culture. Cell fitness scores were calculated by comparing the abundance of the gRNA
413 in cells at the end of erythroid culture to the starting library. The CRISPRO algorithm maps the
414 cell fitness and HbF enrichment score to gene, transcript and protein coordinates and lists
415 associated protein structural domains.

416

417 CUT&RUN

418 CUT&RUN was performed to identify the genome wide ZNF410 / Zfp410 DNA binding profile as
419 previously described¹³. The antibodies used were anti-HA antibody (ThermoFisher #71-5500) in
420 HUDEP-2 cells expressing an HA-tagged ZNF410 construct or anti-ZNF410 (Abcam
421 #ab174204) to detect endogenously expressed ZNF410 in HUDEP-2 and primary human
422 CD34+ cells as well as endogenously expressed Zfp410 in MEL cells. Normal rabbit IgG
423 polyclonal antibody (Millipore Sigma #12370) was used as a control for non-specific sequence
424 enrichment. Anti-H3K27me3 antibody (Cell signaling Technology #9733) was used as a
425 positive control for the steps leading up to the chromatin release. Protein A-MNase was kindly
426 provided by Dr. Steve Henikoff. Sequencing libraries were prepared using the NEBNext Ultra™
427 II DNA Library Prep Kit for Illumina as previously described⁹. Paired-end 42bp sequencing was
428 performed on the NextSeq 500 platform by the Molecular Biology Core Facilities at Dana-Farber
429 Cancer Institute. Sequencing data analyses was adapted from previous protocols^{9,13}. FastQC
430 (Babraham Institute) was performed for all samples to check sequencing quality. Adapter
431 sequences were trimmed with Trimmomatic⁴⁹ with the following settings:
432 “ILLUMINACLIP:\$TRIMMOMATIC/adapters/TruSeq3-PE.fa:2:15:4:4:true
433 SLIDINGWINDOW:4:15 MINLEN:25.” Trimmed reads were aligned to the human reference
434 genome hg19 using bowtie2⁵⁰ with the following settings: “--end-to-end --no-unal --no-mixed --
435 no-discordant --dovetail --phred33 -p 4.” The resulting alignment files (.sam) were converted to
436 sorted, indexed bam files and marked for duplicates using Picard
437 (<https://broadinstitute.github.io/picard/>). Reads were filtered using an alignment score cutoff of
438 10 with samtools⁵¹. Peak calling was performed using macs2⁵² with the following settings:
439 “callpeak -f BAMPE -t [test replicates] -c [control replicates] -B -g [hs or mm] -q 0.05 -n
440 [outputID].” Genomic regions annotated as part of the ENCODE project blacklist⁵³ as

441 problematic regions for alignment of high-throughput functional genomics data were excluded
442 from analysis using files ENCF001TDO (hg19, Birney lab, EBI) and ENCF547MET (mm10,
443 Kundaje lab, Stanford) and BEDtools⁵⁴. Locus footprinting was performed to identify regions of
444 DNA that are relatively protected from MNase cleavage compared to neighboring regions due to
445 occupancy by a transcription factor. Footprint patterns at a locus were determined by
446 enumerating the ends of each fragment sequenced and aligned to the locus. Data was
447 visualized using IGV⁵⁵.

448

449 Genome-wide motif mapping

450 Genome-wide ZNF410 DNA binding motif instances were mapped using the pwmscan webtool
451 (<https://ccg.epfl.ch/pwmtools/pwmscan.php>) and the ZNF410 motif MA0752.1 from the JASPAR
452 CORE 2018 vertebrates motif library. The number of motif instances in the genome was
453 enumerated using a 3 kb sliding window with a 100 bp overlap to determine genomic distribution
454 of motif occurrence for ZNF410. Motifs that fall within the overlapping region between genomic
455 windows are assigned to the adjacent window with the greater number of motifs, or if both
456 adjacent windows have the same number of motifs, motifs are assigned to the first of the two
457 windows.

458

459 ATAC-seq and DNase-seq identification of regions of open chromatin

460 ATAC-seq was performed in HUDEP-2 cells grown in expansion phase medium following the
461 OMNI-ATAC protocol⁵⁶. MEL DNase-seq data was obtained from the ENCODE project^{57,58}
462 (<https://www.encodeproject.org/>) from the lab of John Stamatoyannopoulos, UW (dataset:
463 ENCSR000CNN, file: ENCF990ATO).

464

465 DNA sequence conservation

466 SiPhy rate⁵⁹ (10 mer) from: <http://www.broadinstitute.org/igvdata/hg19/omega.10mers.wig.tdf>.

467

468 GSEA

469 Genes that were differentially expressed in *ZNF410* targeted HUDEP-2 cells compared to
470 *AAVS1* targeted control cells were compared by gene set enrichment analysis (GSEA)^{60,61} to
471 genes that were differentially expressed in *CHD4* targeted HUDEP-2 cells compared to non-
472 targeting control cells⁶. The list of genes differentially expressed in *CHD4* targeted HUDEP-2
473 cells are genes that were differentially expressed in both datasets ($q < 0.05$) when either the
474 helicase domain or the CHDCT2 domain of *CHD4* were perturbed using sgRNA

475 GGTGTCAGTGCCCTGAGCCC or GAATTCGGGCAATGGTAGCT respectively from previously
476 published data⁶. The motivation for combining these two datasets is based on the observation
477 that helicase domain targeting is toxic to cells while CHDCT2 domain targeting is better
478 tolerated and so the combined list of differentially expressed genes better represents gene
479 expression changes due to *CHD4* ablation than either dataset alone.

480

481 Gene dependency correlation

482 Gene dependency scores for 558 cell lines were obtained from the Achilles Avana 20Q2 Public
483 CERES dataset of the Depmap portal (DepMap, Broad (2020): DepMap 20Q2 Public. figshare.
484 Dataset. <https://doi.org/10.6084/m9.figshare.12280541.v4>.)^{16,17}. Project Achilles performs
485 genome scale CRISPR/Cas9 loss of function screening in cancer cell lines and uses CERES to
486 determine a dependency score for each gene in each cell line. Pearson's correlations of
487 dependency scores and p-values were calculated for *ZNF410* and every other gene in the
488 dataset.

489

490 Analysis of gene expression across human tissues

491 *ZNF410* and *CHD4* expression values (TPM) across 54 human tissues were obtained from the
492 GTEx Portal¹⁵ on 10/01/2019. Pearson correlation was used to compare the expression of
493 *ZNF410* and *CHD4*.

494

495 Generation of *CHD4* Δ6.7 kb and Δ6.9 kb clones

496 The genomic region upstream of *CHD4* encompassing the two clusters of *ZNF410* DNA binding
497 motifs was targeted using a pair of sgRNAs (*CHD4*-proximal-gRNA-1:
498 GUGCGGUGGGAUUUCCCGGC and *CHD4*-distal-gRNA-1: CGAGGCUGUGUCAGCGCCGC
499 or *CHD4*-distal-gRNA-2: UUGGUCUGUGGGAUGGACAU) to generate HUDEP-2 clones with
500 biallelic deletion of the intervening sequence. These clones are termed *CHD4* Δ 6.7 kb (for
501 clones generated using *CHD4*-proximal-gRNA-1 and *CHD4*-distal-gRNA-1) or *CHD4* Δ6.9 kb
502 (for clones generated using *CHD4*-proximal-gRNA-1 and *CHD4*-distal-gRNA-2). The bulk
503 population of targeted cells was serially diluted and ~30 cells per plate were plated in 96 well
504 plates to obtain single cell clones. Mono- or biallelic deletion was identified by PCR amplification
505 of the genomic DNA flanking the deletion (outer PCR) and inside the targeted region (inner
506 PCR) using the following primers: *CHD4*-Outer-FP and *CHD4*-Outer-RP1 or *CHD4*-Outer-RP2,
507 *CHD4*-Inner-FP and *CHD4*-Inner-RP (sequences listed in **Supplementary Table**).

508

509 RT-qPCR

510 RNA was isolated using either Trizol (Invitrogen #15596026) or the RNeasy Plus Mini kit
511 (Qiagen #74136) following the manufacturer's protocol. RNA was quantified using the Nanodrop
512 spectrophotometer. cDNA was synthesized using the iScript cDNA synthesis kit (Bio-Rad
513 #1708891) following the manufacturer's recommendations. qPCR was performed using the Sybr
514 Select Master Mix (Thermo Fisher #4472908) on an Applied Biosystems 7300 or Quant Studio 3
515 real-time PCR system. Primers used for RT-qPCR are listed in the **Supplementary Table**. *CAT*
516 was used as a reference gene for human cells and *Gapdh* for mouse cells.

517

518 Generation of *Zfp410* gene-trap allele mice

519 All animal experiments were approved by the Boston Children's Hospital Institutional Animal
520 Care and Use Committee. *Zfp410* gene-trap allele mice were generated as described below.
521 C57BL/6 mice were obtained from Charles River Laboratories (Strain Code 027). Mouse ES
522 cells heterozygous for a *Zfp410* gene-trap allele produced in the EUCOMM (European
523 Conditional Mouse Mutagenesis Program) were purchased from the European Mouse Mutant
524 Cell Repository (EuMMCR, Germany). The ES cells were derived from a C57BL/6N
525 background. The targeting cassette was inserted in intron 5 and contains a splice acceptor site
526 upstream of the *lacZ* gene that disrupts normal splicing and thus expression of *Zfp410*
527 (**Supplementary Fig. 4b**). This allele also has conditional potential with LoxP sites flanking
528 exon 6. We purchased 3 ES cell clones (E06, F06 and F07). Karyotype analysis was performed
529 by EuMMCR. The percentage of cells with normal chromosome count ($2n=40$) for each clone
530 was 77% for E06, 70% for F06 and 90% for F07. Clones E06 and F07 were chosen for
531 blastocyst micro-injections. Chimeric mice were generated by the NIH/NIDDK Center of
532 Excellence in Molecular Hematology, Mouse Embryonic Stem Cell (ESC) and Gene Targeting
533 Core facility. C57BL/6 mice were used as the host for blastocyst micro-injections. For clone
534 E06, there were a total of 13 pups born from 3 pregnant fosters, of which there were 2 male and
535 1 female chimeric mice. For clone F07, there were 3 pups born from 2 pregnant fosters, none of
536 which were chimeras. Of the 3 chimeric mice obtained, one male produced germline
537 transmission of the *Zfp410* gene-trap (Gt) allele upon breeding with wildtype C57BL/6 mice.
538 Mice heterozygous for the *Zfp410* gene-trap allele (*Zfp410* +/Gt) were intercrossed to generate
539 mice homozygous for the *Zfp410* gene-trap allele (*Zfp410* Gt/Gt). Mice carrying the *Zfp410* gene-
540 trap allele were genotyped using primers flanking the LoxP site in intron 6 (LoxP-FP and LoxP-
541 RP). During homologous recombination, the targeting cassette replaces endogenous DNA
542 stretches resulting in a slightly different PCR product from the targeted compared to the wildtype

543 allele. Peripheral blood was collected from mice at 3 months of age. CBCs were performed on
544 the Advia hematology system at the BCH-HSCI Flow Core. Values for the normal range of
545 various hematological parameters for C57BL/6 mice were obtained from the Charles River
546 Laboratories website ([https://animalab.eu/sites/all/pliki/produkty-](https://animalab.eu/sites/all/pliki/produkty-dopobrania/Biochemistry_and_Hematology_for_C57BL6NCrI_Mouse_Colonies_in_North_America_for_January_2008_December_2012.pdf)
547 [dopobrania/Biochemistry and Hematology for C57BL6NCrI Mouse Colonies in North Ameri](https://animalab.eu/sites/all/pliki/produkty-dopobrania/Biochemistry_and_Hematology_for_C57BL6NCrI_Mouse_Colonies_in_North_America_for_January_2008_December_2012.pdf)
548 [can for January 2008 December 2012.pdf](https://animalab.eu/sites/all/pliki/produkty-dopobrania/Biochemistry_and_Hematology_for_C57BL6NCrI_Mouse_Colonies_in_North_America_for_January_2008_December_2012.pdf)). RNA was isolated from whole blood using Trizol
549 following the manufacturer's protocol.

550

551 Xenotransplant

552 NOD.Cg-KitW-41J Tyr + Prkdcscid Il2rgtm1Wjl (NBSGW) mice were obtained from Jackson
553 Laboratory (Stock 026622). CD34+ HSPCs from adult mobilized peripheral blood from de-
554 identified healthy donors were thawed and recovered in X-VIVO complete medium for 24 hours.
555 After recovery, cells were electroporated using the Lonza 4D nucleofector with 3xNLS-Cas9
556 protein and sgRNA. Cells were allowed to recover from electroporation for 24-48 hours in X-
557 VIVO complete medium. Cells were counted and divided equally among 3 or 4 recipient mice
558 per condition. A portion of cells was subjected to *in vitro* erythroid differentiation. Pre-transplant
559 editing efficiency was assessed on day 4 of *in vitro* culture. In each experiment 4 mice received
560 cells that were not subjected to electroporation (mock) as experimental controls. Cells were
561 resuspended in 200 ul DPBS per mouse and infused by retro-orbital injection into non-irradiated
562 NBSGW female mice. 16 weeks post transplantation, mice were euthanized, bone marrow was
563 collected and xenograft analysis was performed as previously described²¹. Analysis of bone
564 marrow subpopulations was performed by flow cytometry. Bone marrow cells were first
565 incubated for 15 minutes with Human TruStain FcX (BioLegend #422302) and TruStain FcX
566 (anti-mouse CD16/32, BioLegend #101320) to block non-specific binding of immunoglobulin to
567 Fc receptors, followed by incubation with anti-human CD45 (V450, clone HI30, BD Biosciences
568 #560367), anti-mouse CD45 (PE-eFluor 610, clone 30-F11, Thermo Fisher #61-0451-82), anti-
569 human CD235a (FITC, BioLegend #349104), anti-human CD33 (PE, BioLegend #366608), anti-
570 human CD19 (APC, BioLegend #302212), anti-human CD3 (PE/Cy7, BioLegend #300420) and
571 anti-human CD34 (FITC, BioLegend #343504) antibodies. Fixable Viability Dye (eFluor 780,
572 Thermo Fisher #65-0865-14) was used to exclude dead cells. The percentage of human
573 engraftment was calculated as $\frac{\text{hCD45}^+ \text{ cells}}{\text{hCD45}^+ \text{ cells} + \text{mCD45}^+ \text{ cells}} \times 100$. B-
574 lymphocyte (CD19+), granulocyte (CD33-dim SSC-high) and monocyte (CD33-bright SSC-low)
575 lineages were gated on the hCD45+ population. HSPCs (CD34+) and T-lymphocyte (CD3+)
576 lineages were gated on the hCD45+ hCD19- hCD33- population. Human erythroid cells

577 (CD235a+) were gated on the hCD45- mCD45- population. The detailed gating strategy is
578 shown in **Supplementary Fig. 6**.

579

580 Statistical analyses

581 All values indicated for replicates (n=x) are biological replicates. p-values were calculated by
582 two-tailed Student's t-test.

583

584 Data availability

585 The datasets generated during the current study are available from the indicated repositories
586 where applicable or are included in this article.

587

588 Code availability

589 The scripts used for analysis of CUT&RUN experiments and motif mapping have been provided
590 in Supplementary Methods.

591

592 **Supplementary Methods**

593

594 CUT&RUN

595 Sequencing data obtained from CUT&RUN experiments were analyzed using the following
596 scripts. The workflow was largely adapted from previous protocols^{9,13}, with the addition of data
597 filtering based on findings of the ENCODE project^{53,57}.

598

599 The prerequisite Software used in our methods are listed below:

600 FastQC 0.11.3

601 Trimmomatic 0.36

602 Bowtie 2 2.2.9

603 Samtools 1.3.1

604 Picard 2.8.0

605 Bedtools 2.27.1

606 Deeptools 3.0.2

607 Macs2 2.1.1.20160309

608

609 1. FastQC (Babraham Institute, version 0.11.3) was performed for all samples to check
610 sequencing quality.

```
611 $zcat *.fastq.gz
612 $fastqc Sample1_Read1.fastq
613 $fastqc Sample1_Read2.fastq
614
615 2. Adapter sequences were trimmed with Trimmomatic49.
616 $java -jar trimmomatic-0.36.jar PE -threads 4 -trimlog trim.log -phred33
617 Sample1_Read1.fastq.gz Sample1_Read2.fastq.gz -baseout path-for-
618 output/Sample1_trimmed.fq.gz ILLUMINACLIP:/adapters/TruSeq3-PE.fa:2:15:4:4:true
619 SLIDINGWINDOW:4:15 MINLEN:25
620
621 3. Trimmed reads were aligned to either the human reference genome hg19 or the mouse
622 reference genome mm10 using bowtie250. Pre-built bowtie2 index files are available at
623 http://bowtie-bio.sourceforge.net/tutorial.shtml.
624 $BOWTIE2_IDX=bowtie2_indexes/hg19 (or bowtie2_indexes/mm10)
625 $bowtie2 --end-to-end --no-unal --no-mixed --no-discordant --dovetail --phred33 -p 4 -x
626 ${BOWTIE2_IDX} -1 Sample1_trimmed_1P.fq.gz -2 Sample1_trimmed_2P.fq.gz -S
627 Sample1.sam
628
629 4. The resulting alignment files (.sam) were converted to sorted, indexed bam files and marked
630 for duplicates using Picard (https://broadinstitute.github.io/picard/).
631 $java -jar picard-2.8.0.jar SortSam I=Sample1.sam O= Sample1.sorted.bam
632 SORT_ORDER=coordinate CREATE_INDEX=true
633 $java -jar picard-2.8.0.jar MarkDuplicates I=Sample1.sorted.bam O=Sample1.dedup.sorted.bam
634 M=Sample1.dedup.txt REMOVE_DUPLICATES=true
635 $bedtools bamtobed -i $sampleID.dedup.sorted.bam > $sampleID.dedup.sorted.bed
636 $cat Sample1.dedup.sorted.bed | awk -v OFS='\t' '{len = $3 - $2; print $0, len }' >
637 Sample1.dedup.sorted.final.bed
638
639 5. Deduplicated and sorted bam files were converted to bigwig files for visualization in IGV.
640 These are the files used to generate the representative IgG or ZNF410 data tracks.
641 $samtools index Sample1.dedup.sorted.bam
642 $bamCoverage --bam Sample1.dedup.sorted.bam -o Sample1.dedup.sorted.bw --binSize 10
643
644 6. Reads were filtered using an alignment score cutoff of 10 with samtools51.
```

645 \$samtools view -b -q 10 Sample_1.dedup.sorted.bam > Sample_1.dedup.filtered.bam

646

647 7. Peak calling was performed using macs2⁵². Biological replicates for test and control samples
648 were grouped together at this stage.

649 \$sampleID=Sample

650 \$controlID=Control

651 \$outputID=Sample_vs_Control

652 \$macs2 callpeak -f BAMPE -t \${sampleID}_Replicate-1.dedup.filtered.bam

653 \${sampleID}_Replicate-2.dedup.filtered.bam \${sampleID}_Replicate-3.dedup.filtered.bam \

654 -c \${controlID}_Replicate-1.dedup.filtered.bam \${controlID}_Replicate-2.dedup.filtered.bam

655 \${controlID}_Replicate-3.dedup.filtered.bam \ -B -g hs -q 0.05 -n \${outputID}

656

657 8. Genomic regions annotated as part of the ENCODE project blacklist^{53,57} as problematic
658 regions for alignment of high-throughput functional genomics data were excluded from analysis
659 using files ENCF001TDO (hg19, Birney lab, EBI) and ENCF547MET (mm10, Kundaje lab,
660 Stanford) and BEDtools⁵⁴.

661 \$bedtools intersect -a Test_vs_control_peaks.bed -b blacklist.bed -v >

662 Test_vs_control_blacklist-filtered.bed

663

664

665 **Acknowledgements**

666

667 We thank Drs. Ryo Kurita and Yukio Nakamura for sharing HUDEP-2 cells (Department of
668 Research and Development, Central Blood Institute, Blood Service Headquarters, Japanese
669 Red Cross Society, Tokyo, Japan and Cell Engineering Division, RIKEN BioResource Research
670 Center, Faculty of Medicine, University of Tsukuba, Ibaraki, Japan); Dr. Ronald Mathieu and the
671 HSCI-BCH Flow Cytometry Facility, supported by the Harvard Stem Cell Institute and the NIH
672 (U54DK110805) for assistance with flow cytometry; Dr. Zachary Herbert from the Molecular
673 Biology Core Facilities at Dana-Farber Cancer Institute for assistance with sequencing; Dr.
674 Yuko Fujiwara from the BCH Mouse Embryonic Stem Cell and Gene Targeting Core (supported
675 by the NIH/NIDDK Center of Excellence in Molecular Hematology U54DK110805) for assistance
676 with transgenic mouse generation; Dr. John Doench for assistance with CRISPR screening; Dr.
677 Steven Henikoff for sharing pA-MNase for CUT&RUN experiments; Dr. Scot Wolfe for sharing
678 3xNLS-SpCas9; Jasmine Bonanno for technical assistance; and Drs. Stuart Orkin, Christian

679 Brendel, Nan Liu, Davide Seruggia, Neekesh Dharia, and members of the Bauer laboratory for
680 helpful discussions. D.S.V. was supported by the Cooley's Anemia Foundation Research
681 Fellowship award (2018-2020); L.P. was supported by NHGRI (R00HG008399 and
682 R35HG010717); D.E.B was supported in part by a Sponsored Research Agreement from
683 Sanofi, NHLBI (DP2HL137300 and P01HL032262), and the Burroughs Wellcome Fund.

684 **Figure Legends**

685

686 **Figure 1. *ZNF410* is a novel HbF repressor.** (a) Schematic of CRISPR/Cas9-based knockout
687 screen in HUDEP-2 cells to identify novel repressors of HbF expression. (b) HbF enrichment
688 and cell fitness scores for each of 1591 transcription factors and 13 genes of the NuRD
689 complex. The gene *ZNF410* was prioritized for further study based on positive HbF enrichment
690 score, neutral cell fitness score and unknown role in erythropoiesis and globin regulation. (c)
691 HUDEP-2/Cas9 cells nontransduced (mock) or transduced with nontargeting (NT) or *ZNF410*
692 targeting sgRNA assayed on day 9 of erythroid differentiation with intra-cellular staining (HbF+
693 cells), RT-qPCR (*HBG1/2* expression, fold change relative to mock) and HPLC (HbF level). Bars
694 indicate mean values and error bars standard deviation (n=3), with $p < 0.05$ for each comparison
695 of NT to *ZNF410* edited. (d) Intra-cellular HbF staining of HUDEP-2 wild-type (wt) cells and
696 three *ZNF410* knockout HUDEP-2 clones without or with (gray bars) re-expression of *ZNF410*.
697 (e) *ZNF410* targeted by RNP electroporation of Cas9 and sgRNA in CD34+ HSPCs and
698 subsequently differentiated to erythroid cells *in vitro*. Bars indicate median value, experiments
699 performed in 4 individual donors including biological triplicate for donor 4 (total n=6 replicates).
700 At the end of erythroid culture (day 18), erythroid maturation was assessed by surface
701 expression of CD71 and CD235a and enucleation frequency by Hoechst staining.
702 Representative FACS plots are shown. Quadrant values indicated are mean \pm SD; t-test
703 comparing *ZNF410* edited to mock for CD71+CD235a+ and CD71-CD235a+ quadrants did not
704 show significant differences ($p > 0.05$). HbF level measured by HPLC was increased in *ZNF410*
705 edited primary erythroid cells compared to mock control cells ($p < 0.0001$).

706

707 **Figure 2. *ZNF410* genomic chromatin occupancy is restricted to two *CHD4* elements with**
708 **densely clustered motifs.** (a) Dense mutagenesis of *ZNF410* coding sequence by pooled
709 screening of 180 sgRNAs (NGG PAM restricted). Each circle represents enrichment score of an
710 individual sgRNA, black line Loess regression. The 5 C2H2 zinc-finger domains (red rectangles)
711 of *ZNF410* appear essential for HbF repression. (b) Genome-wide *ZNF410* chromatin
712 occupancy identified by CUT&RUN in HUDEP-2 samples with *ZNF410*-HA over-expression
713 using anti-HA antibody compared to IgG control (n=4 for each). The two peaks with greatest
714 enrichment of *ZNF410* binding were at the *CHD4* promoter and *CHD4* -6 kb enhancer. The next
715 most enriched peaks, at *CHD4* intron 2 and *TIMELESS* intron 1, showed substantially less
716 enrichment. (c) Genome-wide *ZNF410* motif occurrences (identified from JASPAR and mapped
717 by pwmscan) across 3 kb sliding windows. Only three windows comprised more than two

718 ZNF410 motifs, including the *CHD4* promoter (16 motifs), *CHD4* -6 kb enhancer (11 motifs), and
719 *GALNT18* intron 1 (6 motifs). (d) *CHD4* locus at 100 kb (top panel) or 1.9 kb resolution (bottom
720 panels) indicating ZNF410 binding (red peaks) at the *CHD4* promoter and *CHD4* -6 kb enhancer
721 regions in representative control IgG (n=9) and anti-HA (n=7) samples in HUDEP-2 cells over-
722 expressing HA-tagged ZNF410, control IgG (n=1) and anti-ZNF410 (n=1) in HUDEP-2 cells, and
723 control IgG (n=2) and anti-ZNF410 (n=2) in CD34+ HSPC derived erythroid precursors.
724 Positions of ZNF410 motifs (red rectangles), cleavage frequency (footprint) from ZNF410-HA
725 CUT&RUN (red bars), accessible chromatin by ATAC-seq (gray peaks, n=3) and DNA
726 sequence conservation by SiPhy rate.

727

728 **Figure 3. ZNF410 represses HbF by activating *CHD4*.** (a) RNA-seq differential gene
729 expression analysis of *ZNF410* (n=3) compared to *AAVS1* (n=3) targeted HUDEP-2 cells.
730 Downregulated and upregulated genes defined by $p_{adj} < 0.01$ and $L2FC < -1$ or > 1 respectively. (b)
731 Comparison of genes upregulated in *ZNF410* and *CHD4* mutant cells by GSEA shows
732 enrichment of *CHD4* regulated genes in the *ZNF410* regulated gene set. (c) Pearson correlation
733 between *ZNF410* dependency and *CHD4* dependency across 558 cell lines identifies *CHD4* as
734 the most *ZNF410* codependent gene. (d) *CHD4* expression measured by RT-qPCR in *ZNF410*
735 targeted (n=3) compared to mock and *AAVS1* targeted control HUDEP-2 cells (left panel) and in
736 *ZNF410* targeted primary erythroblasts derived from CD34+ HSPCs (n=3, $p < 0.01$) compared to
737 safe sgRNA targeted control cells on day 7 and day 10 of erythroid culture (right panel). (e)
738 Cas9 paired cleavages with *CHD4*-proximal-gRNA-1 and *CHD4*-distal-gRNA-1 were used to
739 generate an element deletion clone (*CHD4* Δ 6.7 kb), with the biallelic deletion spanning both of
740 the ZNF410 binding regions upstream of *CHD4*. (f) *CHD4* expression measured by RT-qPCR in
741 the *CHD4* Δ 6.7 kb clone compared to 3 individual HUDEP-2 cell clones plated in parallel. *HBG*
742 expression relative to total β -like globin (*HBG*+*HBB*) measured by RT-qPCR in the *CHD4* Δ 6.7
743 kb deletion clone compared to control clones. (g) *CHD4* Δ 6.7 kb clones and HUDEP-2 cells
744 were subjected to control (safe) and *ZNF410* targeting by RNP electroporation. Relative *CHD4*
745 and *HBG* expression measured by RT-qPCR. (h) RNA-seq differential gene expression analysis
746 of *ZNF410* targeted (n=3) compared to *AAVS1* targeted (n=3) *CHD4* Δ 6.7 kb clones.
747 Downregulated and upregulated genes defined by $p_{adj} < 0.01$ and $L2FC < -1$ or > 1 respectively.

748

749 **Figure 4. *Zfp410* deficient mice are viable with unremarkable hematology.** (a) CUT&RUN in
750 mouse erythroleukemia (MEL) cells using anti-Zfp410 antibody (n=3) and IgG control (n=3).
751 Enrichment for Zfp410 binding concentrated at *Chd4* promoter (~77 fold enrichment) and *Chd4* -

752 6 kb enhancer (~45 fold enrichment) peaks. The next most enriched peak was at the *Hist1h2bl*
753 promoter (~14 fold enrichment). (b) *Chd4* locus showing Zfp410 binding (red peaks) at the *Chd4*
754 promoter and *Chd4* -6 kb enhancer in representative IgG control (n=3) and anti-Zfp410 (n=3)
755 samples. Positions of Zfp410 motifs (red rectangles) and accessible chromatin by DNase-seq
756 (gray peaks). (c) Mouse ES cells heterozygous for *Zfp410* gene-trap allele (Gt), obtained from
757 EuMMCR, were used to generate heterozygous (*Zfp410* +/Gt) and homozygous (*Zfp410* Gt/Gt)
758 gene-trap mice, with *Zfp410* +/Gt intercrosses yielding 20 progeny from 4 litters. (d) *Zfp410*
759 expression, measured by RT-qPCR using primers spanning exons 5 and 6, was diminished in
760 *Zfp410* Gt/Gt (n=3) mouse peripheral blood compared to heterozygous (n=7, p<0.05) and
761 wildtype control animals. (e) Mouse weight was measured at indicated time points over the
762 course of 15 weeks. (f) Peripheral blood hematological parameters, with normal ranges for
763 hemoglobin, mean corpuscular volume (MCV), reticulocyte, white blood cell (WBC), neutrophil
764 and platelet count shown by dotted lines.

765

766 **Figure 5. ZNF410 deficient human HSPCs de-repress HbF and retain repopulation**

767 **potential.** (a) Schematic of gene editing and transplant of human CD34+ HSPCs in
768 immunodeficient NBSGW mice. Animals were euthanized 16 weeks post-transplant and bone
769 marrow (BM) was harvested and sorted into various subpopulations by flow cytometry. (b-e)
770 Two independent CD34+ HSPC donors were edited and transplanted into 6 mice for each
771 condition (mock or *ZNF410* edited). Each symbol represents one mouse, recipients of donor 1
772 depicted as circles and donor 2 as triangles. Bars indicate median value. (b) Indel frequency at
773 *ZNF410* was quantified in input cells 4 days after electroporation and in total and sorted
774 engrafted BM cells. Percentage of frameshift alleles is represented in gray and the percentage
775 of in-frame alleles is represented in white for each bar. (c) Engraftment of human hematopoietic
776 cells assessed by hCD45+ compared to total CD45+ cells. (d) B-lymphocytes (CD19+),
777 granulocytes (CD33-dim SSC-high) and monocytes (CD33-bright SSC-low) expressed as
778 fraction of hCD45+ cells. HSPCs (CD34+) and T-lymphocytes (CD3+) expressed as fraction of
779 hCD45+ CD19- CD33- cells. (e) Erythroid cells (hCD235a+) expressed as fraction of hCD45-
780 mCD45- cells. (f) *CHD4* expression measured by RT-qPCR in human erythroid cells from
781 control (n=4) and *ZNF410* edited (n=4) xenografts. (g) HbF measured by HPLC from
782 hemolysates of sorted BM hCD235a+ cells.

783

784

785

786 **Supplementary Figure Legends**

787

788 **Supplementary Figure 1. Absent ZNF410 chromatin occupancy.** (a-c) α -like and β -like
789 globin gene clusters and *GALNT18* intron 1 with a cluster of 6 ZNF410 motifs indicating
790 absence of ZNF410 occupancy in representative CUT&RUN control IgG (n=9) and anti-HA
791 (n=7) in HUDEP-2 cells over-expressing HA-tagged ZNF410, control IgG (n=1) and anti-
792 ZNF410 (n=1) in HUDEP-2 cells, and control IgG (n=2) and anti-ZNF410 (n=2) in CD34+ HSPC
793 derived erythroid precursors. Positions of ZNF410 motifs (red rectangles), accessible chromatin
794 by representative ATAC-seq in HUDEP-2 cells (gray peaks, n=3) and DNA sequence
795 conservation by SiPhy rate.

796

797 **Supplementary Figure 2. ZNF410 chromatin occupancy.** (a) The third most enriched peak
798 for ZNF410 binding (following *CHD4* promoter and -6 kb enhancer) by CUT&RUN with anti-HA
799 antibody in HUDEP-2 cells over-expressing ZNF410-HA was at *TIMELESS* intron 1.
800 Representative CUT&RUN control IgG (n=9) and anti-HA (n=7) in HUDEP-2 cells over-
801 expressing HA-tagged ZNF410, control IgG (n=1) and anti-ZNF410 (n=1) in HUDEP-2 cells, and
802 control IgG (n=2) and anti-ZNF410 (n=2) in CD34+ HSPC derived erythroid precursors.
803 Positions of ZNF410 motifs (red rectangles), accessible chromatin by representative ATAC-seq
804 in HUDEP-2 cells (gray peaks, n=3), DNA sequence conservation by SiPhy rate, and repetitive
805 elements from RepeatMasker. (b) A total of 5 peaks were identified by CUT&RUN with anti-
806 ZNF410 antibody in HUDEP-2 cells. The top 4 peaks were at the *CHD4* promoter or -6 kb
807 enhancer, the fifth was at *DPY19L3* intron 5. (c) A total of 5 peaks were identified by CUT&RUN
808 with anti-ZNF410 antibody in CD34+ HSPC derived erythroid precursors. All 5 peaks were at
809 the *CHD4* promoter or -6 kb enhancer. (d) Peak of ZNF410 occupancy at *DPY19L3* intron 5 in
810 HUDEP-2 cells.

811

812 **Supplementary Figure 3. ZNF410 represses HbF by activating *CHD4*.** (a) Comparison of
813 genes downregulated in *ZNF410* and *CHD4* mutant cells by GSEA shows enrichment of *CHD4*
814 regulated genes in the *ZNF410* regulated gene set. (b) Correlation of *ZNF410* and *CHD4*
815 expression across 54 human tissues from GTEx (Pearson $r=0.77$, $p<0.0001$). (c) Cas9 paired
816 cleavages with *CHD4*-proximal-gRNA-1 and *CHD4*-distal-gRNA-1 (*CHD4* Δ 6.7 kb) or *CHD4*-
817 proximal-gRNA-1 and *CHD4*-distal-gRNA-2 (*CHD4* Δ 6.9 kb) were used to generate HUDEP-2
818 clones with biallelic deletions spanning both of the ZNF410 binding regions upstream of *CHD4*.
819 Positions of ZNF410 motifs (red rectangles) and accessible chromatin by ATAC-seq (gray

820 peaks) shown. (d) CHD4 expression measured by RT-qPCR in *CHD4* Δ 6.9 kb clones
821 compared to HUDEP-2 cells. (e) *CHD4* Δ 6.9 kb clones and HUDEP-2 cells were subjected to
822 AAVS1 (negative control), *ZNF410* and *ZBTB7A* targeting using RNP electroporation of 3X-
823 NLS-Cas9 and sgRNA. Left panel, editing efficiency measured by indel frequency in HUDEP-2
824 cells and *CHD4* Δ 6.9 kb clones targeted with *ZNF410* or *ZBTB7A* sgRNAs. The shaded portion
825 of the bar represents the percentage of indels resulting in frameshift (fs) alleles. The white
826 portion of the bar represents in-frame indels. Right panel, *HBG* expression relative to total β -like
827 globin (*HBG+HBB*) was measured by RT-qPCR in HUDEP-2 cells and *CHD4* Δ 6.9 kb clones
828 targeted with AAVS1 (negative control), *ZNF410* or *ZBTB7A* sgRNAs.

829

830 **Supplementary Figure 4. *Zfp410* is the conserved mouse ortholog of *ZNF410*.** (a)

831 CUT&RUN performed in mouse erythroleukemia (MEL) cells using anti-Zfp410 antibody (n=3)
832 and IgG control (n=3). The third most enriched Zfp410 peak (following *Chd4* promoter and *Chd4*
833 -6 kb enhancer) was at the *Hist1h2bl* promoter. No Zfp410 motifs were identified at this locus,
834 which overlaps accessible chromatin (DNase-seq, gray peaks). (b) Diagram of the *Zfp410* gene
835 trap allele. A targeting cassette including splice acceptor site upstream of *LacZ* was inserted
836 into *Zfp410* intron 5 thus disrupting full-length expression. Schema obtained along with mouse
837 ES cells from EuMMCR, Germany. (c) Exon and domain structure of mouse *Zfp410*.

838

839 **Supplementary Figure 5. *ZNF410* is dispensable for hematopoietic repopulation and
840 erythropoiesis.** CD34⁺ HSPCs from donor 3 were edited by RNP electroporation targeting

841 *ZNF410*, *BCL11A* or *ZBTB7A* and infused to NBSGW mice or subject to in vitro erythroid
842 differentiation. (a) Indel frequency at *ZNF410*, *BCL11A* and *ZBTB7A* was quantified in input
843 cells 4 days after electroporation, and in engrafted total or sorted cells at bone marrow (BM)
844 harvest. The percentage of frameshift alleles is represented in gray and the percentage of in-
845 frame alleles is represented in white. (b) Comparison of engraftment assessed by human
846 CD45⁺ staining compared to total CD45⁺ cells in xenografts of *ZNF410* (n=4), *BCL11A* (n=3)
847 and *ZBTB7A* (n=3) edited and mock control (n=4) CD34⁺ HSPCs. Each symbol represents one
848 mouse. (c, d) Erythroid maturation, evaluated based on CD71 and CD235a immunophenotype
849 and enucleation frequency, was assessed on day 18 of *in vitro* erythroid culture.

850

851 **Supplementary Figure 6. Flow cytometry gating strategy for xenograft experiment.**

852 Hierarchy of FACS gates and representative plots for each gate are shown for a representative
853 control (mock) transplanted bone marrow sample. The first gate was plotted to delineate the cell

854 population of interest (POI) and avoid debris. The second and third gates were plotted to
855 exclude doublets. Values in plots are for respective gates.

856

857

858 **References**

- 859 1. Orkin, S. H. & Bauer, D. E. Emerging Genetic Therapy for Sickle Cell Disease. *Annu. Rev.*
860 *Med.* **70**, 257–271 (2019).
- 861 2. Piel, F. B., Steinberg, M. H. & Rees, D. C. Sickle Cell Disease. *The New England journal of*
862 *medicine* vol. 377 305 (2017).
- 863 3. Weatherall, D. J. The Evolving Spectrum of the Epidemiology of Thalassemia. *Hematol.*
864 *Oncol. Clin. North Am.* **32**, 165–175 (2018).
- 865 4. Sankaran, V. G. *et al.* Human Fetal Hemoglobin Expression Is Regulated by the
866 Developmental Stage-Specific Repressor BCL11A. *Science* **322**, 1839–1842 (2008).
- 867 5. Masuda, T. *et al.* Transcription factors LRF and BCL11A independently repress expression
868 of fetal hemoglobin. *Science* **351**, 285–289 (2016).
- 869 6. Sher, F. *et al.* Rational targeting of a NuRD subcomplex guided by comprehensive in situ
870 mutagenesis. *Nat. Genet.* **51**, 1149–1159 (2019).
- 871 7. Amaya, M. *et al.* Mi2 β -mediated silencing of the fetal γ -globin gene in adult erythroid cells.
872 *Blood* **121**, 3493–3501 (2013).
- 873 8. Martyn, G. E. *et al.* Natural regulatory mutations elevate fetal globin via disruption of
874 BCL11A or ZBTB7A binding. *Nat. Genet.* **50**, 498–503 (2018).
- 875 9. Liu, N. *et al.* Direct Promoter Repression by BCL11A Controls the Fetal to Adult
876 Hemoglobin Switch. *Cell* **173**, 430–442.e17 (2018).
- 877 10. Xu, J. *et al.* Corepressor-dependent silencing of fetal hemoglobin expression by BCL11A.
878 *Proceedings of the National Academy of Sciences* **110**, 6518–6523 (2013).
- 879 11. Maeda, T. *et al.* LRF Is an Essential Downstream Target of GATA1 in Erythroid
880 Development and Regulates BIM-Dependent Apoptosis. *Dev. Cell* **17**, 527–540 (2009).
- 881 12. Benanti, J. A., Williams, D. K., Robinson, K. L., Ozer, H. L. & Galloway, D. A. Induction of
882 extracellular matrix-remodeling genes by the senescence-associated protein APA-1. *Mol.*
883 *Cell. Biol.* **22**, 7385–7397 (2002).
- 884 13. Skene, P. J., Henikoff, J. G. & Henikoff, S. Targeted in situ genome-wide profiling with high
885 efficiency for low cell numbers. *Nat. Protoc.* **13**, 1006–1019 (2018).
- 886 14. Jolma, A. *et al.* DNA-binding specificities of human transcription factors. *Cell* **152**, 327–339

- 887 (2013).
- 888 15. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–
889 585 (2013).
- 890 16. Meyers, R. M. *et al.* Computational correction of copy number effect improves specificity of
891 CRISPR-Cas9 essentiality screens in cancer cells. *Nat. Genet.* **49**, 1779–1784 (2017).
- 892 17. Dempster, J. M. *et al.* Extracting Biological Insights from the Project Achilles Genome-Scale
893 CRISPR Screens in Cancer Cell Lines. 720243 (2019) doi:10.1101/720243.
- 894 18. Liu, P. *et al.* Bcl11a is essential for normal lymphoid development. *Nat. Immunol.* **4**, 525–
895 532 (2003).
- 896 19. O’Shaughnessy-Kirwan, A., Signolet, J., Costello, I., Gharbi, S. & Hendrich, B. Constraint of
897 gene expression by the chromatin remodelling protein CHD4 facilitates lineage
898 specification. *Development* **142**, 2586–2597 (2015).
- 899 20. McIntosh, B. E. *et al.* Nonirradiated NOD,B6.SCID Il2rgamma^{-/-} kitW41/W41 (NBSGW)
900 mice support multilineage engraftment of human hematopoietic cells. *Stem Cell Reports* **4**,
901 171–180 (2015).
- 902 21. Wu, Y. *et al.* Highly efficient therapeutic gene editing of human hematopoietic stem cells.
903 *Nature Medicine* vol. 25 776–783 (2019).
- 904 22. John, C. C. *et al.* Hydroxyurea Dose Escalation for Sickle Cell Anemia in Sub-Saharan
905 Africa. *N. Engl. J. Med.* **382**, 2524–2533 (2020).
- 906 23. Ippolito, G. C. *et al.* Dendritic cell fate is determined by BCL11A. *Proc. Natl. Acad. Sci. U.*
907 *S. A.* **111**, E998–1006 (2014).
- 908 24. Tsang, J. C. *et al.* Single-cell transcriptomic reconstruction reveals cell cycle and multi-
909 lineage differentiation defects in Bcl11a-deficient hematopoietic stem cells. *Genome Biol.*
910 **16**, 178 (2015).
- 911 25. Luc, S. *et al.* Bcl11a Deficiency Leads to Hematopoietic Stem Cell Defects with an Aging-
912 like Phenotype. *Cell Rep.* **16**, 3181–3194 (2016).
- 913 26. Khaled, W. T. *et al.* BCL11A is a triple-negative breast cancer gene with critical functions in
914 stem and progenitor cells. *Nat. Commun.* **6**, 1–10 (2015).
- 915 27. Benitez, C. M. *et al.* An Integrated Cell Purification and Genomics Strategy Reveals
916 Multiple Regulators of Pancreas Development. *PLoS Genet.* **10**, (2014).
- 917 28. Maeda, T. Regulation of hematopoietic development by ZBTB transcription factors. *Int. J.*
918 *Hematol.* **104**, 310–323 (2016).
- 919 29. Yu, X. *et al.* Disruption of the MBD2-NuRD complex but not MBD3-NuRD induces high level
920 HbF expression in human erythroid cells. *Haematologica* **104**, haematol.2018.210963

- 921 (2019).
- 922 30. Krönke, J. *et al.* Lenalidomide causes selective degradation of IKZF1 and IKZF3 in multiple
923 myeloma cells. *Science* **343**, 301–305 (2014).
- 924 31. Lu, G. *et al.* The myeloma drug lenalidomide promotes the cereblon-dependent destruction
925 of Ikaros proteins. *Science* **343**, 305–309 (2014).
- 926 32. Lu, B. *et al.* A Transcription Factor Addiction in Leukemia Imposed by the MLL Promoter
927 Sequence. *Cancer Cell* **34**, 970–981.e8 (2018).
- 928 33. Gillespie, M. A. *et al.* Absolute Quantification of Transcription Factors Reveals Principles of
929 Gene Regulation in Erythropoiesis. *Mol. Cell* **78**, 960–974.e11 (2020).
- 930 34. Jones, W. D. *et al.* De novo mutations in MLL cause Wiedemann-Steiner syndrome. *Am. J.*
931 *Hum. Genet.* **91**, 358–364 (2012).
- 932 35. Sifrim, A. *et al.* Distinct genetic architectures for syndromic and nonsyndromic congenital
933 heart defects identified by exome sequencing. *Nat. Genet.* **48**, 1060–1065 (2016).
- 934 36. Weiss, K. *et al.* De Novo Mutations in CHD4, an ATP-Dependent Chromatin Remodeler
935 Gene, Cause an Intellectual Disability Syndrome with Distinctive Dysmorphisms. *Am. J.*
936 *Hum. Genet.* **99**, 934–941 (2016).
- 937 37. Lambert, S. A. *et al.* The Human Transcription Factors. *Cell* **172**, 650–665 (2018).
- 938 38. Ezer, D., Zabet, N. R. & Adryan, B. Homotypic clusters of transcription factor binding sites:
939 A model system for understanding the physical mechanics of gene expression. *Comput.*
940 *Struct. Biotechnol. J.* **10**, 63–69 (2014).
- 941 39. Kurita, R. *et al.* Establishment of Immortalized Human Erythroid Progenitor Cell Lines Able
942 to Produce Enucleated Red Blood Cells. *PLoS One* **8**, (2013).
- 943 40. Vinjamur, D. S. & Bauer, D. E. Growing and Genetically Manipulating Human Umbilical
944 Cord Blood-Derived Erythroid Progenitor (HUDEP) Cell Lines. *Methods Mol. Biol.* **1698**,
945 275–284 (2018).
- 946 41. Giarratana, M. C. C. *et al.* Proof of principle for transfusion of in vitro-generated red blood
947 cells. *Blood* **118**, 5071–5079 (2011).
- 948 42. Shalem, O. *et al.* Genome-scale CRISPR-Cas9 knockout screening in human cells.
949 *Science* **343**, 84–87 (2014).
- 950 43. Sanjana, N. E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for
951 CRISPR screening. *Nat. Methods* **11**, 783–784 (2014).
- 952 44. Doench, J. G. *et al.* Optimized sgRNA design to maximize activity and minimize off-target
953 effects of CRISPR-Cas9. *Nat. Biotechnol.* **34**, 1–12 (2016).
- 954 45. Li, W. *et al.* MAGeCK enables robust identification of essential genes from genome-scale

- 955 CRISPR/Cas9 knockout screens. *Genome Biol.* **15**, 554 (2014).
- 956 46. Morgens, D. W. *et al.* Genome-scale measurement of off-target activity using Cas9 toxicity
957 in high-throughput screens. *Nat. Commun.* **8**, 15178 (2017).
- 958 47. Canver, M. C. *et al.* BCL11A enhancer dissection by Cas9-mediated in situ saturating
959 mutagenesis. *Nature* **527**, 192–197 (2015).
- 960 48. Schoonenberg, V. A. C. *et al.* CRISPRO: Identification of functional protein coding
961 sequences based on genome editing dense mutagenesis. *Genome Biol.* **19**, 1–19 (2018).
- 962 49. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina
963 sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
- 964 50. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods*
965 **9**, 357–359 (2012).
- 966 51. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–
967 2079 (2009).
- 968 52. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
- 969 53. Amemiya, H. M., Kundaje, A. & Boyle, A. P. The ENCODE Blacklist: Identification of
970 Problematic Regions of the Genome. *Sci. Rep.* **9**, 9354 (2019).
- 971 54. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic
972 features. *Bioinformatics* **26**, 841–842 (2010).
- 973 55. Robinson, J. T. *et al.* Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
- 974 56. Corces, M. R. *et al.* An improved ATAC-seq protocol reduces background and enables
975 interrogation of frozen tissues. *Nat. Methods* **14**, 959–962 (2017).
- 976 57. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human
977 genome. *Nature* **489**, 57–74 (2012).
- 978 58. Davis, C. A. *et al.* The Encyclopedia of DNA elements (ENCODE): data portal update.
979 *Nucleic Acids Res.* **46**, D794–D801 (2018).
- 980 59. Garber, M. *et al.* Identifying novel constrained elements by exploiting biased substitution
981 patterns. *Bioinformatics* **25**, i54–62 (2009).
- 982 60. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for
983 interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15545–
984 15550 (2005).
- 985 61. Mootha, V. K. *et al.* PGC-1alpha-responsive genes involved in oxidative phosphorylation
986 are coordinately downregulated in human diabetes. *Nat. Genet.* **34**, 267–273 (2003).

Figure 1

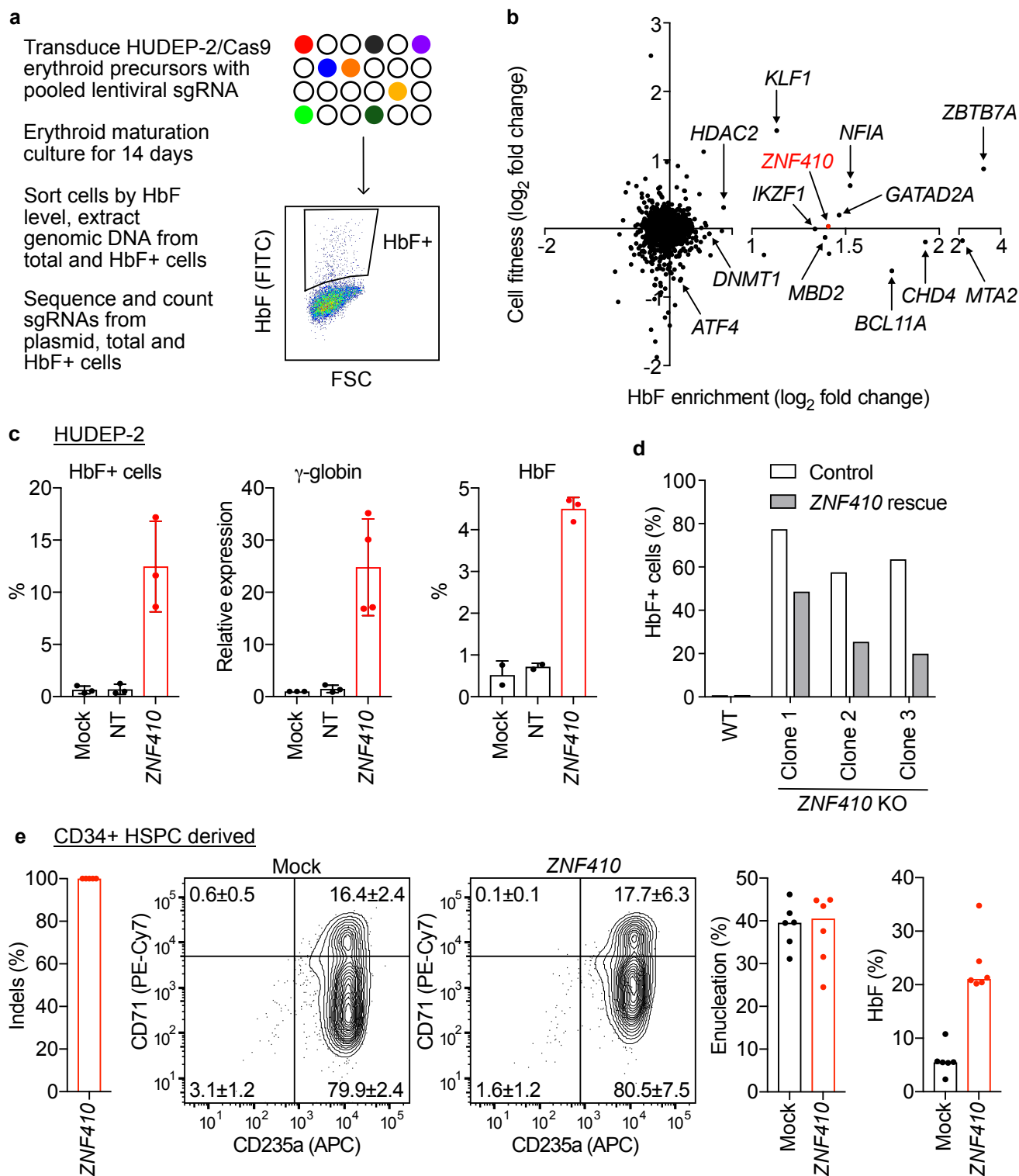


Figure 2

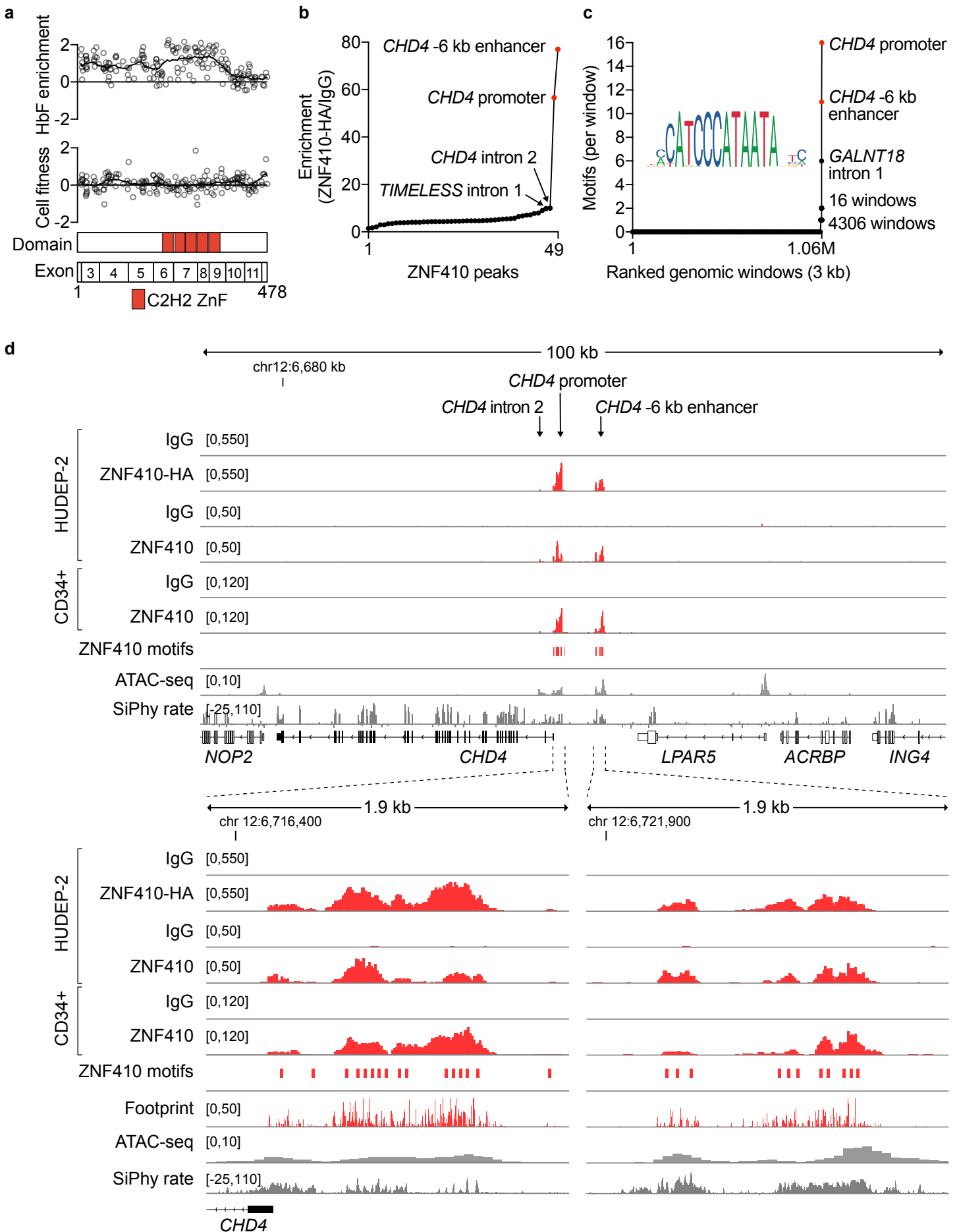


Figure 3

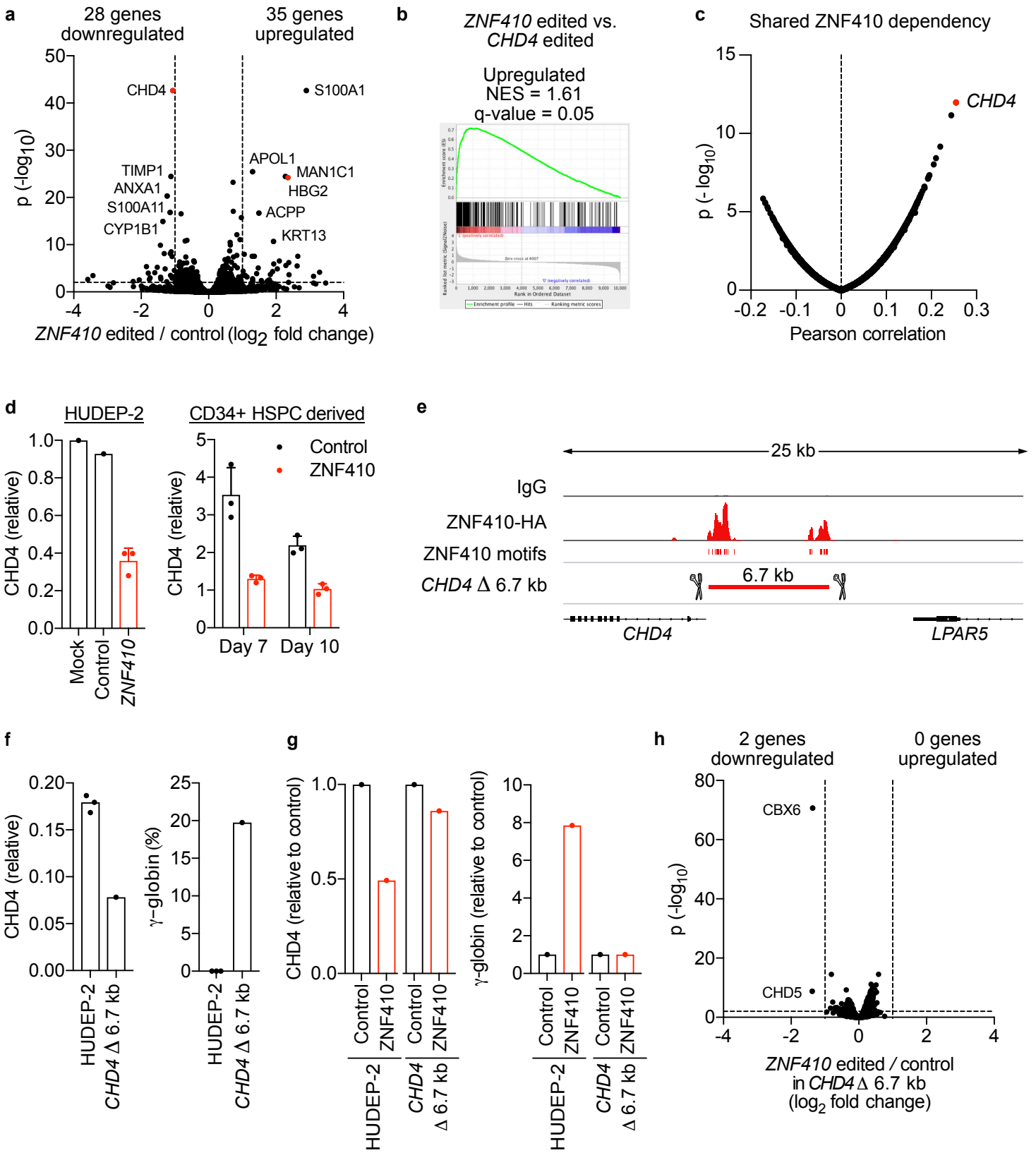


Figure 4

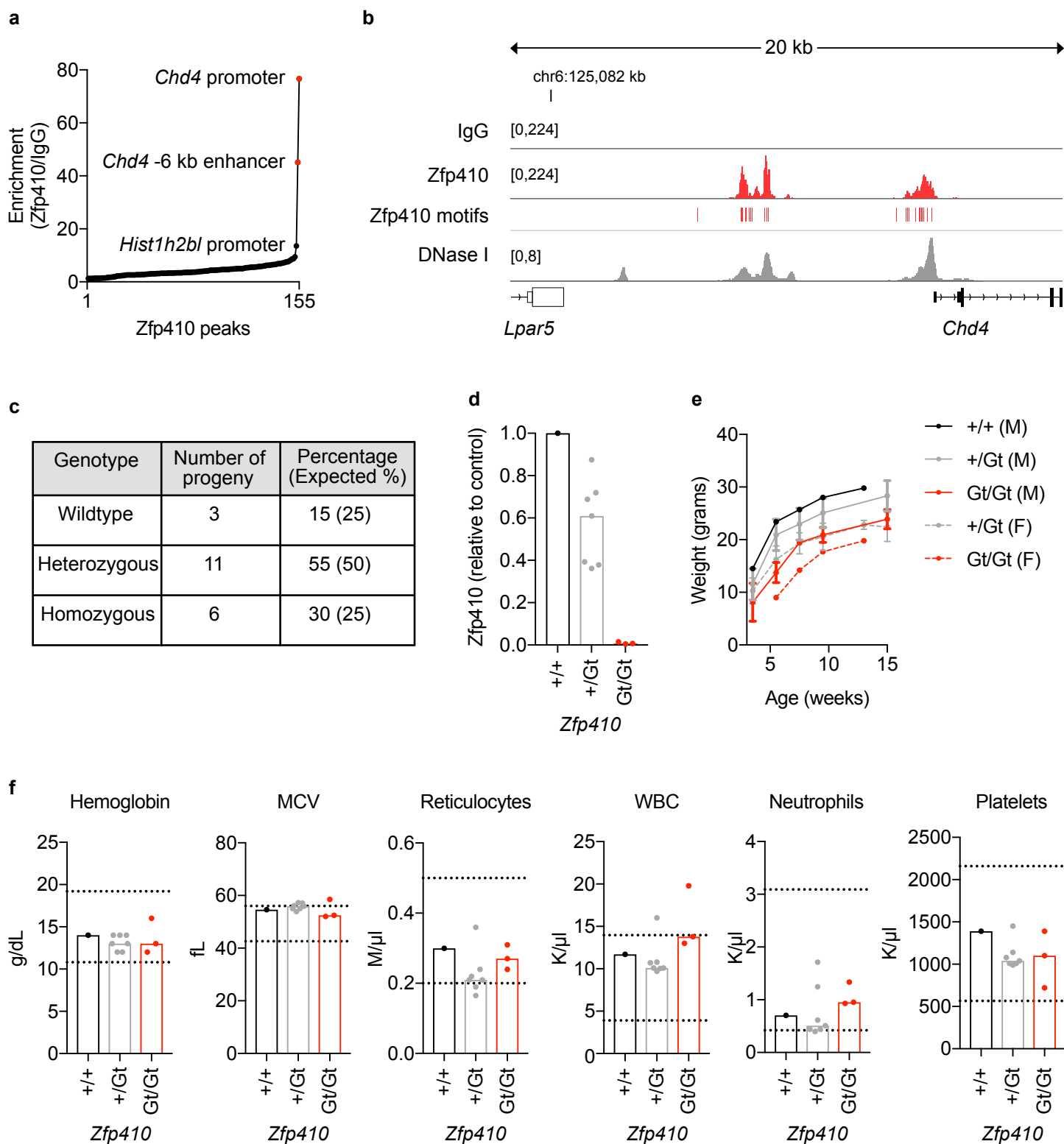
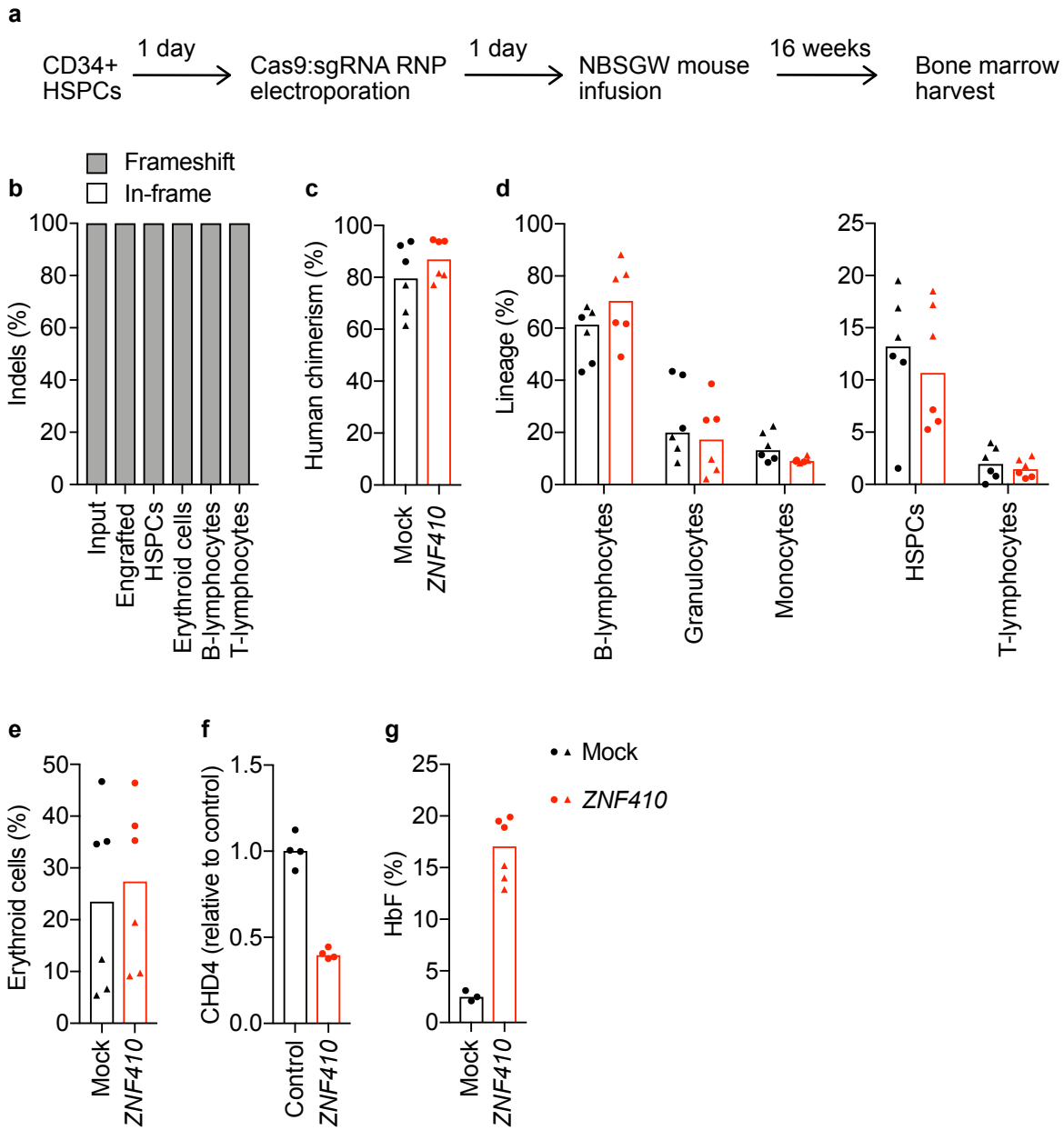
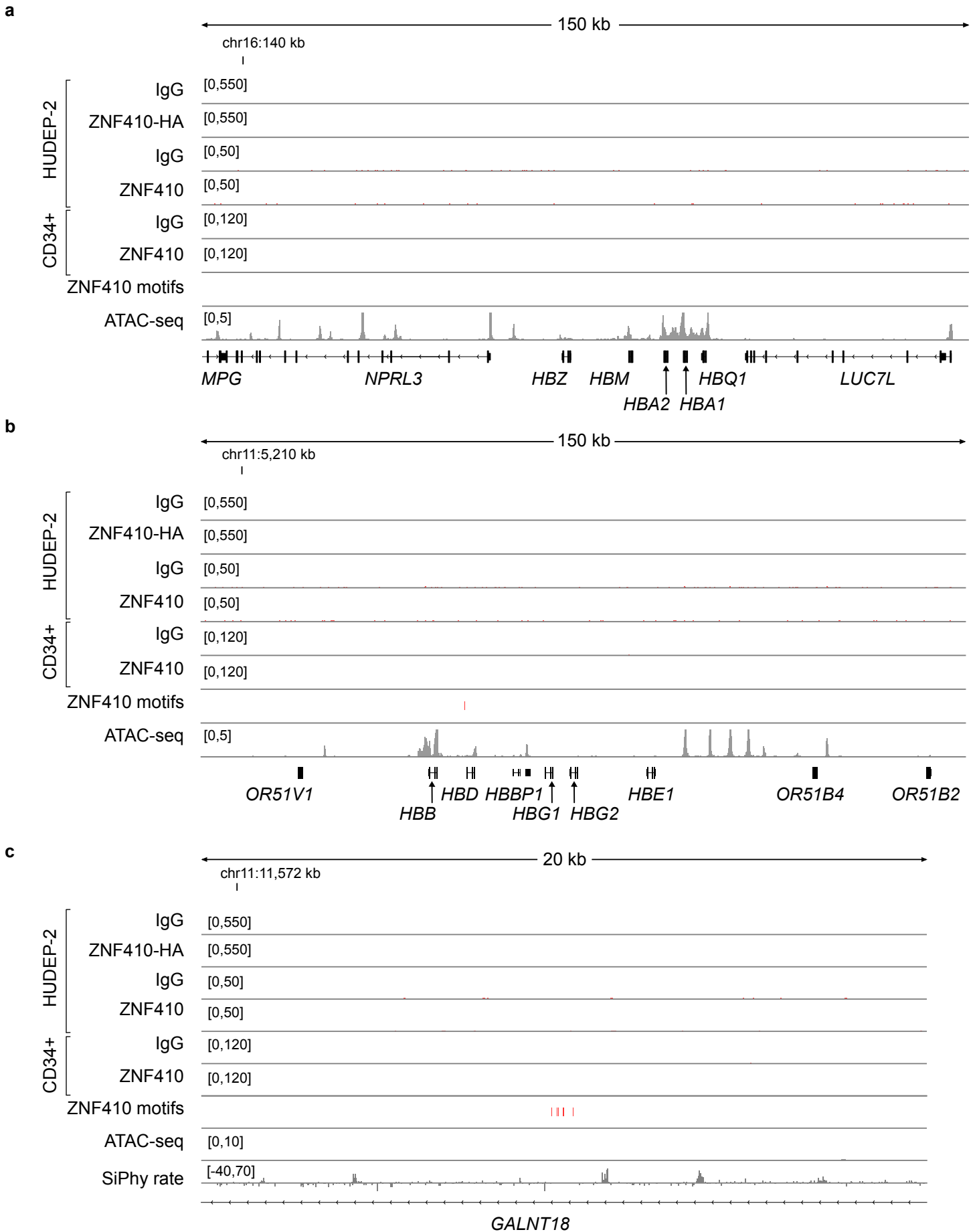


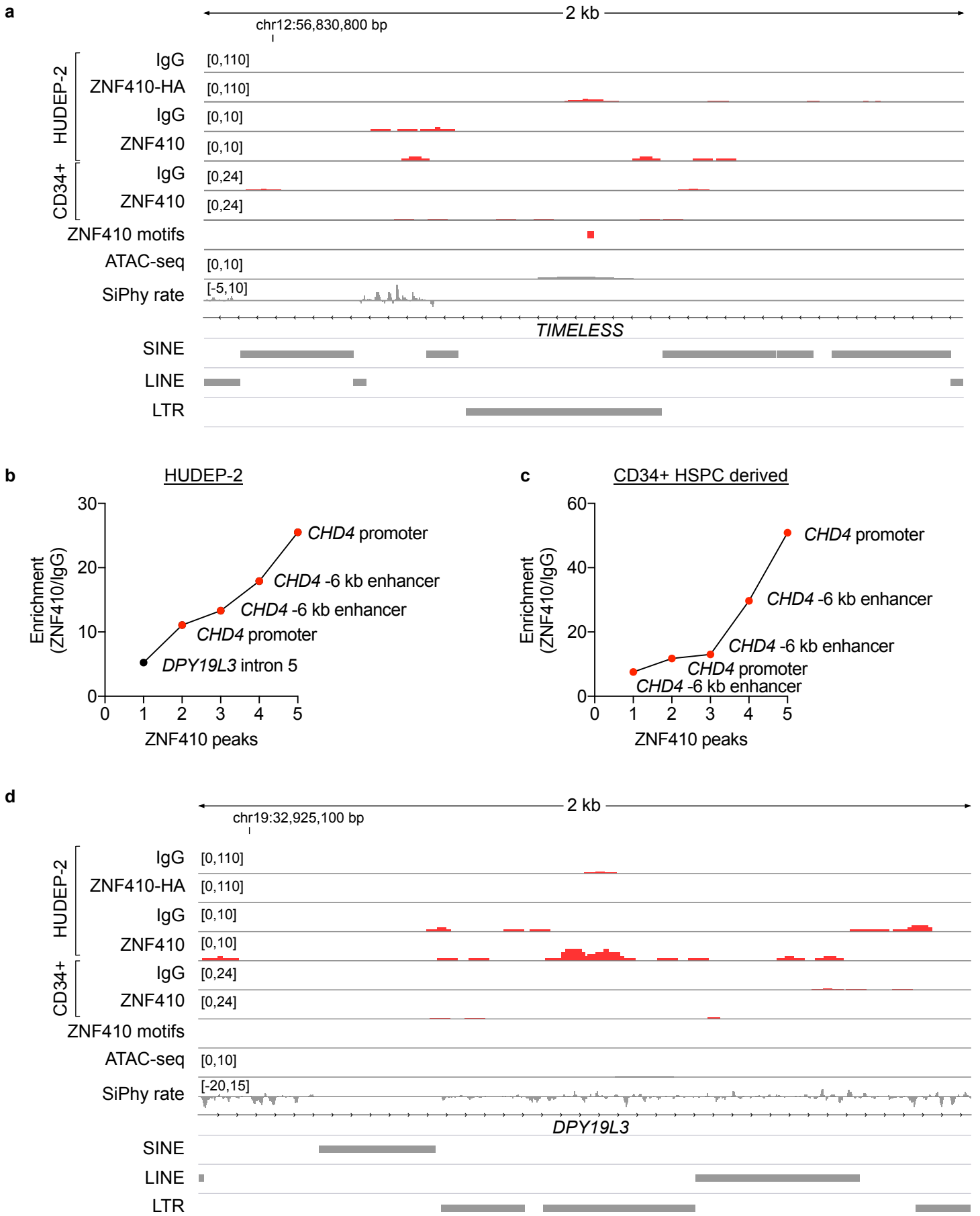
Figure 5



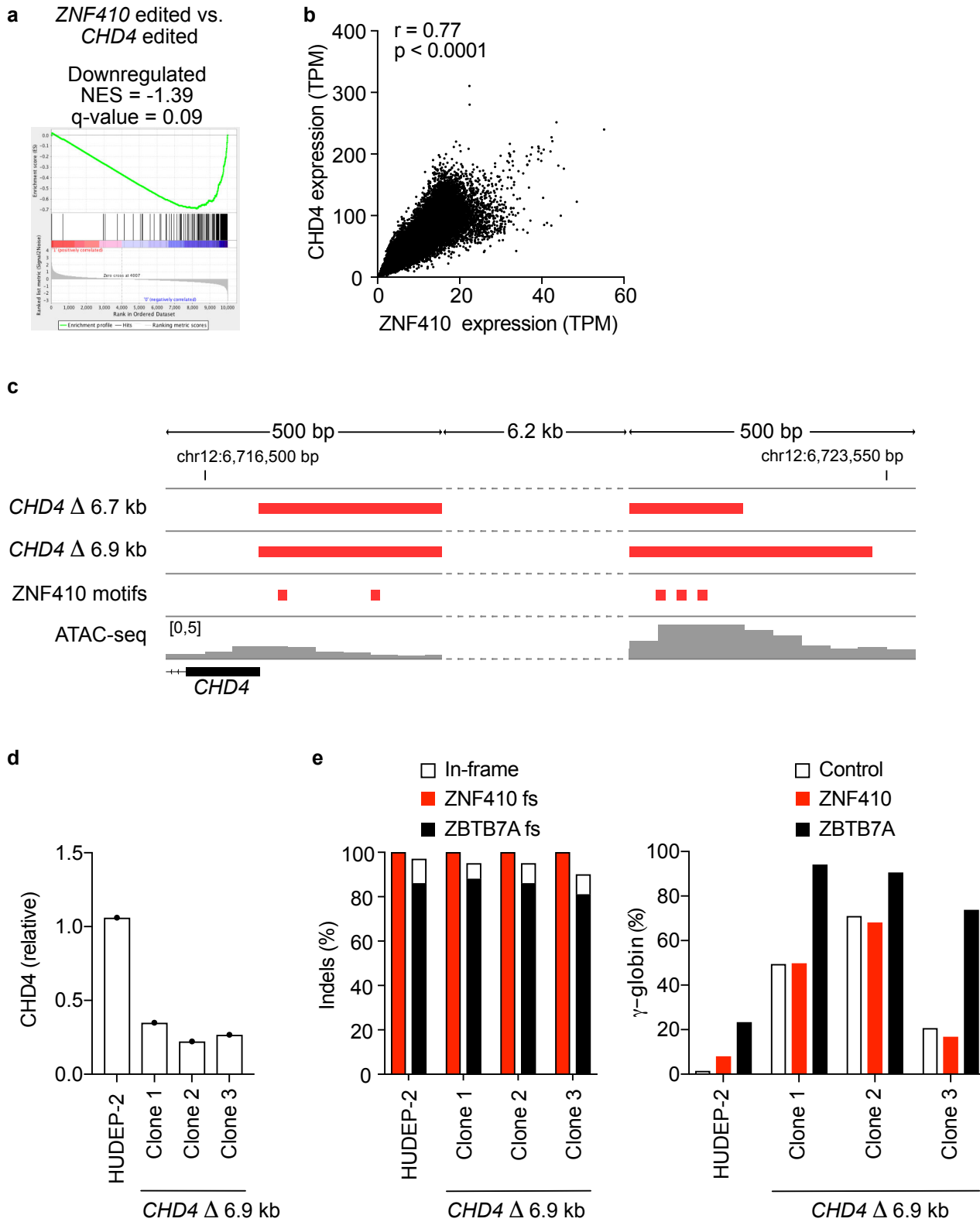
Supplementary Figure 1



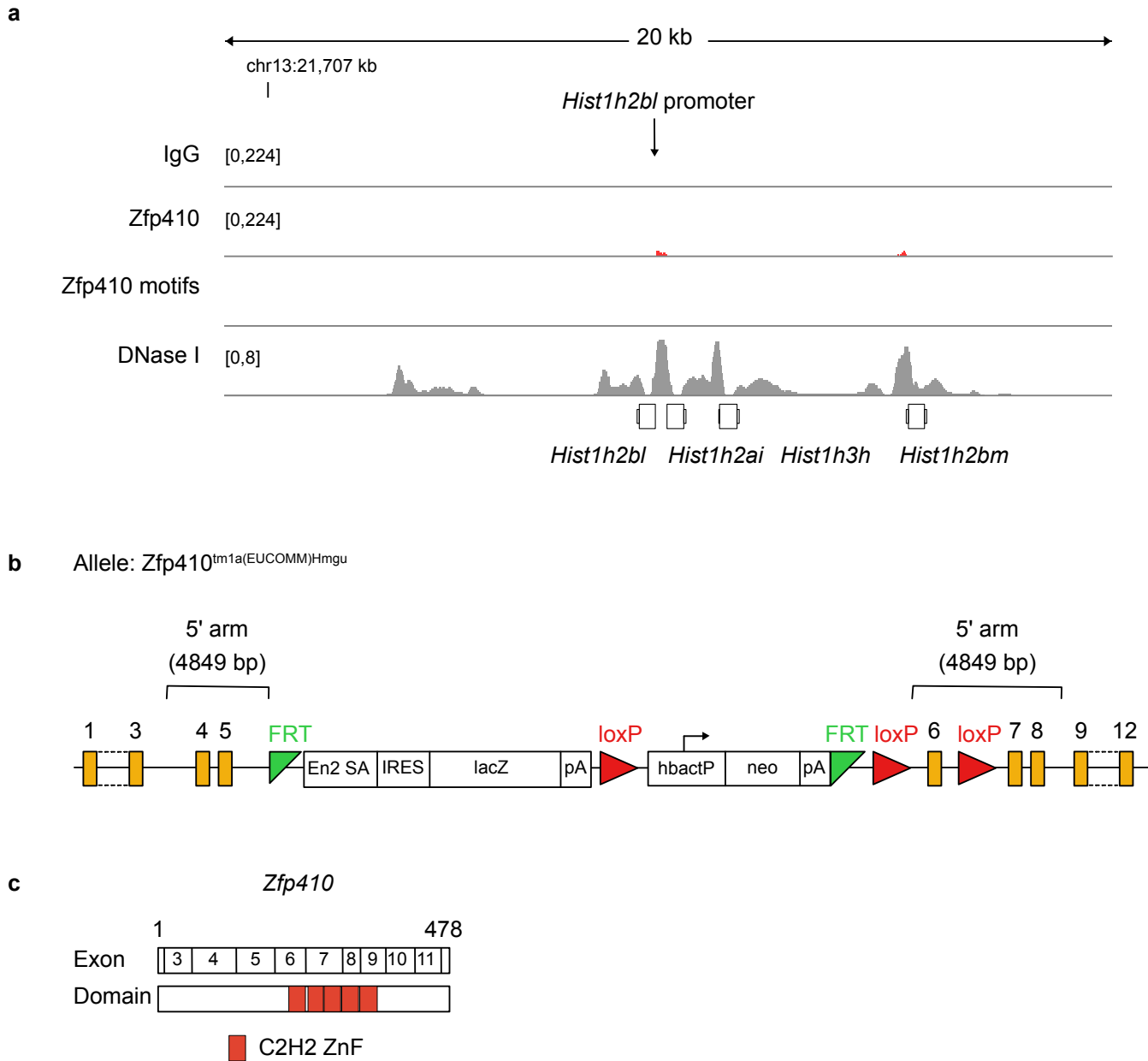
Supplementary Figure 2



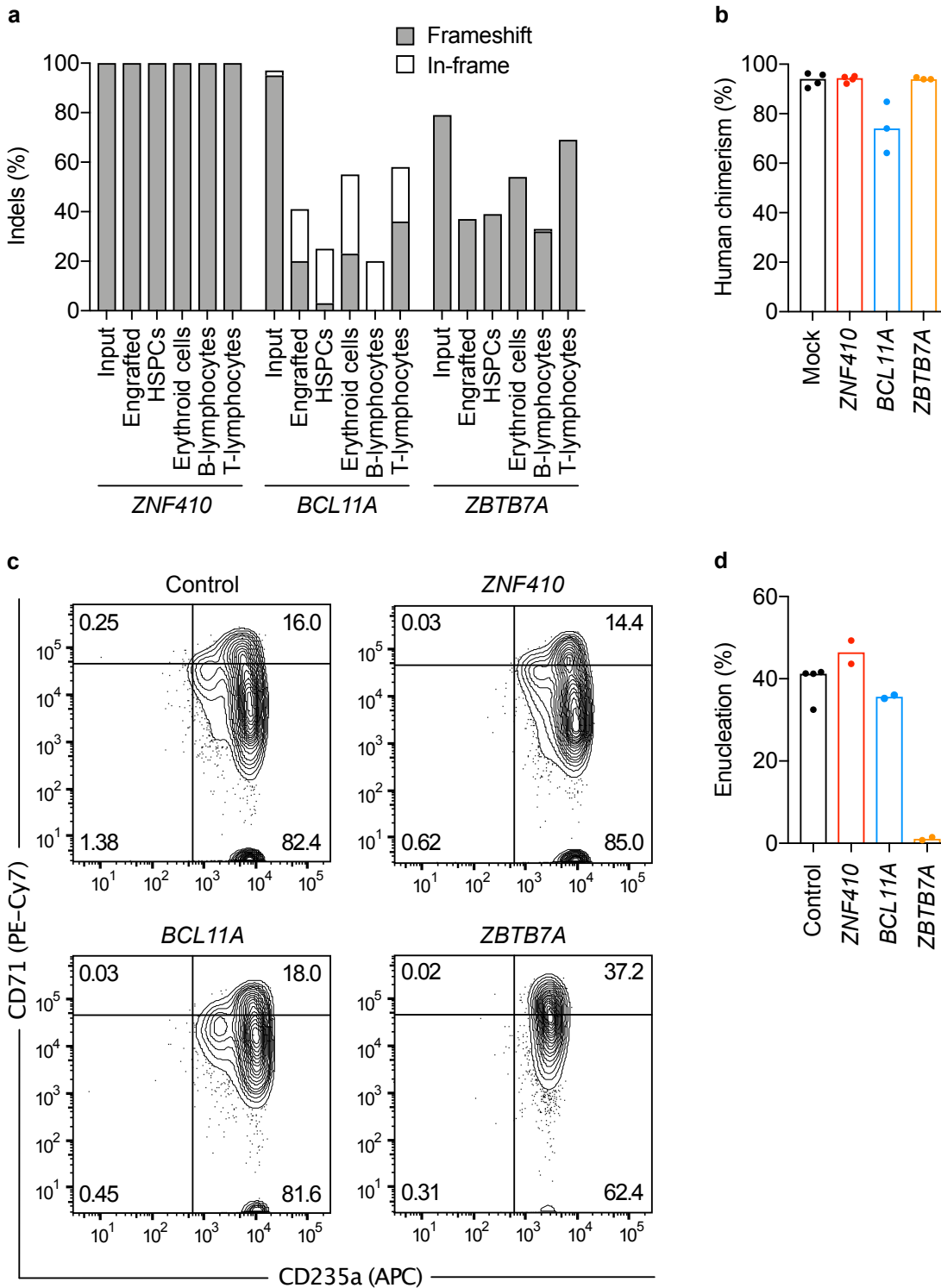
Supplementary Figure 3



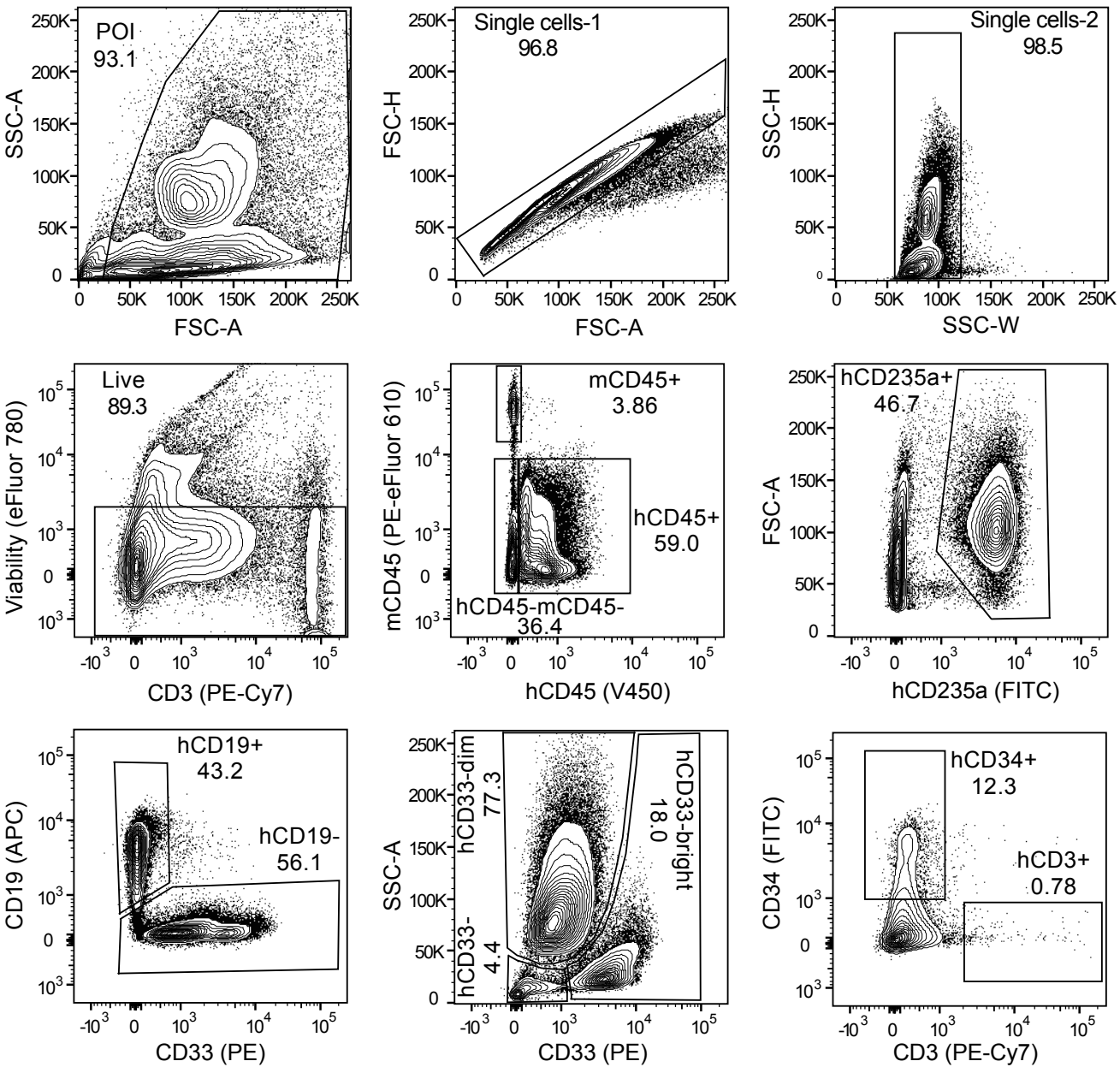
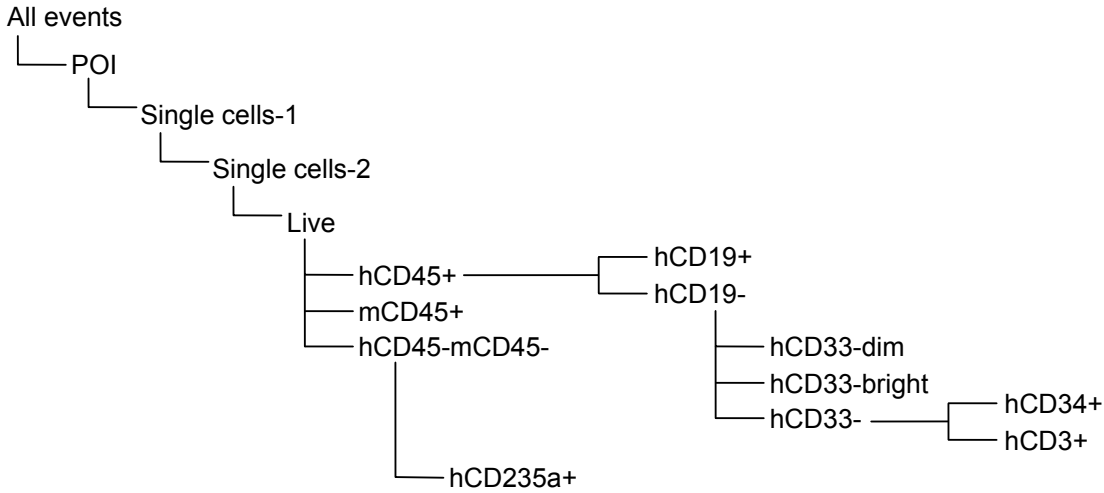
Supplementary Figure 4



Supplementary Figure 5



Supplementary Figure 6



Supplementary Table 1

I Guide RNAs			
Gene/genomic region	gRNA sequence		
AAVS1	CUCCUCCAGGAUCCUCUC		
Safe sgRNA	GUAAGCUUAAAACAUUAGUA		
ZNF410	GUACAGUUGAAGGUUGUGAC		
BCL11A	gACAGAUUGAACCAGACCA		
ZBTB7A	GUAGGCGAAGUCAUGAGCG		
CHD4-proximal-gRNA-1	GUGCGGUGGGAAUUCGCCGC		
CHD4-distal-gRNA-1	CGAGGUGUGUGCAGCGCCGC		
CHD4-distal-gRNA-2	UUGGUCUGGGGAUGGACAU		
ZNF410-del-5f-tgt1	ACGAGCATTTTCTCAAGCAC		
ZNF410-del-5f-tgt1	ATACCAGTAAACTGAACCG		
ZNF410-del-5f-tgt2	CGGGTACTACCTAACCGAGG		
ZNF410-del-3f-tgt2	ATCTCAAATGCGTATACT		
II RT-qPCR primers			
Gene	Forward Primer	Reverse Primer	
HBG	TGGATGATCTCAAGGGCAC	TCAGTGGTATCTGGAGGACA	
HBB	CTGAGGAGAAGTCTGCCGTTA	AGCATCAGGAGTGGACAGAT	
CHD4	CACCGAATCTCAACCACAG	TCTCCACATCTCACTCTCC	
mZfp410 exon 5 / exon 6	CAGGAACTGGCCCATGAC	GTCTTTTGGTTGCCATCAC	
mZfp410 exon 6 / exon 7	CCAGCTCACTCAAGTACCAC	TGTCTCATGTGCACCTTC	
III PCR primers			
Name	Sequence 5'-3'		
CHD4-Outer-FP	GGGCGTCTTTGGGAAC		
CHD4-Outer-RP1	GGAGAACCAAGCAATGGAGAA		
CHD4-Outer-RP2	GTGTTTCCGAAGACCTGAA		
CHD4-Inner-FP	CCCATGGCCATCCATAATAA		
CHD4-Inner-RP	AGGATGTGGTAGGAGAGTAAA		
ZNF410-outer-FP	TGAGATCCCCACAGTACACTTG		
ZNF410-outer-RP	TGATAGCAAGAGTGCTGCTC		
ZNF410-inner-FP	TTGGCAAACAGAACCAAAA		
ZNF410-inner-RP	CTTCCCATCTTGTTTTCCA		
LoxP-FP	GTGAGCAGTGTAGAGGACTTTAT		
LoxP-RP	GATACTGCCAAGCTGACTTA		
IV PCR primers for sgRNA library amplification from genomic DNA			
Step 1 PCR			
Name	Sequence 5'-3'		
PCR_1_FP	AATGGACTATCATATGCTTACCCTAACTTGAAGTATTTTCG		
PCR_1_RP	CTTTAGTTTGTATGCTGTGCTATTATGCTACTATTCTTTCCC		
Name	Sequence 5'-3'		
Forward primers			
PCR_2_F01	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTCCGATCTtAAGTAGAGtcttggaaaggacgaaacaccg		
PCR_2_F02	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTCCGATCTatACACGATcttggaaaggacgaaacaccg		
PCR_2_F03	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTCCGATCTgatCGCGCGTcttggaaaggacgaaacaccg		
PCR_2_F04	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTCCGATCTcgatCATGATCGtcttggaaaggacgaaacaccg		
PCR_2_F05	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTCCGATCTcgatCGTTACCAtcttggaaaggacgaaacaccg		
PCR_2_F06	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTCCGATCTatcgatTCCTTGGTcttggaaaggacgaaacaccg		
PCR_2_F07	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTCCGATCTgatcgatAACGCATtcttggaaaggacgaaacaccg		
PCR_2_F08	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTCCGATCTcgatcgatACAGGTATcttggaaaggacgaaacaccg		
PCR_2_F09	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTCCGATCTacgatAGGTAAGTcttggaaaggacgaaacaccg		
PCR_2_F10	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTCCGATCTtAAATGtcttggaaaggacgaaacaccg		
PCR_2_F11	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTCCGATCTatACTGTATcttggaaaggacgaaacaccg		
PCR_2_F12	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTCCGATCTgatAGGTTCGAtcttggaaaggacgaaacaccg		
Reverse primers			
PCR_2_R01	CAAGCAGAAGACGGCATAACGAGATAAGTAGAG GTGACTGGAGTTCAGACGTGTGCTCTCCGATCTtTACTATTCTTTCCCTGCACCTGT		
PCR_2_R02	CAAGCAGAAGACGGCATAACGAGATAACGATCGTACTGGAGTTCAGACGTGTGCTCTCCGATCTatTACTATTCTTTCCCTGCACCTGT		
PCR_2_R03	CAAGCAGAAGACGGCATAACGAGATCGCGCGTGTGACTGGAGTTCAGACGTGTGCTCTCCGATCTgatTACTATTCTTTCCCTGCACCTGT		
PCR_2_R04	CAAGCAGAAGACGGCATAACGAGATCATGATCGGTGACTGGAGTTCAGACGTGTGCTCTCCGATCTcgatTACTATTCTTTCCCTGCACCTGT		
PCR_2_R05	CAAGCAGAAGACGGCATAACGAGATCGTTACAGTACTGGAGTTCAGACGTGTGCTCTCCGATCTcgatTACTATTCTTTCCCTGCACCTGT		
PCR_2_R06	CAAGCAGAAGACGGCATAACGAGATTCCTTGGTGTGACTGGAGTTCAGACGTGTGCTCTCCGATCTatcgatTACTATTCTTTCCCTGCACCTGT		
PCR_2_R07	CAAGCAGAAGACGGCATAACGAGATAACGATTCGACTGGAGTTCAGACGTGTGCTCTCCGATCTgatcgatTACTATTCTTTCCCTGCACCTGT		
PCR_2_R08	CAAGCAGAAGACGGCATAACGAGATACAGGTATGACTGGAGTTCAGACGTGTGCTCTCCGATCTcgatcgatTACTATTCTTTCCCTGCACCTGT		
PCR_2_R09	CAAGCAGAAGACGGCATAACGAGATAGGTAAGGGTGTGACTGGAGTTCAGACGTGTGCTCTCCGATCTacgatcgatTACTATTCTTTCCCTGCACCTGT		
PCR_2_R10	CAAGCAGAAGACGGCATAACGAGATAACAATGGGTGACTGGAGTTCAGACGTGTGCTCTCCGATCTtTACTATTCTTTCCCTGCACCTGT		
PCR_2_R11	CAAGCAGAAGACGGCATAACGAGATACTGTACTGACTGGAGTTCAGACGTGTGCTCTCCGATCTatTACTATTCTTTCCCTGCACCTGT		
PCR_2_R12	CAAGCAGAAGACGGCATAACGAGATAGTTCGAGTACTGGAGTTCAGACGTGTGCTCTCCGATCTgatTACTATTCTTTCCCTGCACCTGT		