HHS Public Access

Author manuscript

*Nat Genet*. Author manuscript; available in PMC 2023 February 20.

Published in final edited form as:

*Nat Genet*. 2022 September ; 54(9): 1417–1426. doi:10.1038/s41588-022-01152-6.

# HIC2 controls developmental hemoglobin switching by repressing *BCL11A* transcription

**Peng Huang**[1,✉], **Scott A. Peslak**[1,2], **Ren Ren**[3], **Eugene Khandros**[1], **Kunhua Qin**[1], **Cheryl A. Keller**[4,5], **Belinda Giardine**[4], **Henry W. Bell**[6], **Xianjiang Lan**[1], **Malini Sharma**[1], **John R. Horton**[3], **Osheiza Abdulmalik**[1], **Stella T. Chou**[1], **Junwei Shi**[7], **Merlin Crossley**[6], **Ross C. Hardison**[4], **Xiaodong Cheng**[3], **Gerd A. Blobel**[1,7,✉]

[1]Division of Hematology, The Children's Hospital of Philadelphia, Philadelphia, PA, USA.

[2]Division of Hematology/Oncology, Department of Medicine, Hospital of the University of Pennsylvania, Philadelphia, PA, USA.

[3]Department of Epigenetics and Molecular Carcinogenesis, University of Texas MD Anderson Cancer Center, Houston, TX, USA.

[4]Department of Biochemistry and Molecular Biology, Pennsylvania State University, University Park, PA, USA.

[5]Genomics Research Incubator, Pennsylvania State University, University Park, PA, USA.

[6]School of Biotechnology and Biomolecular Sciences, UNSW Sydney, Sydney, New South Wales, Australia.

[7]Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA.

## Abstract

The fetal-to-adult switch in hemoglobin production is a model of developmental gene control with relevance to the treatment of hemoglobinopathies. The expression of transcription factor BCL11A,



✉**Correspondence and requests for materials** should be addressed to Peng Huang or Gerd A. Blobel. huangp1@chop.edu; blobel@chop.edu.

Author contributions

P.H. and G.A.B. conceived the project. P.H., S.A.P., E.K., K.Q., X.L., M.S., C.A.K., B.G., H.W.B., O.A., S.T.C., J.S. and R.C.H. conducted the experiments. R.R. and J.R.H. performed protein purification, DNA-binding, X-ray data collection and structure determination. J.R.H. performed computational modeling. X.C. (who is a CPRIT Scholar in Cancer Research) organized and designed the scope of the structural study. P.H., M.C., X.C. and G.A.B. designed the experiments, performed data analyses and wrote the manuscript. All authors contributed to editing of the manuscript.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-022-01152-6.

Code availability

Codes used in the present study were obtained from published pipelines and packages.

Competing interests

The authors declare no competing interests.

**Extended data** are available for this paper at https://doi.org/10.1038/s41588-022-01152-6.

**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s41588-022-01152-6.

which represses fetal β-type globin (*HBG*) genes in adult erythroid cells, is predominantly controlled at the transcriptional level but the underlying mechanism is unclear. We identify HIC2 as a repressor of *BCL11A* transcription. HIC2 and BCL11A are reciprocally expressed during development. Forced expression of HIC2 in adult erythroid cells inhibits *BCL11A* transcription and induces *HBG* expression. HIC2 binds to erythroid *BCL11A* enhancers to reduce chromatin accessibility and binding of transcription factor GATA1, diminishing enhancer activity and enhancer–promoter contacts. DNA-binding and crystallography studies reveal direct steric hindrance as one mechanism by which HIC2 inhibits GATA1 binding at a critical *BCL11A* enhancer. Conversely, loss of HIC2 in fetal erythroblasts increases enhancer accessibility, GATA1 binding and *BCL11A* transcription. HIC2 emerges as an evolutionarily conserved regulator of hemoglobin switching via developmental control of *BCL11A*.

The β-globin gene cluster is among the most deeply studied loci and has served as a paradigm for the developmental regulation of gene expression and nuclear architecture[1–6]. The genetics that underpin diseases caused by insufficient expression of β-globin genes, such as various forms of thalassemia, have greatly contributed to our understanding of the β-globin locus[7–10]. A strong driving force behind these studies has been the recognition that elevated expression of the fetal β-like globin genes (*HBG1/2)* provides clinical benefit to patients with sickle cell disease (SCD) and β-thalassemia. One major direct *HBG* transcriptional repressor is transcription factor BCL11A[11,12], which occupies the *HBG* promoters to inhibit their activity in adult erythroblasts[4,13,14]. Hence, a key question is how *BCL11A* is regulated.

Although there is evidence to support a role for a post-transcriptional mechanism for *BCL11A* regulation via fetal stage-expressed *LIN28B*[15,16] and *IGF2BP1*[17], several lines of evidence indicate that *BCL11A* is primarily regulated at the transcriptional level. *BCL11A* messenger RNA is more abundant in adult than in fetal erythroblasts based on transcriptome analyses, RNA polymerase II (Pol II) profiling, quantitative PCR with reverse transcription (RT–qPCR) analysis, histone modifications and chromatin accessibility measurements[3,5,18–20] (the present study). The difference in mRNA synthesis thus seems to account in large part for corresponding changes in BCL11A protein levels.

Although several transcription factors are known to drive *BCL11A* expression in adult erythroid cells, such as Kruppel-like factor 1 (*KLF1*)[21], GATA1 (ref.[22]), the heme-regulated inhibitor (HRI)-regulated factor ATF4[23,24] and *NFIA*/-*X*[25], there is no evidence that any of these account for the differences in *BCL11A* levels between fetal and adult stages. Therefore, a critical unresolved issue in the control of globin gene switching is the developmental control mechanism(s) of *BCL11A* transcription.

In the present study, using a CRISPR (clustered regularly inter-spaced short palindromic repeats)-based genetic screen, we identify the fetal stage-expressed transcription factor HIC2 as a repressor of *BCL11A* transcription, and demonstrate that HIC2 directly disables GATA1-driven *BCL11A* enhancers to impair *BCL11A* mRNA synthesis at the fetal stage of erythroid development.

## Results

### *BCL11A* is primarily regulated at the transcriptional level.

Several transcriptome studies have indicated that *BCL11A* mRNA levels are higher in adult compared with fetal erythroblasts[3,5,18–20], although one study did not detect major differences[16]. We therefore carefully quantified the *BCL11A* mRNA levels during development in normal human erythroid cells. We expanded $CD34^+$ hematopoietic stem and progenitor cells (HSPCs) from three distinct developmental stages—fetal liver, adult peripheral blood and newborn cord blood, the latter representing populations transitioning from fetal to adult stages—and differentiated them in vitro toward the erythroid lineage. Adult (*HBB*) and fetal (*HBG*) globin levels were highest in adult and fetal cells, respectively, with intermediate levels in cord blood cells, validating the assay (Fig. 1a,b). Importantly, *BCL11A* mRNA levels were lowest in fetal erythroblasts, increased in newborn erythroblasts and peaked in adult cells (Fig. 1a,b), consistent with previous RNA-sequencing (RNA-seq) results and Pol II profiling (Extended Data Fig. 1a,b). We noted that Pol II occupancy was higher at the *BCL11A* locus in adult cells in a previous study[5], and sought to further examine whether the adult stage increase in mRNA levels is due to a gain in *BCL11A* transcription in our hands. To this end, we profiled the chromatin accessibility landscape (via assay for transposase-accessible chromatin with high-throughput sequencing (ATAC-seq)[26]) along with the active histone mark H3K27ac, and the erythroid master transcription factor GATA1 (via CUT&RUN[27]) (Extended Data Fig. 1c–f). We observed that, at the key *BCL11A* erythroid enhancers, +55, +58 and +62, all of these features were markedly higher in adult erythroblasts compared with their fetal counterparts (Fig. 1c and Extended Data Fig. 1c–f), reflecting a considerably higher enhancer activity in adult cells and accounting for the increased levels in *BCL11A* mRNA production.

To compare the activities of the *BCL11A* enhancers, we inserted 523-bp, 613-bp and 429-bp fragments spanning the +55, +58 and +62 regions, respectively, into a lentiviral vector next to a luciferase reporter gene. We measured the *BCL11A* enhancer activity using reporter assays in fetal-type HUDEP1 and adult-type HUDEP2 erythroid cell lines[28]. All three enhancers were more active in HUDEP2 cells than in HUDEP1 cells, with the +55 enhancer showing the greatest difference in activity between the two stages (Fig. 1d), further confirming that the *BCL11A* enhancers are under developmental control.

Altogether, these results indicate that *BCL11A* is primarily regulated at the transcriptional level during development through the dynamic activities of the *BCL11A* enhancers.

### A CRISPR screen identifies HIC2 as a fetal globin regulator.

To identify candidate regulators of the fetal-to-adult hemoglobin switch and *BCL11A* regulation, we mined our recent CRISPR–Cas9 genetic screens conducted in the adult-type erythroid cell line HUDEP2 (refs.[24,29]). Our search focused on the identification of single guide (sg)RNAs enriched in cells with low levels of fetal hemoglobin (HbF) because these might uncover repressors of *BCL11A*. All six sgRNAs in the library targeting HIC2 were strongly enriched in the HbF-low cells (Extended Data Fig. 2a). HIC2 is a penta-dactyl zinc finger DNA-binding protein of the POK (POZ domain Krüppel-like zinc finger) family

that contains a BTB/POZ domain critical for transcription repression[30]. HIC2 sgRNA read counts persisted throughout cell cultures similar to control sgRNAs, indicating that HIC2 loss does not affect cell fitness (Extended Data Fig. 2b)[31].

The screening results were validated by depleting HIC2 in HUDEP2 cells using two independent sgRNAs. Although these cells express low levels of *HIC2* and *HBG*, depletion of HIC2 further decreased *HBG* mRNA levels (Extended Data Fig. 2c,d).

We measured HIC2 levels in fetal liver, cord blood and adult peripheral blood-derived erythroid cultures and found that both HIC2 mRNA and protein levels were highest in fetal cells, inversely correlating with those of *BCL11A*, but not with the other major fetal globin repressor *ZBTB7A* (Fig. 1a,b and Extended Data Fig. 2e). The *HIC2* gene is evenly expressed through pro-erythroblast, early basophilic, late basophilic and polychromatic stages, and ultimately declines at the orthochromatic stage of erythroid maturation[32] (Extended Data Fig. 2f).

We next examined whether HIC2 promotes fetal globin expression or even imparts a global fetal phenotype on adult erythroid cells by forcing its expression in HUDEP2 cells. HIC2 overexpression (OE) was achieved by transduction with a lentiviral vector containing HIC2 fused to a self-cleaving P2A sequence and green fluorescent protein (GFP) (Extended Data Fig. 2g). We enriched for cells expressing HIC2 protein at levels similar to endogenous HIC2 levels in primary fetal erythroblasts and supraphysiological HIC2 levels by gating on GFP. By comparing GFP-low and GFP-high cells we observed a HIC2 dosage-dependent upregulation of *HBG* (Fig. 2a,b and Extended Data Fig. 2g). By RNA-seq we identified 322 upregulated and 224 downregulated genes (false discovery rate (FDR) < 0.01 and fold-change >2) compared with control HUDEP2 cells (Fig. 2c). Gene set enrichment analysis (GSEA)-based comparisons with RNA-seq data from primary fetal and adult erythroblasts[3] revealed that upregulated genes were enriched for fetal gene expression signatures whereas downregulated genes were consistent with adult gene expression patterns (Fig. 2d). Notably, we did not observe differential expression of previously identified fetal genes *LIN28B* and *IGF2BP1*, suggesting that HIC2 causes only a partial adoption of the fetal cell transcriptional program. The most highly upregulated genes were *HBG* (>90-fold) and *HBZ* (>70-fold) (Fig. 2a–c), both of which are repressed by *BCL11A*.

ATAC-seq and H3K27ac chromatin immunoprecipitation sequencing (ChIP–seq) revealed marked increases in chromatin accessibility and histone acetylation at the *HBG* genes in HIC2 OE cells (Fig. 2e). We further investigated the chromatin contact probabilities of the distal enhancer using Capture-C[33] and found significant increases in locus control region (LCR)–*HBG* gene contacts (Fig. 2e,f), consistent with elevated *HBG* transcription levels. In concert, these results reveal that HIC2 can potently induce fetal globin transcription and partially reprogram HUDEP2 cells toward a fetal state.

## HIC2 OE augments HbF production.

We overexpressed HIC2 via lentiviral transduction in primary peripheral blood $CD34^+$ HSPC-derived erythroid cells. Consistent with the findings in HUDEP2 cells, HIC2 OE significantly elevated *HBG* transcripts, the proportion of cells expressing HbF (F-cells)

and HbF protein levels (Fig. 3a–d). HIC2 OE modestly delayed erythroid differentiation (Extended Data Fig. 3a,b). *HBB* transcripts were also diminished, which might be due in part to the maturation delay or the result of reciprocal expression of the fetal and adult globin genes (Extended Data Fig. 3c,d). Notably, the *HBG* increases were seen at all examined erythroid maturation stages (Extended Data Fig. 3a,b,e).

To examine the effects of HIC2 in a whole animal model system, we performed xenotransplantation of HIC2 OE HSPCs into immunodeficient NBSGW mice. Although the engraftment efficiencies were low for both control and HIC2 OE, we observed a trend of *HBG* induction on HIC2 OE in human erythroid cells isolated from mouse bone marrow (Extended Data Fig. 3f). Together, these results suggest that HIC2 positively regulates *HBG* levels in primary adult erythroblasts and that the effects are not accounted for by changes in cell maturation.

Elevation of HbF levels can attenuate hemoglobin polymerization and thus reduce sickling of erythroid cells in patients with SCD. To test whether HIC2 OE can counteract cellular sickling, we overexpressed HIC2 in cells derived from patients with SCD[23,34]. Consistent with the results in HUDEP2 and normal primary adult erythroblasts, HIC2 OE strongly induced *HBG* transcription and raised the number of F-cells (Extended Data Fig. 4a–d). When grown at 2.5% $O_2$ to induce cell sickling, HIC2 OE SCD cells exhibited dramatically reduced sickling compared with controls (Extended Data Fig. 4e,f), suggesting that raising HIC2 expression induces HbF to levels sufficient to mitigate hemoglobin polymerization.

## HIC2 induces *HBG* production by repressing *BCL11A* transcription.

HIC2 functions as a transcriptional repressor, suggesting that it activates *HBG* transcription indirectly. *BCL11A* mRNA production was significantly lower in HUDEP2 and primary erythroid cells on HIC2 OE (Figs. 2a,c and 3a and Extended Data Fig. 3d). Given that *BCL11A* mRNA can also be post-transcriptionally regulated[15,17], we assessed whether HIC2 OE inhibits *BCL11A* transcription by measuring the amounts of nascent *BCL11A* transcripts. The reduction in *BCL11A* primary transcripts largely mirrored that of the mature *BCL11A* mRNA in both HUDEP2 and primary human erythroblasts (Extended Data Fig. 2g and Fig. 3a), indicating that HIC2 OE inhibits *BCL11A* transcription. Expression of *ZBTB7A* and other known *HBG* regulators did not change significantly on HIC2 OE (Extended Data Fig. 3d). Conversely, depletion of HIC2 in HUDEP2 cells further elevated *BCL11A* (Extended Data Fig. 2c,d). Of note, *BCL11A* mRNA levels were generally commensurate with BCL11A protein levels and declined in a HIC2 dosage-dependent manner (Figs. 2c,d and 3a,b) in accordance with *BCL11A* transcriptional regulation being the predominant mechanism of overall *BCL11A* expression control.

We considered the possibility that HIC2 might affect the expression of other transcription (co-)factors impacting the levels of *BCL11A*, or the transcription factors working independently of *BCL11A*[29,31]. To address this question, we restored *BCL11A* levels in HIC2 OE HUDEP2 cells by expressing *BCL11A* fused to the estrogen receptor (ER) ligand-binding domain[13]. On treatment with tamoxifen, *HBG* transcription was almost completely silenced (a 94% reduction compared with controls; Fig. 2g,h), indicating that *BCL11A* is the critical HIC2 target with regard to globin gene regulation.

### HIC2 directly decommissions erythroid *BCL11A* enhancers.

We profiled chromatin occupancy of endogenous HIC2 in primary erythroblasts derived from newborn $CD34^+$ cells, as well as of an ectopically expressed conditional form of HIC2 (HIC2-ER) in HUDEP2 cells. The most highly enriched core motif under HIC2 or HIC2-ER peaks was TGCCA/C in both cell types, which is the same as the motif found in vitro using SELEX[35] (Extended Data Fig. 5a–c). Importantly, strong HIC2 peaks were observed at the previously described erythroid *BCL11A* +55 and +62 enhancers in both cell types (Fig. 4a), suggesting that HIC2 represses *BCL11A* transcription directly.

To examine the consequences of HIC2 occupancy, we profiled chromatin signatures at the *BCL11A* enhancers. HIC2 OE strongly reduced the active histone mark H3K27ac at all enhancers, and significantly diminished enhancer–promoter contacts (Fig. 4a–c), arguing in favor of a model in which HIC2 directly decommissions enhancer activity.

We measured the activities of the +55, +58 and +62 enhancers in HUDEP2 cells by linking them to a firefly luciferase reporter. The HS2 region from the LCR served as a positive control. As expected, all enhancers strongly induced luciferase expression compared with the promoter-only control. In agreement with the HIC2 chromatin occupancy profile, HIC2 OE repressed all the *BCL11A* enhancers, but not HS2, with the strongest impact on the +55 enhancer (Fig. 4d). Together, these results suggest that HIC2 functions mainly by inhibiting *BCL11A* +55 erythroid enhancer activity.

As HIC2 strongly binds to and represses the +55 enhancer, we inspected the DNA sequence encompassing this region. Two consensus HIC2-binding motifs (motif-1 and motif-2) reside under the HIC2 summit in close proximity to GATA::E-box and GATA motifs, respectively (Extended Data Fig. 6a). To investigate the functions of these motifs, we mutated one nucleotide in one or both motifs and measured enhancer activities using a luciferase reporter assay. Mutating either HIC2 motif alone partially impaired HIC2's ability to repress the enhancer, whereas mutating both motifs abolished it (Extended Data Fig. 6b). Increased activity of the double mutant reporter in the absence of HIC2 OE probably reflects basal level HIC2 activity. These results suggest that both HIC2 motifs convey repression with the HIC2 motif-2 playing a stronger role.

The HIC2 motif-2 resides close to a composite GATA::E-box element and an individual GATA element at the +55 enhancer (Extended Data Fig. 6a). To assess the contributions of these motifs to enhancer activity, we mutated each motif by G-to-C conversion of the GATA core element. Disruption of the sole GATA motif had little effect on enhancer activity and HIC2 was fully capable of repression (Extended Data Fig. 6c). In contrast, disruption of the GATA sequence in the composite element dramatically reduced enhancer activity, which was not further repressed by HIC2. Hence the GATA::E-box motif plays a major role in enhancer activity and is subject to HIC2 regulation.

To investigate the function of HIC2 occupancy at the endogenous *BCL11A* locus, we mutated both HIC2-binding motifs at the endogenous +55 enhancer in HUDEP2 cells, while ensuring intactness of the neighboring GATA elements (Extended Data Fig. 6a,d). ChIP–qPCR confirmed loss of HIC2 binding to the +55 enhancer (Extended Data Fig.

6e). Of note, these mutations also diminished HIC2 occupancy at the +58 enhancer. In spite of clonal variation, a well-known phenomenon in HUDEP2 cells[20], HIC2 motif mutations consistently increased *BCL11A* basal transcription in independently derived clones (Extended Data Fig. 6f), similar to the observations in HIC2-depleted HUDEP2 cells (Extended Data Fig. 2c,d), suggesting that HIC2 functions at least partially through binding to the two motifs. Consistent with the loss of HIC2 binding, the mutations diminished the effects of HIC2 OE on *BCL11A* repression (Extended Data Fig. 6f). The elevated *BCL11A* levels triggered by loss of HIC2 binding led to a strong reduction of *HBG* expression, with only minimal increases even in the presence of HIC2 OE (Extended Data Fig. 6f). Together, these results suggest that HIC2 binding at the +55 enhancer represses *BCL11A* and thereby de-represses *HBG* expression.

### HIC2 imparts developmental control on to the *BCL11A* enhancers.

Given that HIC2 overexpression reduced H3K27ac, chromatin accessibility and enhancer–promoter contacts, we tested whether these findings were linked to the presence of GATA1 because (1) GATA1 is known to interact with the acetyltransferase CREB-binding protein (CBP)[36,37], (2) GATA1 is involved in the formation of long-range contacts[38] and (3) the HIC2-binding motif-2 is very close to the GATA::E-box and GATA motifs (Extended Data Fig. 6a). GATA1 ChIP–seq experiments in HIC2 OE cells revealed only a minimal reduction in GATA1 chromatin occupancy genome wide (Extended Data Fig. 5c). However, GATA1-binding intensity was markedly reduced at the +55 enhancer with modest effects at +58 and +62 in HIC2 OE cells (Fig. 4a and Extended Data Fig. 5c). Hence, HIC2 impairs *BCL11A* enhancer activity in a manner commensurate with loss of GATA1 binding.

Although key developmental regulatory mechanisms of the globin genes are conserved between human and mouse, there are notable exceptions[24], which can be relevant when considering mice as preclinical models. HIC2 proteins are highly conserved with 88.1% identity between human and mouse. To investigate whether the function of HIC2 is also conserved, we determined the binding profiles of overexpressed human HIC2 in differentiated murine erythroid cells (G1E-ER4)[39]. We intersected the two replicates and detected 3,675 high-confidence peaks in G1E-ER4 cells. The motif enriched under the peaks is highly similar to that in human cells (Extended Data Fig. 7a). The higher peak number in murine versus human cells could be due to cell type or species differences, or HIC2 protein levels. The core sequences of the human and murine *BCL11A* +55 enhancers are structured similarly, including the two HIC2, GATA::E-box, and GATA motifs (Extended Data Fig. 6a). Overexpressed human HIC2 bound strongly to the mouse +55 and +62 enhancers and impaired GATA1 binding[40], resulting in repression of *Bcl11a* transcription and upregulation of the murine embryonic globin genes *Hbb-bh1* and *Hbb-y* (Extended Data Fig. 7b,c), consistent with a conserved function of HIC2 in *BCL11A* repression.

As *HIC2* is expressed in a manner opposite to that of *BCL11A* in fetal versus adult erythroid cells (Fig. 1a,b), we performed HIC2 loss-of-function experiments in erythroblasts derived from newborn $CD34^+$ cells. HIC2 depletion significantly increased *BCL11A* mRNA levels and reduced *HBG* transcription without affecting *ZBTB7A* levels (Fig. 5a,b). By RNA-seq we identified 205 upregulated and 80 downregulated genes (FDR < 0.01 and fold-change

>1.5) on HIC2 depletion (Fig. 5c). GSEA-based comparisons of the differentially expressed genes showed high correlation with the changes on HIC2 OE in HUDEP2 cells (Fig. 5d). We observed a strong association between the upregulated (92 of 205, 44.8%) genes and HIC2 binding in newborn erythroblasts (Fig. 5e), suggesting direct repression by HIC2. In contrast, the downregulated genes were not associated with HIC2 binding (Fig. 5e), suggesting indirect regulation. Together, these results suggest that HIC2 directly represses *BCL11A* at the fetal stage of development.

To further examine whether the HIC2-dependent developmental *BCL11A* expression pattern is governed through the erythroid enhancers, we profiled GATA1 binding and chromatin accessibility in fetal-type HUDEP1 cells. Remarkably, at the +55 enhancer GATA1 binding was virtually absent in HUDEP1 when compared with HUDEP2 cells (Fig. 4a). Moreover, the +55 enhancer displayed little to no ATAC signal in HUDEP1 cells, contrasting with that in HUDEP2 cells. Depletion of HIC2 in HUDEP1 cells reversed this pattern with strong increases in GATA1 binding and chromatin accessibility (Fig. 4a and Extended Data Fig. 5c). Importantly, these changes were associated with significantly increased *BCL11A* primary and mature transcript levels (Fig. 5f and Extended Data Fig. 8a). The loss of HIC2 led to strong silencing of the embryonic *HBE1*, a known target of *BCL11A*[29], with lesser effects on *HBG* genes (Fig. 5f and Extended Data Fig. 8b). We also observed a reciprocal induction of the *HBB* gene (Fig. 5f). Together, these results suggest that HIC2 regulates hemoglobin switching by repressing *BCL11A* transcription via decommissioning of *BCL11A* erythroid enhancers in fetal erythroblasts.

### HIC2 competes with GATA1 binding at the *BCL11A* +55 enhancer.

To gain insight into the molecular basis of how HIC2 recognizes its target DNA sequence, we measured DNA-binding affinity and performed crystallization of the HIC2 zinc finger (ZF) domain in complex with DNA. HIC2 contains five ZFs at its carboxyl terminus, with a long linker between ZF1 and ZF2 that is rich in negatively charged glutamate and aspartate (Fig. 6a,b). We measured binding affinities of HIC2 ZF1–5 and ZF2–5 to motif-1 and motif-2 by fluorescence polarization (Extended Data Fig. 8c)[41]. The ZF2–5 displayed the strongest binding to motif-2 with a dissociation constant ($K_D$) of 50 nM, whereas ZF1–5 bound with ~2–3-fold lower affinity (Extended Data Fig. 8d,e), probably due to repulsion between the negatively charged linker and negatively charged deoxyribose-phosphate backbone. For comparison, ZNF410 (ref.[29]) ZF domain binding was >12-fold lower under the same conditions (Extended Data Fig. 8e), confirming the specificity of HIC2.

Next, we determined the crystal structure of the ZF2–5 domain in complex with an oligonucleotide from the +55 enhancer spanning the HIC2 motif and the GATA::E-box element (Supplementary Table 9). Among the four ZFs used in the crystallization, ZF2 is mainly involved in deoxyribose-phosphate interactions (similar to that observed in ZFP568 (ref.[42])), ZF3 and ZF4 are responsible for basepair recognition, whereas ZF5 was not observed in the current structure (Fig. 6c–e), which is common for the ends of ZF arrays when they are not directly involved in base contacts. As in conventional C2H2 ZF proteins, ZF3 and ZF4 follow the rule of three basepairs per ZF, recognizing the TGC (positions

7–9) and CAA (positions 10–12) triplets, respectively (Fig. 6c). The direct base-specific interactions observed are the Gln544 (ZF3) with A7, Thr547 (ZF3) with C8, Arg550 (ZF3) with G9, and Arg572 and Tyr574 of ZF4 with G:C basepair at position 10 (Fig. 6f–i). These interactions are in accordance with the most dominant appositions of glutamine for adenine recognition and arginine for guanine recognition[43]. In addition, the guanidine group of Arg575 of ZF4 bends to span the space of three basepairs at positions 11–13. Arg575 makes a van der Waals contact with the methyl groups of T12 and an electrostatic interaction with the phosphate group of G13 (Fig. 6j). In sum, the two fingers ZF3 and ZF4 protect six basepairs of the HIC2 motif. Importantly, as mentioned above, one single basepair substitution from T:A to A:T at position 7 is sufficient to decrease HIC2 binding, and results in increased luciferase signal in the reporter assay (Extended Data Fig. 6b).

The marked loss of GATA1 occupancy on HIC2 OE and the close proximity of GATA::E-box and HIC2 motif at the +55 enhancer suggested that HIC2 may displace GATA1 from DNA. To test this hypothesis, we modeled a GATA DNA-binding domain (PDB accession no. 4HC9)[44] by superimposing the six common basepairs AGATAA used in the two studies (Fig. 6k). Next, we superimposed the DNA-binding domain on to the respective full-length protein structure predicted by AlphaFold[45]. It is evident that the full-length proteins of GATA1 and HIC2 have much larger footprints (Fig. 6l,m), resulting in mutual repulsion at the +55 enhancer.

To directly test the hypothesis of mutually exclusive binding, we performed electrophoretic mobility shift assays (EMSA) using full length GATA1 and the HIC2 ZF domain (HIC2-ZnF). Both GATA1 and HIC2-ZnF can bind to the *BCL11A* +55 probes individually, but not the probes containing mutations in the respective motifs, confirming specificity (Fig. 6n and Extended Data Fig. 8f). However, no complexes containing both factors were observed (Fig. 6n). When a 24-bp spacer was inserted between the GATA and HIC2 motifs, we observed a larger complex co-bound by GATA1 and HIC2, as both GATA1 and FLAG antibodies super-shifted the bands (Fig. 6n). Together, these results suggest that HIC2 and GATA1 bind to the *BCL11A* +55 enhancer in a mutually exclusive manner.

## Discussion

Several decades have passed since the discovery of the fetal-to-adult developmental switch of the human β-like globin genes, yet many of the key control points of this switch have remained unclear. In the present study, we show that the transcriptional repressor HIC2 regulates the hemoglobin switch by occupying developmental stage-specific enhancer elements in the *BCL11A* gene to repress its activity in fetal erythroid cells. Downregulation of HIC2 levels in adult erythroid cells relieves this repression, allowing for *BCL11A* expression and HbF silencing (Extended Data Fig. 9).

Several lines of evidence support this model: (1) *HIC2* and *BCL11A* are reciprocally expressed in erythroid cells during development; (2) OE of HIC2 in adult cells inhibits *BCL11A* transcription and increases *HBG* expression; (3) forced expression of *BCL11A* reverses *HBG* de-repression in HIC2 OE cells; (4) impairing HIC2 production in fetal cells increases *BCL11A* transcription; (5) removal of HIC2-binding sites in the *BCL11A*

enhancer elevates *BCL11A* expression and reduces the response to HIC2 OE; and (6) the binding of HIC2 and GATA1 is mutually exclusive at the *BCL11A* +55 enhancer site in vitro and in vivo.

The intragenic *BCL11A* enhancer region emerges as a complex regulatory module with some elements, such as GATA1-binding sites, conveying tissue-specific expression of *BCL11A*, whereas others contributing to developmental control. We find that the *BCL11A* erythroid enhancers +55 and +58 display different profiles during development, including histone acetylation, chromatin accessibility and transcription factor binding. This is reminiscent of the signal responsive transcription factor ATF4, which binds to the +55, but not the +58 enhancer[24]. HIC2 preferentially binds to the +55 and +62 enhancers at the fetal stage, but has the strongest repressive effects on the +55 enhancer. Loss of one regulatory element, such as the +55 enhancer, may influence the function of a neighboring regulatory element such as the +58 enhancer, as can happen at complex enhancers[46].

Mechanistically, HIC2 deactivates the *BCL11A* erythroid enhancers by triggering loss of H3K27ac, reducing chromatin accessibility, and reducing GATA1 binding. Although the hierarchical order of these events remains to be elucidated, our observation that HIC2 can directly compete with GATA1 for binding at the +55 enhancer, based on X-ray crystallography and in vitro DNA-binding studies, suggests that GATA1 displacement may be the initiating event. However, HIC2-binding motifs are not in the same proximity to GATA sites at the other *BCL11A* enhancers, suggesting that HIC2 may function by different mechanisms. For example, HIC2 might recruit co-repressors via its BTB/POZ domain that alter chromatin accessibility to impede binding of GATA1 or additional transcription factors, which in turn may reduce enhancer–promoter contacts.

HIC2 is itself expressed in a developmentally dynamic pattern with higher levels in fetal cells compared with adult cells. Of note, even more dramatic differences in HIC2 expression were observed at the protein level, suggesting an additional layer of post-transcriptional regulation of HIC2, which remains to be explored.

Gene expression profiling in response to HIC2 OE revealed changes beyond the *BCL11A* and globin genes. Up- and downregulated genes are enriched in fetal and adult signatures, respectively, suggesting that HIC2 can promote a broad fetal transcriptional program. However, the HIC2-induced gene expression pattern does not represent a full reversion to a fetal-like state because neither *LIN28B* nor *IGF2BP1* was measurably increased on HIC2 OE. This fetal state-promoting function has also been detected in other tissues. For example, HIC2 is crucial for embryonic cardiovascular development[30] and *Hic2* knockout in murine embryonic cardiac tissue triggers premature expression of fetal/adult genes[47].

Although HIC2 depletion in fetal erythroid cells increased *BCL11A* transcription, we did not observe complete silencing of the *HBG* genes and full activation of the *HBB* gene. It is likely that additional fetal stage-specific genes are involved in maintaining the cells in a fetal state. Activation of *HBB* might also require additional adult stage-specific factors. Therefore, depletion of HIC2 alone in fetal cells is not sufficient to enable full adult-type globin expression.

In sum, our study uncovered HIC2 as a key regulator of the fetal-to-adult transition in erythroid development. HIC2 exerts developmental control over the β-globin cluster by inactivating developmental stage-specific *BCL11A* erythroid enhancers, thus placing it upstream of the transcriptional circuitry controlling hemoglobin switching.

## Methods

All the mouse experiments were approved by the Institutional Animal Care and Use Committee at the Children's Hospital of Philadelphia.

### HUDEP cell culture.

HUDEP1 and HUDEP2 cells were maintained in StemSpan SFEM medium (STEMCELL Technologies) supplemented with 1 μg $ml^{-1}$ of doxycycline, 3 units $ml^{-1}$ of erythropoietin, 1 μM dexamethasone, 100 ng $ml^{-1}$ of human stem cell factor and 1% penicillin–streptomycin[23]. Cells were differentiated for 3–5 d in Iscove's modified Dulbecco's medium (IMDM) supplemented with 330 μg $ml^{-1}$ of holotransferrin, 10 μg $ml^{-1}$ of heparin, 10 μg $ml^{-1}$ of insulin, 5% fetal bovine serum (FBS), 3 units $ml^{-1}$ of erythropoietin, 1 μg $ml^{-1}$ of doxycycline and 1% penicillin–streptomycin.

### Human primary CD34$^+$ HSPC culture.

Human $CD34^+$ HSPCs from fetal liver, cord blood and adult peripheral blood were obtained from the Stem Cell and Xenograft Core at the University of Pennsylvania, STEMCELL Technologies and Fred Hutchinson Cancer Research Center, respectively. Human $CD34^+$ HSPCs were cultured in a three-phase culturing system[23]. For phase 1 medium, IMDM was supplemented with 100 ng $ml^{-1}$ of human stem cell factor, 2% penicillin–streptomycin, 1 ng $ml^{-1}$ of interleukin 3 (IL-3), 3 units $ml^{-1}$ of erythropoietin, 200 μg $ml^{-1}$ of holotransferrin, 10 μg $ml^{-1}$ of insulin, 5% human A/B plasma and 10 μg $ml^{-1}$ of heparin. For phase 2 medium, the IL-3 was withdrawn after 9 d of culture and cells were cultured for another 4 d. For phase 3, the cells were cultured with IMDM supplemented with 3 units $ml^{-1}$ of erythropoietin, 10 μg $ml^{-1}$ of insulin, 2% penicillin–streptomycin, 1 mg $ml^{-1}$ of holotransferrin, 5% human A/B plasma and 10 μg $ml^{-1}$ of heparin for 2 d. Cells were collected for ATAC-seq, CUT&RUN and ChIP–seq at day 11 of culture, for reverse transcription (RT)-qPCR and RNA-seq at day 13 of culture and, for HbF staining, western blotting and high-performance–liquid chromatography (HPLC) at day 15 of culture.

### Human SCD CD34$^+$ HSPC isolation and culture.

Human SCD peripheral blood mononuclear cells were isolated from de-identified, discarded, peripheral blood from red blood cell exchange transfusions. The use of de-identified apheresis waste products does not meet the definition of human subject research as determined by the Children's Hospital of Philadelphia Institutional Review Board. SCD $CD34^+$ HSPCs were subsequently isolated using MACS CD34 MicroBead kit (catalog no. 130-100-453, Miltenyi Biotec) according to the manufacturer's instructions[34]. SCD HSPCs were cultured and differentiated using the same protocol for normal human HSPCs.

### COS-7 cell culture.

COS-7 cells were cultured in Dulbecco's modified Eagle's medium (Gibco) supplemented with 10% (v:v) FBS (Bovogen Biologicals) and 1% penicillin–streptomycin–glutamine (Gibco). Cells were lifted for passaging by incubation in 0.05% trypsin–EDTA (Gibco) at 37 °C for 5 min.

### Mouse G1E-ER4 cell culture.

G1E-ER4 cells were cultured with IMDM supplemented with 15% FBS, 2% penicillin–streptomycin, 0.15 mM monothioglycerol, 0.6% Kit ligand-conditioned medium and 2 units $ml^{-1}$ of erythropoietin[24]. Cells were induced with 100 nM estradiol for 24 h.

### OE of HIC2 and BCL11A-ER.

*HIC2* and *BCL11A-ER* complementary DNAs were cloned into lentiviral vector by Gibson cloning. Genes were driven by the *EF1a* promoter. HUDEP2 cells and HSPCs were transduced with lentivirus and sorted for GFP or selected with 2 μg $ml^{-1}$ of puromycin.

### OE of HIC2-ZnF-FLAG and GATA1 for EMSA.

DNA encoding the HIC2 zinc fingers was cloned between the BamHI and EcoRI sites of pcDNA3-FLAG. Recombinant HIC2-ZnF-FLAG and mouse GATA1 (pRcCMV-GATA1) were generated by transfecting COS-7 cells in 100-mm plates with 5 μg of mammalian expression plasmid using FuGENE 6 (Promega) according to the manufacturer's instructions[48]. Transfected cells were incubated at 37 °C for 48 h before harvest. Nuclear extractions were performed using buffer A (10 mM Hepes-KOH, pH 7.9, 10 mM KCl, 1.5 mM $MgCl_2$, 0.5 mM dithiothreitol and 0.2 mM phenylmethylsulfonyl fluoride) and buffer C (20 mM Hepes-KOH pH 7.9, 1.5 mM $MgCl_2$, 25% glycerol, 420 mM NaCl, 0.2 mM EDTA, 0.5 mM dithiothreitol and 0.2 mM phenylmethylsulfonyl fluoride)[49].

### HbF staining and flow cytometry.

Cells were fixed in 0.05% glutaraldehyde for 10 min at room temperature and then permeabilized with 0.1% Triton X-100 for 3 min. Cells were stained with HbF-AF647-conjugated antibody (catalog no. NB110–41084; Novus Biologicals; 1:250 dilution) for 30 min in the dark. Flow cytometry was performed on a BD FACSCanto instrument and data were collected with FACSDiva 8 (Extended Data Fig. 10a,b). Flow cytometry results were analyzed using Flowjo (v.10.3).

### Sickling assay.

SCD cells at day 21 of differentiation were stained with CD71-biotin antibody (Invitrogen, catalog 13-0711-82, clone R17217; 1:200). $CD71^-$ cells were isolated using streptavidin magnetic beads (BioLegend, catalog no. 480016). Cells were then resuspended in 100 μl of HEMOX buffer supplemented with 20 mM glucose and 0.32% bovine serum albumin (BSA). Cells were subsequently incubated under 2.5% $O_2$ at 37 °C for 1 h. Cells were immediately fixed with 200 μl of 2% glutaraldehyde solution. Fixed-cell suspensions were spread on to glass microslides and subjected to microscopic morphological analysis of bright-field images of single-layer cells on an Olympus BX40 microscope fitted with

an Infinity Lite B camera (Olympus). The cells with irregular shapes (sickled shape or protruding structures) were counted as sickled cells. Cell shapes were measured by software CellProfiler[50]. Irregular shape was determined by the cell area compactness (cutoff 1.3). Over 1,300 cells were analyzed for each condition.

### Luciferase reporter assay.

The HS2 (chr11:5280565–5280983, hg38), *BCL11A* +55 (chr2:60498101–60498623, hg38), +58 (chr2:60495079–60495691, hg38) and +62 (chr2:60490852–60491280, hg38) enhancers were PCR amplified from genomic DNA and cloned into a lentiviral vector with firefly luciferase reporter driven by a TATA minimal promoter from pGL4.24 (Promega). *Renilla* luciferase driven by *EF1a* promoter was used as the internal control. The luciferase signals were measured using the Dual-Glo Luciferase Assay kit (Promega, catalog no. E2920) according to the manufacturer's specifications.

### RT–qPCR and RNA-seq.

Cells were homogenized using TRIzol (Life Technologies). Total RNA was extracted using RNeasy Mini Kit (QIAGEN). Reverse transcription was performed using iScript Reverse Transcription Supermix (BioRad). Primers for RT–qPCR are listed in Supplementary Table 1.

RNA-seq libraries were constructed from 500 ng of DNase-treated total RNA. RNA was treated using the NEBNext rRNA Depletion Kit (Human/Mouse/Rat) (New England Biolabs, catalog no. E6310L) for ribosomal RNA depletion, followed by the TruSeq Stranded Total kit (Illumina, catalog no. 20020598) for cDNA synthesis and library preparation according to the manufacturer's specifications. Briefly, first-strand cDNA was synthesized from rRNA-depleted RNA using reverse transcriptase and random primers, followed by second-strand synthesis, end repair, 3′-adenylation and adaptor ligation. Completed libraries were amplified by PCR for 15 cycles. The quality and size (mean 286 bp) of each library were evaluated using the Agilent Bioanalyzer 2100 and the DNA 7500 kit (5067–1504), followed by quantification with real-time PCR using the KAPA Library Quant Kit for Illumina (KAPA Biosystems, catalog no. KK4835). Libraries were then pooled and sequenced in paired-end mode using a P2 flow cell on the NextSeq 2000 to generate 2 × 51 bp reads using Illumina-supplied kits as appropriate.

The sequence reads were processed using the ENCODE3 long RNA-seq pipeline (https://www.encodeproject.org/pipelines/ENCPL002LPE). Differentially expressed genes were identified using DESeq2.

### CRISPR gene knockout in HUDEP1 and HUDEP2 cells.

SgRNAs are listed in Supplementary Table 2. For gene knockout, sgRNAs were cloned into lentiviral vectors[23]. HUDEP1 and HUDEP2 cells with stable spCas9 integration were transduced and sorted based on fluorescent reporters. For the HIC2 motif mutations, sgRNAs and single-strand template DNA were transfected into HUDEP2 cells by electroporation (Amaxa $CD34^+$ kit and program U-008). Single-cell clones were screened by PCR followed by Sanger sequencing confirmation.

### ChIP–qPCR and ChIP–seq.

Undifferentiated HUDEP1, HUDEP2 and G1E-ER4 cells were crosslinked with 1% formaldehyde at room temperature for 10 min. Cells were lysed using 1 ml of cell lysis buffer (10 mM Tris, pH 8, 10 mM NaCl, 0.2% NP-40/Igepal) on ice for 10 min. Nuclei were extracted by incubation with 0.3 ml of nuclear lysis buffer (50 mM Tris, pH 8, 2 mM EDTA, 1% sodium dodecylsulfate (SDS)). Chromatin was fragmented by sonication for 8 min (Bioruptor Pico, Diagenode). Samples were diluted and incubated with antibody pre-bound protein A/G agarose beads overnight at 4 °C. Beads were washed with immunoprecipitation wash buffer I (20 mM Tris, pH 8.0, 2 mM EDTA, 50 mM NaCl, 1% Triton X-100 and 0.1% SDS), high-salt buffer (20 mM Tris, pH 8.0, 2 mM EDTA, 500 mM NaCl, 1% Triton X-100 and 0.01% SDS), immunoprecipitation wash buffer II (10 mM Tris, pH 8, 2 mM EDTA, 0.25 M LiCl, 1% NP-40/Igepal and 1% sodium deoxycholate) and TE (10 mM Tris, pH 8.0 and 1 mM EDTA). Chromatin was eluted from the beads with elution buffer (100 mM $NaHCO_3$, and 1% SDS). Chromatin was de-crosslinked overnight at 65 °C. DNA was purified using the QIAquick PCR Purification kit per the manufacturer's instructions[24]. Antibodies used for ChIP are as follows: hemagglutinin (HA) (Bethyl Laboratories, catalog no. A190–108A; 1:100), GATA1 (human; Abcam, catalog no. Ab11852; 1:100), GATA1 (mouse; Santa Cruz Biotechnology, catalog no. SC-265X, clone N6; 1:100) and H3K27ac (Active Motif, catalog no. 39133; 1:100). ChIP–qPCR primers are listed in Supplementary Table 3.

For ChIP–seq, samples were processed for library construction using Illumina's TruSeq ChIP Sample Preparation Kit (Illumina, catalog no. IP-202–1012) according to the manufacturer's instructions. In brief, DNA fragments were repaired to generate blunt ends, purified using Agencourt AMPure XP Beads (Beckman Coulter, catalog no. A63881), and a single 'A' nucleotide was added to each end. Double-stranded Illumina adaptors containing unique dual indexes (Illumina, catalog no. 20020591) were ligated to the fragments. Ligation products were purified using Agencourt AMPure XP Beads and subjected to size selection using SPRIselect Beads (Beckman Coulter, catalog no. B23318), in which both a left-side size selection was performed at 0.9× volume, and a right-side size selection was performed at 0.6× volume according to manufacturer's specifications. Library fragments were then amplified with 16 cycles of PCR and the products were purified using Agencourt AMPure XP Beads. Constructed libraries were evaluated using the Agilent Bioanalyzer 2100 with the DNA 7500 kit (5067–1504), followed by quantification using real-time PCR with the KAPA Library Quant Kit for Illumina (KAPA Biosystems, catalog no. KK4835). Libraries were then pooled and sequenced in paired-end mode using a P2 flow cell on the NextSeq 2000 using Illumina-supplied kits as appropriate.

Reads were aligned to the reference genome using Bowtie1 using parameters --best --sam --chunkmbs 256 -X 800. Peaks were called using MACS2 (Model-based Analysis of ChIP-Seq data 2). Peaks overlapped with the blacklist and nonspecific regions were further filtered out using bedtools.

### Capture-C.

Capture-C was performed using NlaIII enzyme digestion[24]. Biotin-labeled DNA probes are listed in Supplementary Table 4. The sequencing reads were processed using published scripts[33]. Capture-C interactions of biological replicates were combined and normalized to total interactions.

### Xenotransplantation mouse studies.

NBSGW (NOD.Cg-*Kit*$^{W-41J}$*Tyr*$^{+}$ *Prkdc*$^{scid}$ *Il2rg*$^{tm1Wjl}$/ThomJ) mice were obtained from the Jackson Laboratory. Animals were housed in a barrier facility at the Children's Hospital of Philadelphia. Standard conditions included a 12-h light:dark cycle, 20 °C ambient temperature, 50% humidity and free access to food and water. All experimental protocols were approved by the Institutional Animal Care and Use Committee at Children's Hospital of Philadelphia. Human CD34$^{+}$ HSPCs from adult peripheral blood were obtained from the Fred Hutchinson Cancer Research Center and cultured in StemSpan SFEM medium (STEMCELL Technologies) supplemented with 100 ng ml$^{-1}$ of recombinant human stem cell factor, 100 ng ml$^{-1}$ of recombinant human thrombopoietin, 100 ng ml$^{-1}$ of recombinant human Flt-3 and 1% penicillin–streptomycin. The *HIC2* gene was driven by HS2 enhancer and *ANK1* promoter for erythroid expression. On day 1 of culture, CD34$^{+}$ cells were transduced with either empty vector control or HIC2-overexpressing lentivirus. Transduced cells were selected with 2 μg ml$^{-1}$ of puromycin on days 2 and 3. CD34$^{+}$ cells were then resuspended in 200 μl of phosphate-buffered saline/0.1% BSA and injected retro-orbitally into nonirradiated 6- to 8-week-old male and female NBSGW mice at a dose of 1–1.5 × 10$^{5}$ cells per mouse. At 16 weeks post-transplantation, mice were euthanized using $CO_2$ narcosis and bone marrow extracted to assess for human CD45$^{+}$ engraftment using human-specific CD45-V450 (BD, catalog no. 560367; 1:200) and mouse-specific CD45-PE-eFluor610 (eBioscience, catalog no. 61-0451-82; 1:200). To assess for *HIC2* as well as gene expression levels in engrafted human erythroid cells, CD235a$^{+}$ cells were isolated using phycoerythrin (PE)-positive selection with the EasySep PE Positive Selection Kit II (STEMCELL Technologies, catalog no. 17684) as per manufacturer's instructions in combination with the human-specific CD235a-PE antibody (BD, catalog no. 555570; 1:200). Total RNA was extracted using the RNeasy Mini Kit (QIAGEN, catalog no. 74104) with on-column DNase digestion (QIAGEN, catalog no. 79256). *HIC2*, *HBG* and *HBB* transcript levels were determined by RT–qPCR analysis as described above.

### Antibodies for western blotting.

Primary antibodies used for western blots: HA (Bethyl Laboratories, catalog no. A190–108A; 1:2,000), HIC2 (Proteintech, catalog no. 22788–1-AP; 1:500), BCL11A (Abcam, catalog no. 19487, clone 14B5; 1:1,000), HBG (Novus Biologicals, catalog no. NB110–41084; 1:2,000), HBB (Santa Cruz Biotechnology, catalog no. sc-21757; clone 37–8; 1:1,000), ZBTB7A (Thermo Fisher Scientific, catalog no. 14-3309-82, clone 13E9; 1:1,000) and ACTB (Santa Cruz Biotechnology, catalog no. 47778, clone C4; 1:500).

### ATAC-seq.

ATAC-seq was performed using the Omni-ATAC protocol[26]. Briefly, 50,000–100,000 cells were lysed with 0.1% NP-40, 0.1% Tween-20 and 0.01% digitonin, and incubated with Tagment DNA Enzyme (Illumina) for 30 min at 37 °C. DNA was purified with QIAGEN MinElute Reaction Cleanup Kit. Library fragments were amplified using Phusion High-Fidelity PCR Master Mix with HF Buffer (Thermo Fisher Scientific, catalog no. F531S) and customized primers with unique dual indexes. Libraries were purified using AMPure XP beads (Beckman Coulter, catalog no. A63881) at a 1:1 ratio and according to the manufacturer's instructions. Constructed libraries were evaluated using the Agilent Bioanalyzer 2100 with the DNA 7500 kit (catalog no. 5067–1504), followed by quantification using real-time PCR with the KAPA Library Quant Kit for Illumina (KAPA Biosystems, catalog no. KK4835). Libraries were then pooled and sequenced in paired-end mode using a P2 flow cell on the NextSeq 2000 using Illumina-supplied kits as appropriate.

Reads were aligned to the reference genome using Bowtie1 using parameters --best --sam --chunkmbs 256 -X 800. Peaks were called using MACS2. Peaks overlapped with the blacklist were further filtered out using bedtools.

### CUT&RUN.

CUT&RUN was performed using the CUT&RUN Kit from Cell Signaling (catalog no. 86652S) according to the manufacturer's instructions. Briefly, 0.2–0.8 million cells were harvested and washed twice in 1× wash buffer at room temperature and bound to 15 μl of Concanavalin A beads. Primary antibody incubation was performed in PCR tubes for 2 h at 4 °C with rotation. The following antibodies were used: H3K27ac (Invitrogen, catalog no. MA5–23516), GATA1 (Abcam, catalog no. ab11852). After two washes, pAG-MNase was added and incubated for 1 h at 4 °C with rotation. After three washes, cells were resuspended in 1× wash buffer and pAG-MNase digestion was activated by adding 2 mM $CaCl_2$. Samples were incubated on ice for 45 min, then 2× stop buffer was added to end the reaction. The samples were incubated at 37 °C for 10 min to release the captured chromatin fragments. The sequencing libraries were prepared using the NEBNext Ultra II DNA Library (New England Biolabs, catalog no. E7645) with 15–30 ng of input DNA. Sequencing libraries were pooled and pair-end (2 × 50 bp) sequenced on an Illumina Nextseq 2000 platform.

Reads were aligned to the reference genome using Bowtie1 with parameters --best --sam --chunkmbs 256 -X 800. Peaks were called using MACS2. Peaks overlapped with the blacklist were further filtered out using bedtools.

### Protein purification and crystallography.

The fragments of human HIC2 (NP_055909) comprising ZF1–5 (residues 440–615, pXC2268) and ZF2–5 (residues 503–615, pXC2269) were cloned into pGEX-6P-1 vector with a GST fusion tag[51]. The proteins were kept at −80 °C in the buffer of 20 mM Tris-HCl, pH 7.5, 500 mM NaCl, 5% glycerol and 0.5 mM tris(2-carboxyethyl)phosphine.

The protein–DNA complexes were prepared by mixing purified ZF2–5 with the 16-bp oligonucleotides (5′-AGATAATGCCAACAGT-3′ and 3′-TCTATTACGGTTGTCA-5′) in a 1:1.2 molar ratio followed by dialysis in 20 mM Tris-HCl, pH 7.5, 150 mM NaCl overnight. The complex was further concentrated to ~0.8 mM. The complex crystals were grown using the sitting drop vapor diffusion method at 19 °C with 18% Tacsimate, pH 7.0 (Hampton Research).

Crystals were flash frozen using the crystallization condition supplemented with 20% (v:v) ethylene glycol for cryoprotection. The X-ray diffraction data were collected at SER-CAT 22-ID beamline of the Advanced Photon Source at Argonne National Laboratory utilizing X-ray beam at 1.0-Å (0.1-nm) wavelength and processed by HKL2000 (ref.[52]).

For phasing, Bijvoet pairs were kept separated and ZnSAD with molecular replacement was utilized by PHENIX[53]. PHENIX Xtriage reported a Zn anomalous signal to 5–6 Å. Although an initial 5-Å map allowed identification of DNA, utilization of an ideal 16-mer B-DNA as search model allowed use of the full resolution of the dataset and gave a better map for side-chain identification. Molecular replacement was performed with the PHENIX PHASER module. Zn atom positions were verified from the anomalous map. PHENIX REFINE was used for refinement with 5% randomly chosen reflections for validation by the $R_{free}$ value. COOT was further used for model building and corrections between refinement rounds[54]. Structure quality was analyzed during PHENIX refinements and later validated by the Protein Data Bank (PDB) validation server. Molecular graphics were generated using PyMol (Schrödinger, LLC).

### Fluorescence polarization DNA-binding assay.

The fluorescence polarization method was used to measure the binding affinity with Synergy 4 Microplate Reader (Bio Tek)[41]. Aliquots (10 nM) of 6-carboxyfluorescein-labeled HIC2 motif-1 and motif-2 oligonucleotides were incubated with varied amount of proteins (0–2.5 μM) in 20 mM Tris-HCl, pH 7.5, 150 mM NaCl and 5% glycerol for 10 min at room temperature before measurement. The data were processed with equation (mP) = (maximum mP) × $[C]/(K_D + [C])$ + (baseline mP), in which mP is millipolarization and $[C]$ is protein concentration. The $K_D$ value for each protein–DNA interaction was derived from two or three replicated experiments.

### Electrophoretic mobility shift assays.

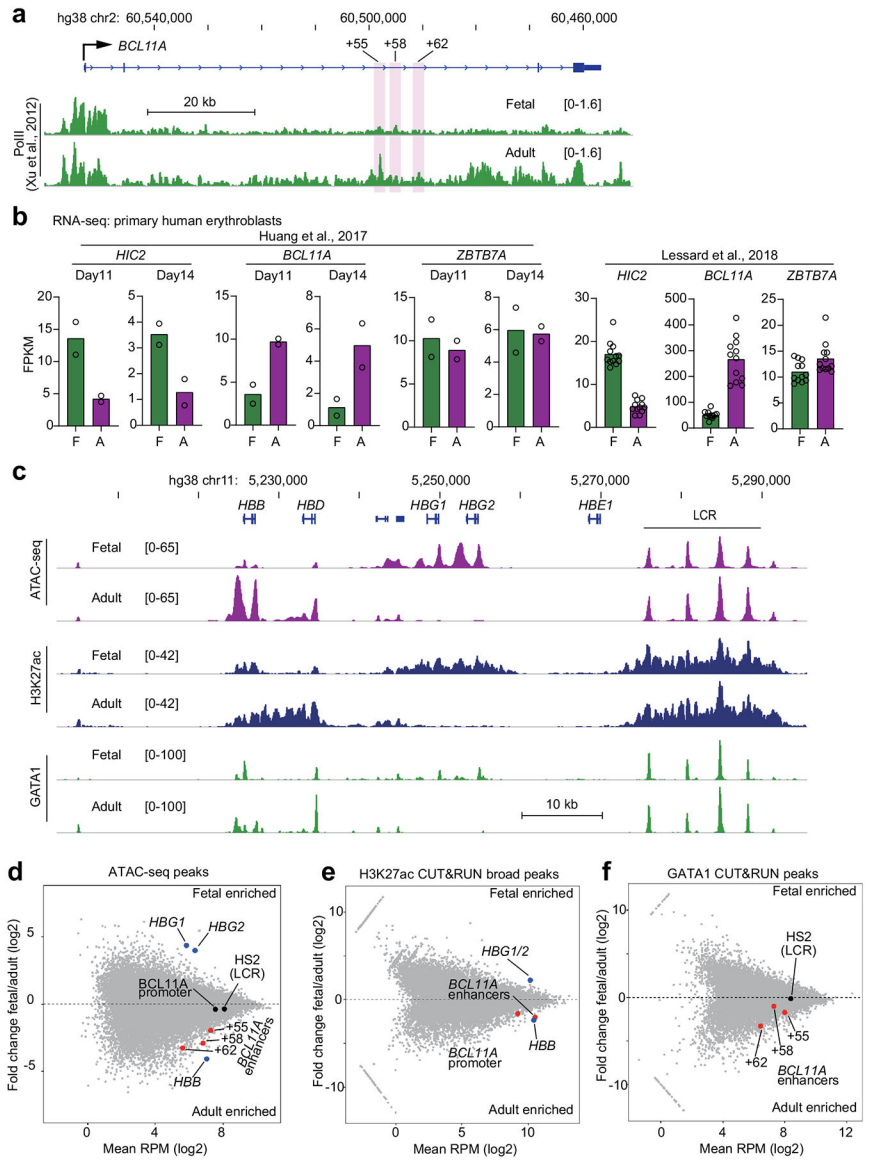
Oligonucleotides used in radiolabeled probes are listed in Supplementary Table 8. The top strand for each probe was labelled with $^{32}$P from $\gamma$–$^{32}$P-labeled ATP (Perkin Elmer) using T4 PNK (New England Biolabs), before annealing the bottom strand by slow cooling from 100 °C to room temperature. The annealed probes were purified using quick spin columns for radiolabeled DNA purification (Roche). Transcription factors were overexpressed and harvested from COS-7 cells, and ‘empty’ extract containing no overexpressed protein was used to aid identification of background bands caused by endogenous protein binding. Antibody for GATA1 (Santa Cruz Biotechnology, catalog no. N6 sc-265) and FLAG (Sigma-Aldrich, catalog no. F1804) were used as indicated to identify their respective targets on the gel. Complexed samples were loaded on 6% native polyacrylamide gel in TBE buffer (45

mM Tris, 45 mM boric acid and 1 mM EDTA). Electrophoresis was performed at 4 °C and 250 V for 100 min and vacuum dried before exposing a FUJIFILM BAS Cassette2 phosphor screen overnight. Imaging was performed on a GE Typhoon FLA 9500 fluorescent image analyzer.

### Reporting summary.

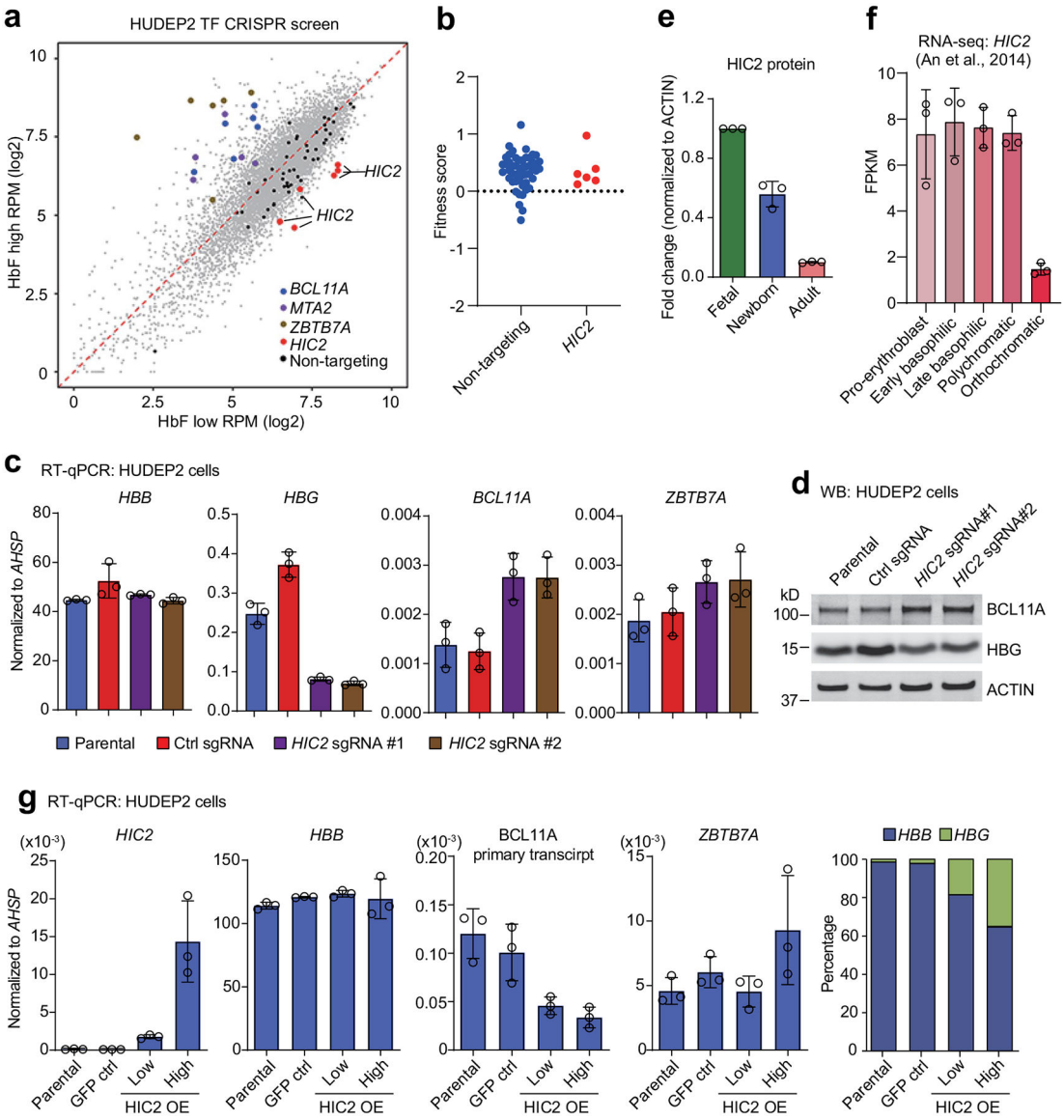
Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Extended Data




**Extended Data Fig. 1 |. Developmental regulation of *BCL11A* occurs primarily at the transcriptional level.**

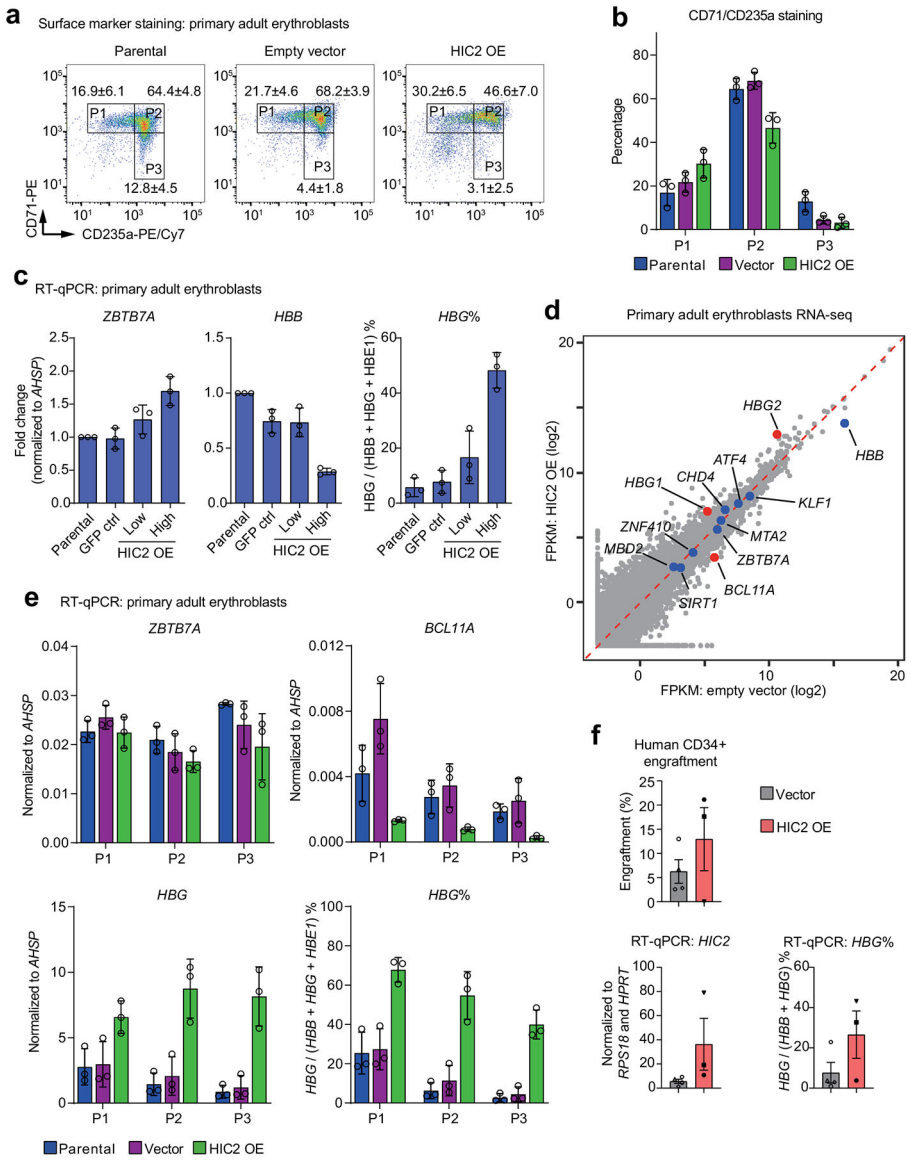
**a**, PolII ChIP-seq tracks from primary human fetal and adult erythroblasts of the *BCL11A* locus. Results were obtained from published data. The *BCL11A* erythroid enhancers +55, +58 and +62 are highlighted in pink. **b**, *HIC2*, *BCL11A* and *ZBTB7A* mRNA levels in primary human fetal and adult erythroblasts. RNA-seq data were obtained from previously published results. FPKM, fragments per kilobase per million; F, fetal; A, adult. **c**, Representative tracks of ATAC-seq, H3K27ac and GATA1 CUT&RUN from primary human fetal and adult erythroblasts of the *β-globin* locus. Results are normalized to reads in peaks. **d**, MA plot of read counts of ATAC-seq of fetal and adult erythroblasts. *BCL11A* enhancers are highlighted in red. *BCL11A* promoter, *HBG1/2*, *HBB* and HS2 from LCR (locus control region) enhancer are highlighted in blue. **e**, MA plot of read counts of H3K27ac CUT&RUN of fetal and adult erythroblasts. *BCL11A* promoter and enhancers are highlighted in red. *HBG1/2* and *HBB* are highlighted in blue. **f**, MA plot of read counts of GATA1 CUT&RUN of fetal and adult erythroblasts. *BCL11A* enhancers are highlighted in red. HS2 from LCR is highlighted in blue.




**Extended Data Fig. 2 |. A CRISPR screen identifies HIC2 as a *BCL11A* repressor and *HBG* activator.**

**a**, Scatter plot of CRISPR screen in HUDEP2 cells. Each dot represents a sgRNA. sgRNAs targeting to *HIC2* are highlighted in red. RPM, reads per million reads. **b**, Cell fitness scores of non-targeting and *HIC2* sgRNAs. Each dot represents a sgRNA. Fitness scores were calculated according to the abundance of the sgRNAs in the cells of Day13 compared to Day3 post sgRNA transduction. Statistical test compares non-targeting and *HIC2* sgRNAs, $P = 0.8931$ by two-tailed Student's *t*-test. **c**, *HBB* (sgRNA #1, $P = 0.8051$; sgRNA #2, $P = 0.9997$), *HBG* (sgRNA #1, $P < 0.0001$; sgRNA #2, $P < 0.0001$), *BCL11A* (sgRNA #1, $P = 0.0112$; sgRNA #2, $P = 0.0118$) and *ZBTB7A* (sgRNA #1, $P = 0.1816$; sgRNA #2, $P = 0.1502$) mRNA measured by RT-qPCR in HIC2 depleted HUDEP2 cells. Results are normalized to *AHSP* and shown as mean ± s.d. ($n = 3$). Statistical tests compare HIC2 sgRNAs and parental samples. P-values were calculated by one-way ANOVA test. **d**, Western blot of extracts from HUDEP2 cells upon HIC2 depletion. Results are

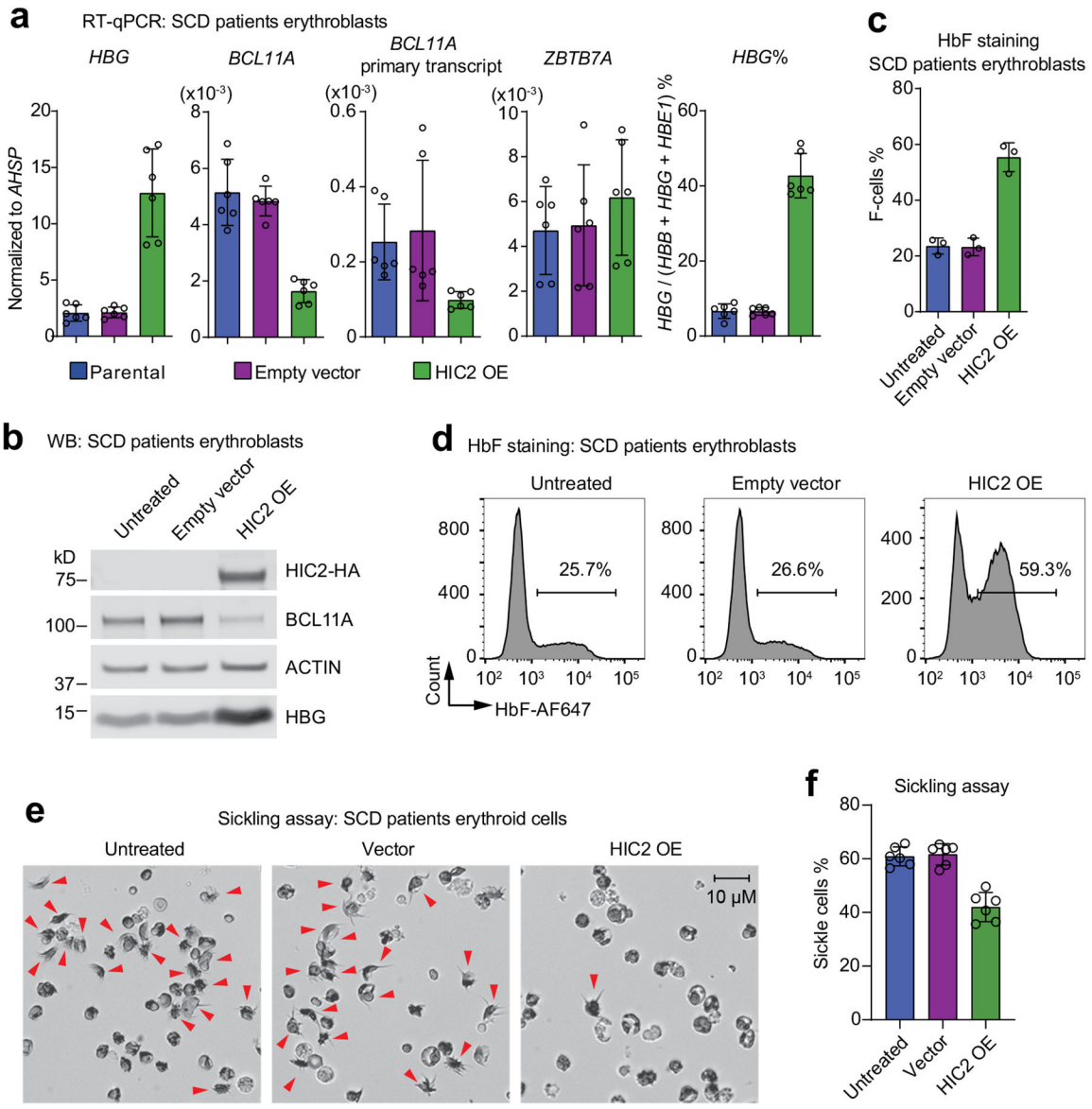
representative of three biological replicates. **e**, Quantitative of HIC2 protein levels of WB blots in primary human erythroblasts from three sequential developmental stages. Results are normalized to ACTIN and shown as mean ± s.d. **f**, *HIC2* expression levels during erythropoiesis. Results were obtained from published RNA-seq data. FPKM, fragments per kilobase per million. **g**, *HIC2*, *HBB* (low, $P = 0.3948$; high, $P = 0.7553$), *BCL11A* primary transcript (low, $P = 0.0058$; high, $P = 0.0023$), ZBTB7A (low, $P > 0.9999$; high, $P = 0.0918$) and percentage of *HBG* mRNA measured by RT-qPCR in HIC2 OE HUDEP2 cells. Results are normalized to *AHSP* and shown as mean ± s.d. ($n = 3$). Statistical tests compare HIC2 OE and parental samples. *P*-values were calculated by one-way ANOVA test.




**Extended Data Fig. 3 |. HIC2 overexpression represses *BCL11A* transcription and augments fetal globin production in primary erythroid cells.**

**a**, Representative flow cytometry results of surface markers CD71 and CD235a staining in primary human adult erythroblasts. **b**, Percentage of each population in panel **a)**. Results are

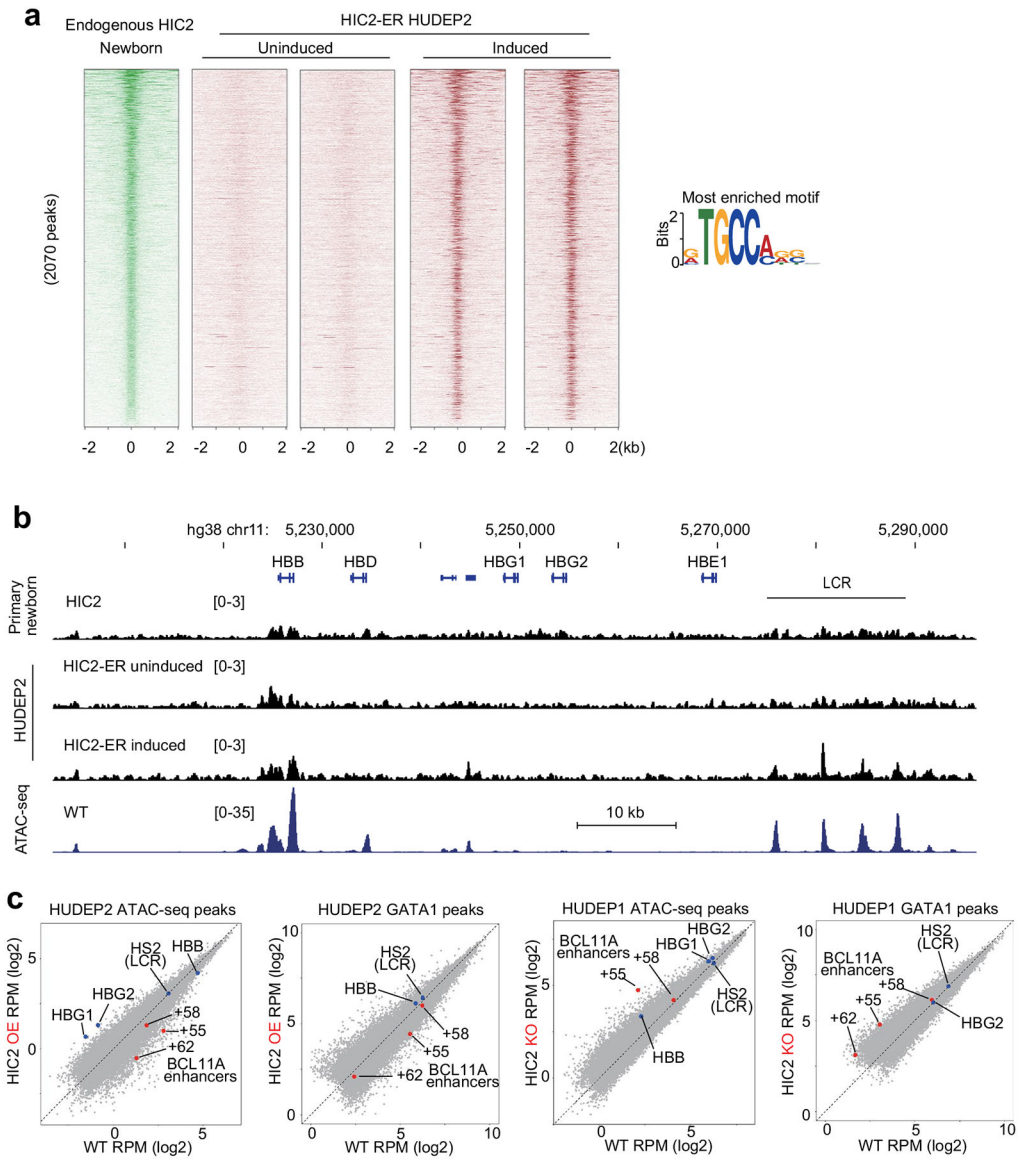
shown as mean ± s.d. ($n$ = 3). Statistical tests compare parental and HIC2 OE samples; P1, $P$ = 0.0605; P2, $P$ = 0.0222; P3, $P$ = 0.0314 by two-tailed Student's $t$-test. **c**, *ZBTB7A* (low, $P$ = 0.2160; high, $P$ = 0.0031), *HBB* (low, $P$ = 0.0133; high, $P$ < 0.0001) and percentage of *HBG* (low, $P$ = 0.1596; high, $P$ = 0.0001) measured by RT-qPCR in primary adult human erythroblasts. Results are normalized to *AHSP* and shown as mean ± s.d. ($n$ = 3). Statistical tests compare HIC2 OE and parental samples. $P$-values were calculated by one-way ANOVA test. **d**, Scatter plot of the RNA-seq results in primary human adult erythroblasts. FPKM, fragments per kilobase per million. **e**, *ZBTB7A* (P1, $P$ = 0.9319; P2, $P$ = 0.0948; P3, $P$ = 0.0889), *BCL11A* (P1, $P$ = 0.0441; P2, $P$ = 0.0298; P3, $P$ = 0.0040), *HBG* (P1, $P$ = 0.0243; P2, $P$ = 0.0063; P3, $P$ = 0.0055) and percentage of *HBG* (P1, $P$ = 0.0040; P2, $P$ = 0.0028; $P$3, P = 0.0011) measured by RT-qPCR in the stage-matched populations in panel **a)**. Results are normalized to *AHSP* and shown as mean ± s.d. ($n$ = 3). Statistical tests compare parental and HIC2 OE samples. P-values were calculated by two-tailed Student's $t$-test. **f**, NBSGW mice xenotransplantation experiment. Results are shown as mean ± s.e.m., each symbol represents one mouse (Vector, $n$ = 4 and HIC2 OE, $n$ = 3).




**Extended Data Fig. 4 |. HIC2 overexpression represses *BCL11A* transcription, augments fetal globin production, and reduces SCD cell sickling.**

**a**, *HBG* ($P < 0.0001$), *BCL11A* mature ($P < 0.0001$) and primary transcript ($P = 0.0043$), *ZBTB7A* ($P = 0.2917$) and percentage of *HBG* mRNA ($P < 0.0001$) measured by RT-qPCR in primary human erythroblasts derived from SCD patient CD34+ cells. Results are normalized to *AHSP* and shown as mean ± s.d. (3 independent donors with 2 technical replicates for each donor). Statistical tests compare HIC2 OE and parental samples. *P*-values were calculated by two-tailed Student's *t*-test. **b**, Western blot of extracts from primary human erythroblasts derived from SCD patient CD34+ cells upon HIC2 OE. Results are representative of three biological replicates. **c**, Flow cytometry results of HbF staining in erythroblasts derived from SCD patient HSPCs. Results are shown as mean ± s.d. ($n = 3$). Statistical tests compare HIC2 OE and parental samples by two-tailed Student's t-test, $P = 0.0007$. **d**, Representative flow cytometry results of HbF staining in erythroblasts derived

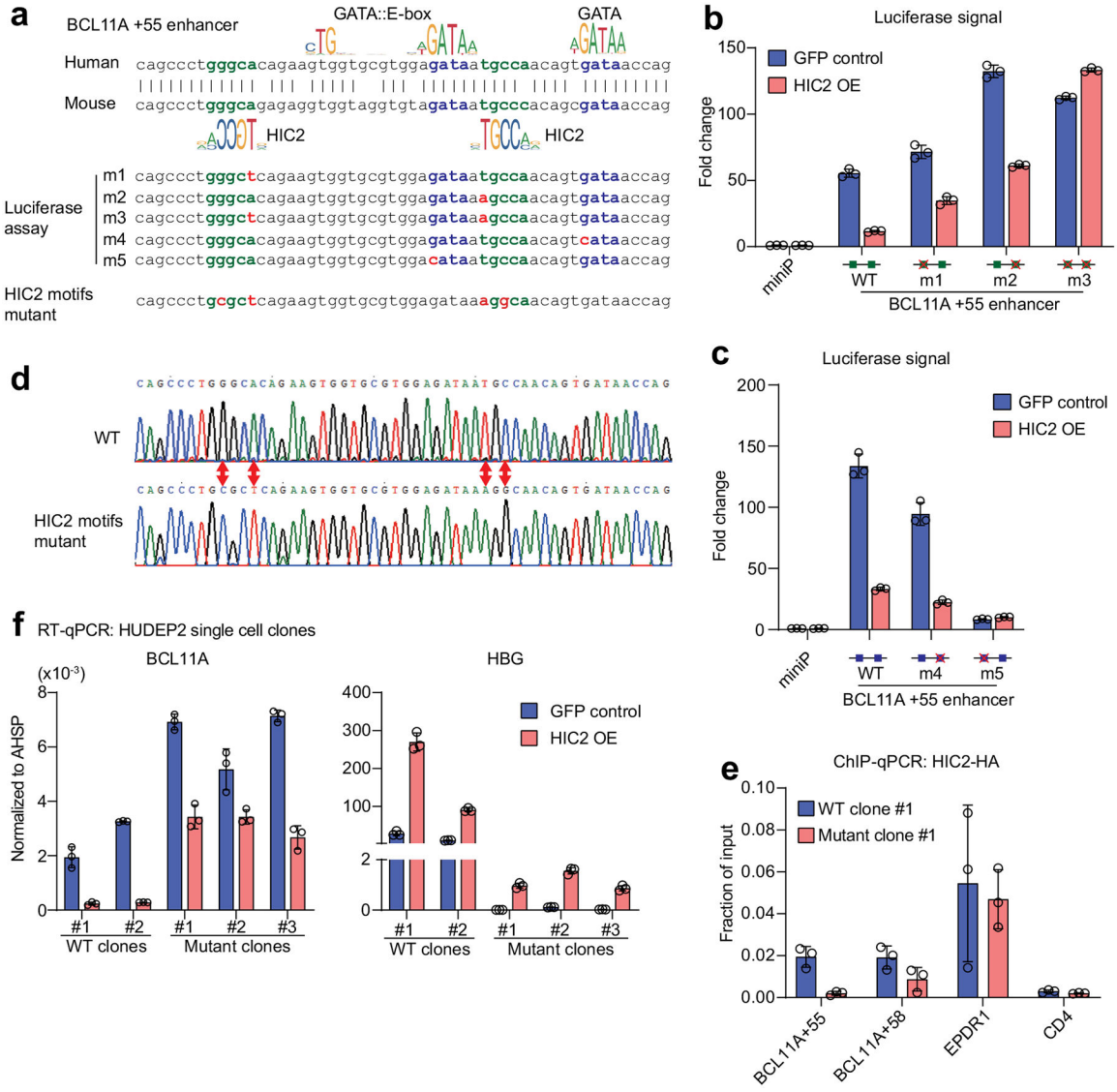
from SCD patient HSPCs. **e**, Representative brightfield images of terminally differentiated SCD patient-derived erythroid cells following incubation at 2.5% O2 for 1 hour followed by fixation. Sickled cells are indicated by red arrowheads. **f**, Percentage of sickled cells quantified following low-O2 sickling assay. Results are mean ± s.d. (3 independent donors with 2 technical replicates for each donor). Statistical tests compare HIC2 OE and parental samples by two-tailed Student's *t*-test, $P < 0.0001$.




**Extended Data Fig. 5 |. Chromatin binding profiles of HIC2 and its effects on GATA1 occupancy and chromatin accessibility.**

**a**, Heatmaps of endogenous HIC2 ChIP–seq in primary newborn erythroblasts and HIC2-ER ChIP-seq in HUDEP2 cells, and the most enriched motif. **b**, Tracks of endogenous HIC2 ChIP-seq in primary newborn erythroblasts, HIC2-ER and ATAC-seq in HUDEP2 cells of the β-globin locus. The HIC2 and HIC2-ER ChIP-seq were normalized to total reads. LCR, locus control region. **c**, Scatter plots of ATAC-seq and GATA1 ChIP-seq in HUDEP2 and

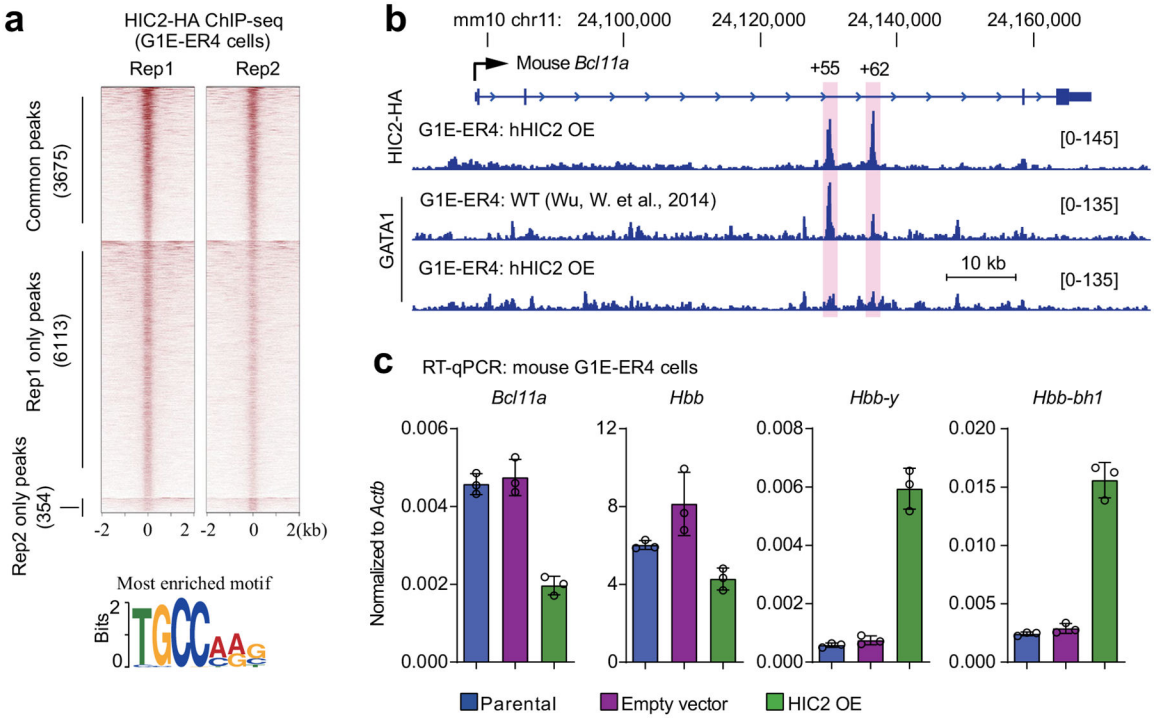
HUDEP1 cells upon HIC2 perturbations. *BCL11A* enhancers are highlighted in red. *HBG1*, *HBG2*, *HBB* and HS2 from LCR are highlighted in blue. RPM, reads per million reads.




**Extended Data Fig. 6 |. HIC2 imparts developmental control onto the *BCL11A* enhancers.**

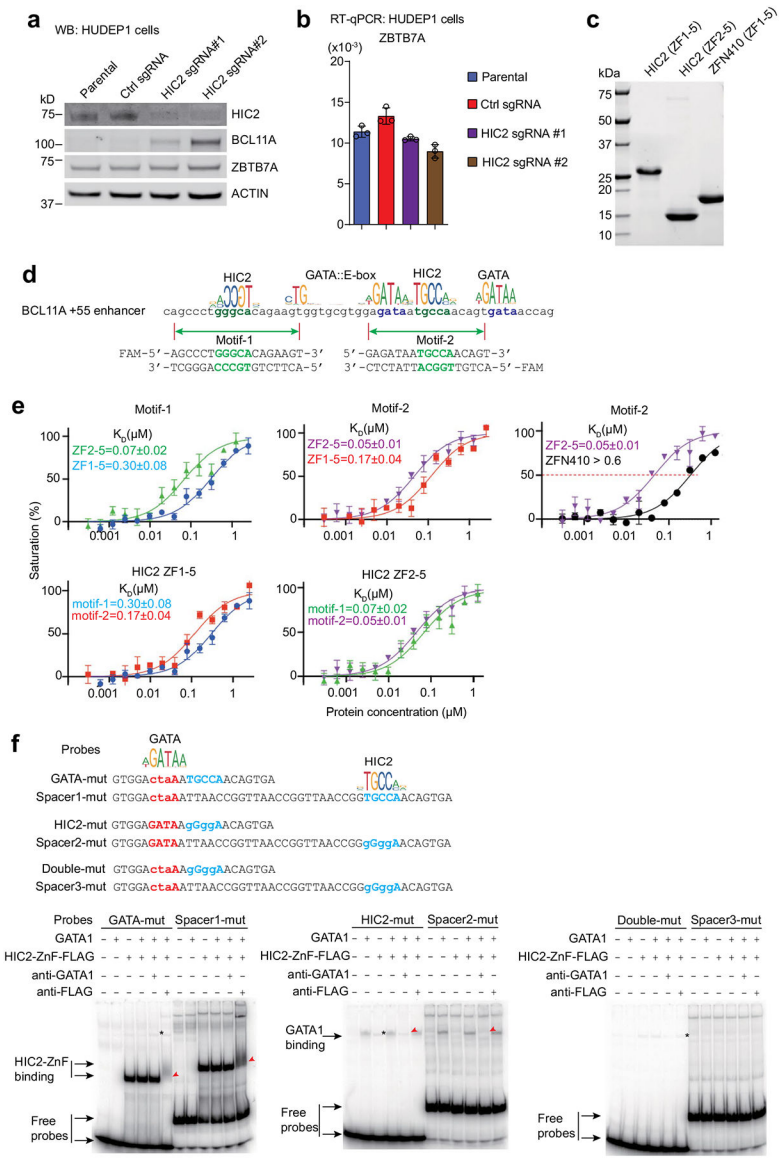
**a**, The core sequence of human and mouse *BCL11A* +55 enhancer. The point mutations in the luciferase reporter constructs in panels **b)** and **c)**, and CRISPR-mediated knock-in mutations in HUDEP2 cells are highlighted in red. **b**, Luciferase reporter assay of human *BCL11A* +55 enhancers in HUDEP2 cells. The two HIC2 motifs in (a) are indicated by green boxes. Red X-crosses indicate the mutated HIC2 motifs. Results are mean ± s.d. ($n = 3$). Statistical tests compare GFP control and HIC2 OE samples by two-tailed Student's *t*-test; WT, $P < 0.0001$; m1, $P = 0.0004$; m2, $P < 0.0001$; m3, $P < 0.0001$. **c**, Luciferase reporter assay of human *BCL11A* +55 enhancers in HUDEP2 cells. The two GATA motifs in panel **a)** are indicated by blue boxes. Red X-crosses indicate the mutated GATA motifs. Results are mean ± s.d. ($n = 3$). Statistical tests compare GFP control

and HIC2 OE samples by two-tailed Student's *t*-test; WT, $P < 0.0001$; m4, $P = 0.0002$; m5, $P = 0.0087$. **d**, Representative Sanger sequencing traces of WT and HIC2 motifs edited HUDEP2 single cell clones. Mutations are indicated by red arrows. **e**, ChIP-qPCR of HIC2-HA in representative WT and HIC2 motifs edited HUDEP2 single cell clones. Results are mean ± s.d. ($n = 3$). Statistical tests compare WT clone #1 and Mutant clone #1 samples by two-tailed Student's *t*-test; *BCL11A*+55, $P = 0.0040$; *BCL11A*+58, $P = 0.0833$; *EPDR1*, $P = 0.7585$; *CD4*, $P = 0.0778$. **f**, *BCL11A* and *HBG* mRNA measured by RT-qPCR in independent WT and HIC2 motifs edited HUDEP2 single cell clones. Results are normalized to AHSP and shown as mean ± s.d. ($n = 3$). Statistical tests compare GFP control and HIC2 OE samples by two-tailed Student's *t*-test; *BCL11A* (WT clone #1, $P = 0.0016$; WT clone #2, $P < 0.0001$; Mutant clone #1, $P = 0.0003$; Mutant clone #2, $P = 0.0192$; Mutant clone #3, $P < 0.0001$); *HBG* (WT clone #1, $P < 0.0001$; WT clone #2, $P < 0.0001$; Mutant clone #1, $P < 0.0001$; Mutant clone #2, $P < 0.0001$; Mutant clone #3, $P = 0.0001$).




**Extended Data Fig. 7 |. The repression of *BCL11A* by HIC2 is conserved between human and mouse.**

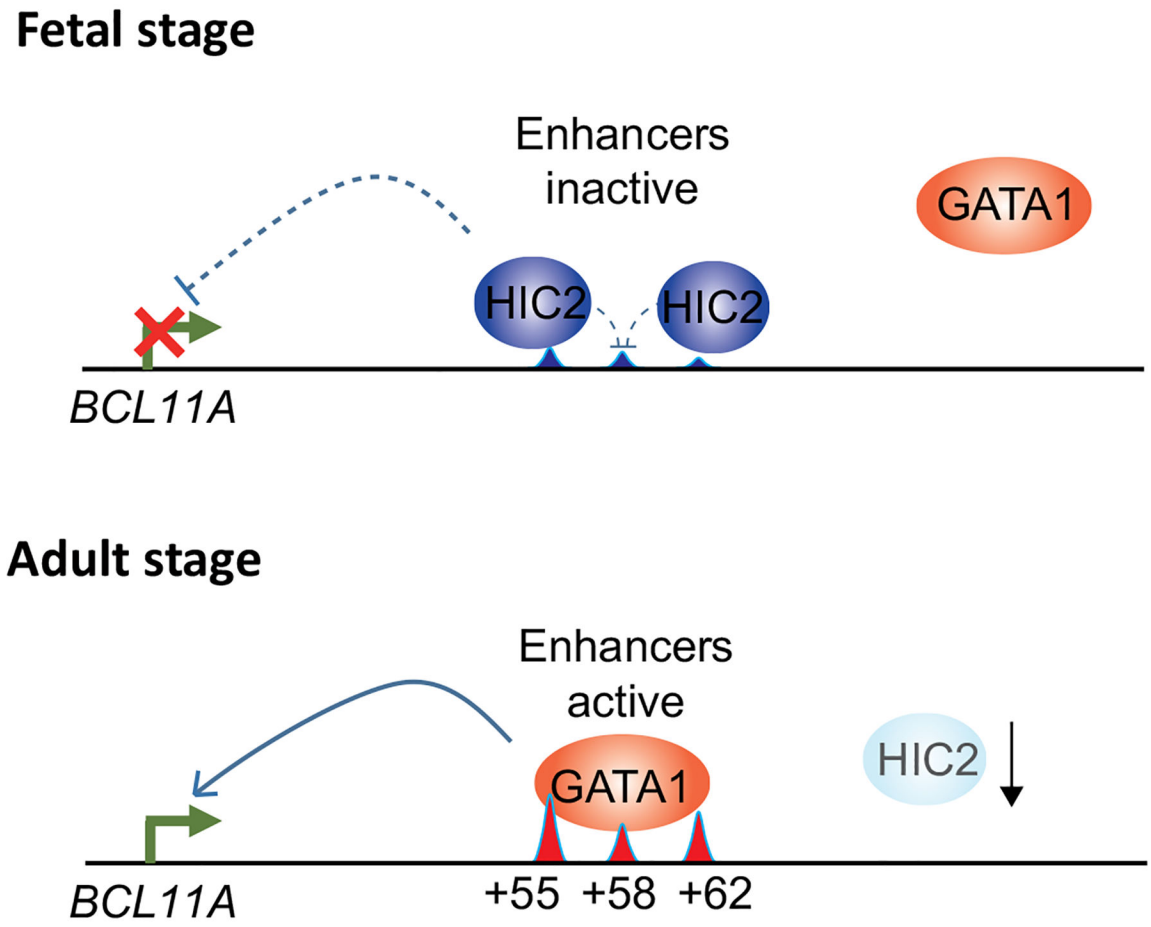
**a**, Heatmaps of HIC2-HA ChIP-seq in induced mouse G1E-ER4 cells and the most enriched motif. **b**, ChIP-seq tracks of hHIC2-HA and GATA1 in G1E-ER4 cells. The *mBcl11a* enhancers homologous to human +55 and +62 are highlighted in pink. GATA1 ChIP-seq of WT G1E-ER4 cells was obtained from published data. **c**, *Bcl11a* ($P = 0.0002$), *Hbb* ($P = 0.0076$), *Hbb-y* ($P = 0.0002$) and *Hbb-bh1* ($P = 0.0001$) mRNA levels as measured by RT-qPCR in the hHIC2 OE G1E-ER4 cells. Results are normalized to *Actb* and shown as mean ± s.d. ($n = 3$). Statistical tests compare HIC2 OE and parental samples. *P*-values were calculated by two-tailed Student's *t*-test.



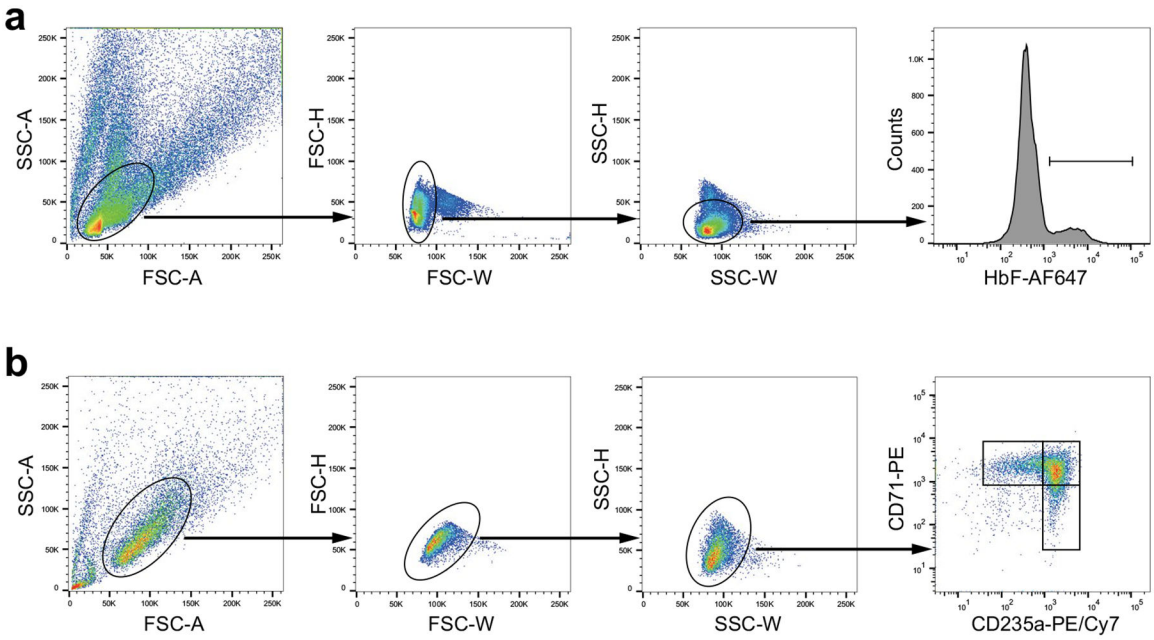

**Extended Data Fig. 8 |. HIC2 developmentally controls *HBG* by repressing *BCL11A* transcription.**

**a**, Western blot with indicated antibodies of extracts from HUDEP1 cells upon HIC2 depletion. Results are representative of two biological replicates. **b**, *ZBTB7A* mRNA (sgRNA #1, $P = 0.4015$; sgRNA #2, $P = 0.0110$) measured by RT-qPCR in HUDEP1 cells upon HIC2 depletion. Results are normalized to *AHSP* and shown as mean ± s.d. ($n = 3$). Statistical tests compare *HIC2* sgRNAs and parental samples. *P*-values were calculated by one-way ANOVA test. **c**, Purified recombinant proteins used in the crystallography and fluorescence polarization experiments. Results were performed with one biological replicate. **d**, The FAM-labeled oligonucleotides of HIC2 binding motif-1 and motif-2 at *BCL11A* +55 enhancer. **e**, Comparison of binding affinities between HIC2 ZF1–5 and ZF2–5, and ZNF410 ZF1–5 at the two HIC2 binding motifs. **f**, EMSA of GATA1 and HIC2 using probes from the *BCL11A* +55 enhancer, wild type or with mutated binding motifs. Red arrowheads indicate

the super-shifted bands; asterisks indicate non-specific bands. Results were performed with one biological replicate.




**Extended Data Fig. 9 |. Model of HIC2 regulating *BCL11A* transcription during development.** At fetal stage, HIC2 binds to *BCL11A* enhancers and blocks GATA1 binding. At adult stage, HIC2 is downregulated and GATA1 binds to the enhancers promoting *BCL11A* transcription.




**Extended Data Fig. 10 |. Flow cytometry gating strategies.**

**a**, Gating strategy for HbF staining experiments (Fig. 3d and Extended Data Fig. 4d). **b**, Gating strategy for CD71/CD235a staining experiments (Extended Data Fig. 3c).

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements

We thank the CHOP flow cytometry core for assistance with cell sorting and members of the Blobel laboratory for helpful comments and discussions. We thank R. Kurita and Y. Nakamura for sharing HUDEP1 and HUDEP2 cells. This work was supported by National Institutes of Health (NIH) grants from the National Heart, Lung, and Blood Institute (grant no. HL119479) and research funding from Pfizer (to G.A.B.); the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK; grant no. R24DK106766 to G.A.B. and R.C.H.); NIH/NIDDK Mentored Clinical Scientist Research Career Development Award (no. K08 DK129716), the Doris Duke Charitable Foundation Physician Scientist Fellowship grant (no. 2020062), and American Society of Hematology Scholar Award (to S.A.P.); a fellowship by the Cooley's Anemia Foundation (to X.L.); T32 training grant (no. HL007150-42) and an American Society of Hematology Research Training Award for Fellows (to E.K.); a National Health and Medical Research Council Australia grant (to M.C.); an Australian government Research Training Program Scholarship (toH.W.B.); the St. Jude Children's Research Hospital Collaborative Research Consortium on Novel Gene Therapies for Sickle Cell Disease; and the National Institute of General Medical Sciences grant (no. R35GM134744) and the Cancer Prevention & Research Institute of Texas grant (no. RR160029) (to X.C.). Cooperative Centers of Excellence in Hematology at Fred Hutchinson Cancer Research Center was supported by NIDDK (DK106829). We thank the DiGaetano family for their generous support.

## Data availability

The ChIP–seq, RNA-seq, ATAC-seq, CUT&RUN and Capture-C data have been deposited into the Gene Expression Omnibus database (accession no. GSE173587). The X-ray structures of the HIC2 ZF2–5 domain with bound DNA have been submitted to PDB (accession no. 7TXC). Publicly available datasets used in this manuscript: structure of DNA binding by GATA3: PDB, accession no. 4HC9; Pol II ChIP–seq, accession no. GSE36994;

RNA-seq of fetal and adult erythroblasts, accession nos. GSE102201 and GSE90878; RNA-seq of erythroid differentiation, accession no. GSE53983; G1E-ER4 GATA1 ChIP–seq, accession no. GSE36029. Source data are provided with this paper.

## References


1. Orkin SH Globin gene regulation and switching: circa 1990. Cell 63, 665–672 (1990). [PubMed: 2225071]
2. Sankaran VG & Orkin SH The switch from fetal to adult hemoglobin. Cold Spring Harb. Perspect. Med 3, a011643 (2013). [PubMed: 23209159]
3. Huang P et al. Comparative analysis of three-dimensional chromosomal architecture identifies a novel fetal hemoglobin regulatory element. Genes Dev. 31, 1704–1713 (2017). [PubMed: 28916711]
4. Liu N et al. Transcription factor competition at the γ-globin promoters controls hemoglobin switching. Nat. Genet 53, 511–520 (2021). [PubMed: 33649594]
5. Xu J et al. Combinatorial assembly of developmental stage-specific enhancers controls gene expression programs during human erythropoiesis. Dev. Cell 23, 796–811 (2012). [PubMed: 23041383]
6. Doerfler PA et al. Activation of γ-globin gene expression by GATA1 and NF-Y in hereditary persistence of fetal hemoglobin. Nat. Genet 53, 1177–1186 (2021). [PubMed: 34341563]
7. Forget BG Molecular basis of hereditary persistence of fetal hemoglobin. Ann. N.Y. Acad. Sci 850, 38–44 (1998). [PubMed: 9668525]
8. Driscoll MC, Dobkin CS & Alter BP Gamma delta beta-thalassemia due to a de novo mutation deleting the 5' beta-globin gene activation-region hypersensitive sites. Proc. Natl Acad. Sci. USA 86, 7470–7474 (1989). [PubMed: 2798417]
9. Kioussis D, Vanin E, deLange T, Flavell RA & Grosveld FG β-Globin gene inactivation by DNA translocation in γβ-thalassaemi. Nature 306, 662–666 (1983). [PubMed: 6318113]
10. Topfer SK et al. Disrupting the adult globin promoter alleviates promoter competition and reactivates fetal globin gene expression. Blood 139, 2107–2118 (2022). [PubMed: 35090172]
11. Menzel S et al. A QTL influencing F cell production maps to a gene encoding a zinc-finger protein on chromosome 2p15. Nat. Genet 39, 1197–1199 (2007). [PubMed: 17767159]
12. Lettre G et al. DNA polymorphisms at the *BCL11A*, *HBS1L-MYB*, and beta-globin loci associate with fetal hemoglobin levels and pain crises in sickle cell disease. Proc. Natl Acad. Sci. USA 105, 11869–11874 (2008). [PubMed: 18667698]
13. Martyn GE et al. Natural regulatory mutations elevate the fetal globin gene via disruption of *BCL11A* or *ZBTB7A* binding. Nat. Genet 50, 498–503 (2018). [PubMed: 29610478]
14. Liu N et al. Direct promoter repression by *BCL11A* controls the fetal to adult hemoglobin switch. Cell 173, 430–442.e17 (2018). [PubMed: 29606353]
15. Lee YT et al. LIN28B-mediated expression of fetal hemoglobin and production of fetal-like erythrocytes from adult human erythroblasts ex vivo. Blood 122, 1034–1041 (2013). [PubMed: 23798711]
16. Basak A et al. Control of human hemoglobin switching by *LIN28B*-mediated regulation of *BCL11A* translation. Nat. Genet 52, 138–145 (2020). [PubMed: 31959994]
17. de Vasconcellos JF et al. IGF2BP1 overexpression causes fetal-like hemoglobin expression patterns in cultured human adult erythroblasts. Proc. Natl Acad. Sci. USA 114, E5664–E5672 (2017). [PubMed: 28652347]
18. Sankaran VG et al. Human fetal hemoglobin expression is regulated by the developmental stage-specific repressor BCL11A. Science 322, 1839–1842 (2008). [PubMed: 19056937]
19. Lessard S, Beaudoin M, Orkin SH, Bauer DE & Lettre G 14q32 and let-7 microRNAs regulate transcriptional networks in fetal and adult human erythroblasts. Hum. Mol. Genet 27, 1411–1420 (2018). [PubMed: 29432581]
20. Khandros E et al. Understanding heterogeneity of fetal hemoglobin induction through comparative analysis of F and A erythroblasts. Blood 135, 1957–1968 (2020). [PubMed: 32268371]

21. Zhou D, Liu K, Sun C-W, Pawlik KM & Townes TM KLF1 regulates BCL11A expression and γ- to β-globin gene switching. Nat. Genet 42, 742–744 (2010). [PubMed: 20676097]
22. Canver MC et al. BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. Nature 527, 192–197 (2015). [PubMed: 26375006]
23. Grevet JD et al. Domain-focused CRISPR screen identifies HRI as a fetal hemoglobin regulator in human erythroid cells. Science 361, 285–290 (2018). [PubMed: 30026227]
24. Huang P et al. The HRI-regulated transcription factor ATF4 activates BCL11A transcription to silence fetal hemoglobin expression. Blood 135, 2121–2132 (2020). [PubMed: 32299090]
25. Qin K et al. Dual function NFI factors control fetal hemoglobin silencing in adult erythroid cells. Nat. Genet 54, 874–884 (2022). [PubMed: 35618846]
26. Corces MR et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. Nat. Methods 14, 959–962 (2017). [PubMed: 28846090]
27. Skene PJ & Henikoff S An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. eLife 6, e21856 (2017). [PubMed: 28079019]
28. Kurita R et al. Establishment of immortalized human erythroid progenitor cell lines able to produce enucleated red blood cells. PLoS ONE 8, e59890 (2013). [PubMed: 23533656]
29. Lan X et al. ZNF410 uniquely activates the NuRD component CHD4 to silence fetal hemoglobin expression. Mol. Cell 81, 239–254.e8 (2021). [PubMed: 33301730]
30. Dykes IM et al. HIC2 Is a novel dosage-dependent regulator of cardiac development located within the distal 22q11 deletion syndrome region. Circ. Res 115, 23–31 (2014). [PubMed: 24748541]
31. Vinjamur DS et al. ZNF410 represses fetal globin by singular control of CHD4. Nat. Genet 53, 719–728 (2021). [PubMed: 33859416]
32. An X et al. Global transcriptome analyses of human and murine terminal erythroid differentiation. Blood 123, 3466–3477 (2014). [PubMed: 24637361]
33. Davies JO et al. Multiplexed analysis of chromosome conformation at vastly improved sensitivity. Nat. Methods 13, 74–80 (2016). [PubMed: 26595209]
34. Peslak SA et al. HRI depletion cooperates with pharmacologic inducers to elevate fetal hemoglobin and reduce sickle cell formation. Blood Adv. 4, 4560–4572 (2020). [PubMed: 32956454]
35. Jolma A et al. DNA-binding specificities of human transcription factors. Cell 152, 327–339 (2013). [PubMed: 23332764]
36. Blobel GA, Nakajima T, Eckner R, Montminy M & Orkin SH CREB-binding protein cooperates with transcription factor GATA-1 and is required for erythroid differentiation. Proc. Natl Acad. Sci. USA 95, 2061–2066 (1998). [PubMed: 9482838]
37. Letting DL, Rakowski C, Weiss MJ & Blobel GA Formation of a tissue-specific histone acetylation pattern by the hematopoietic transcription factor GATA-1. Mol. Cell. Biol 23, 1334–1340 (2003). [PubMed: 12556492]
38. Vakoc CR et al. Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1. Mol. Cell 17, 453–462 (2005). [PubMed: 15694345]
39. Welch JJ et al. Global regulation of erythroid gene expression by transcription factor GATA-1. Blood 104, 3136–3147 (2004). [PubMed: 15297311]
40. Wu W et al. Dynamic shifts in occupancy by TAL1 are guided by GATA factors and drive large-scale reprogramming of gene expression during hematopoiesis. Genome Res. 24, 1945–1962 (2014). [PubMed: 25319994]
41. Patel A, Hashimoto H, Zhang X & Cheng X Characterization of how DNA modifications affect DNA binding by C2H2 zinc finger proteins. Methods Enzymol. 573, 387–401 (2016). [PubMed: 27372763]
42. Patel A et al. DNA conformation induces adaptable binding by tandem zinc finger proteins. Cell 173, 221–233.e12 (2018). [PubMed: 29551271]
43. Luscombe NM, Laskowski RA & Thornton JM Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. Nucleic Acids Res. 29, 2860–2874 (2001). [PubMed: 11433033]
44. Chen Y et al. DNA binding by GATA transcription factor suggests mechanisms of DNA looping and long-range gene regulation. Cell Rep. 2, 1197–1206 (2012). [PubMed: 23142663]

45. Tunyasuvunakool K et al. Highly accurate protein structure prediction for the human proteome. Nature 596, 590–596 (2021). [PubMed: 34293799]
46. Shin HY et al. Hierarchy within the mammary STAT5-driven Wap super-enhancer. Nat. Genet 48, 904–911 (2016). [PubMed: 27376239]
47. Dykes IM, van Bueren KL & Scambler PJ HIC2 regulates isoform switching during maturation of the cardiovascular system. J. Mol. Cell. Cardiol 114, 29–37 (2018). [PubMed: 29061339]
48. Crossley M & Orkin SH Phosphorylation of the erythroid transcription factor GATA-1. J. Biol. Chem 269, 16589–16596 (1994). [PubMed: 8206977]
49. Andrews NC & Faller DV A rapid micropreparation technique for extraction of DNA-binding proteins from limiting numbers of mammalian cells. Nucleic Acids Res. 19, 2499 (1991). [PubMed: 2041787]
50. Stirling DR et al. CellProfiler 4: improvements in speed, utility and usability. BMC Bioinf. 22, 433 (2021).
51. Yang Y et al. Structural basis for human ZBTB7A action at the fetal globin promoter. Cell Rep. 36, 109759 (2021). [PubMed: 34592153]
52. Otwinowski Z, Borek D, Majewski W & Minor W Multiparametric scaling of diffraction intensities. Acta Crystallogr. A 59, 228–234 (2003). [PubMed: 12714773]
53. Adams PD et al. PHENIX: building new software for automated crystallographic structure determination. Acta Crystallogr. D Biol. Crystallogr 58, 1948–1954 (2002). [PubMed: 12393927]
54. Emsley P & Cowtan K Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr 60, 2126–2132 (2004). [PubMed: 15572765]

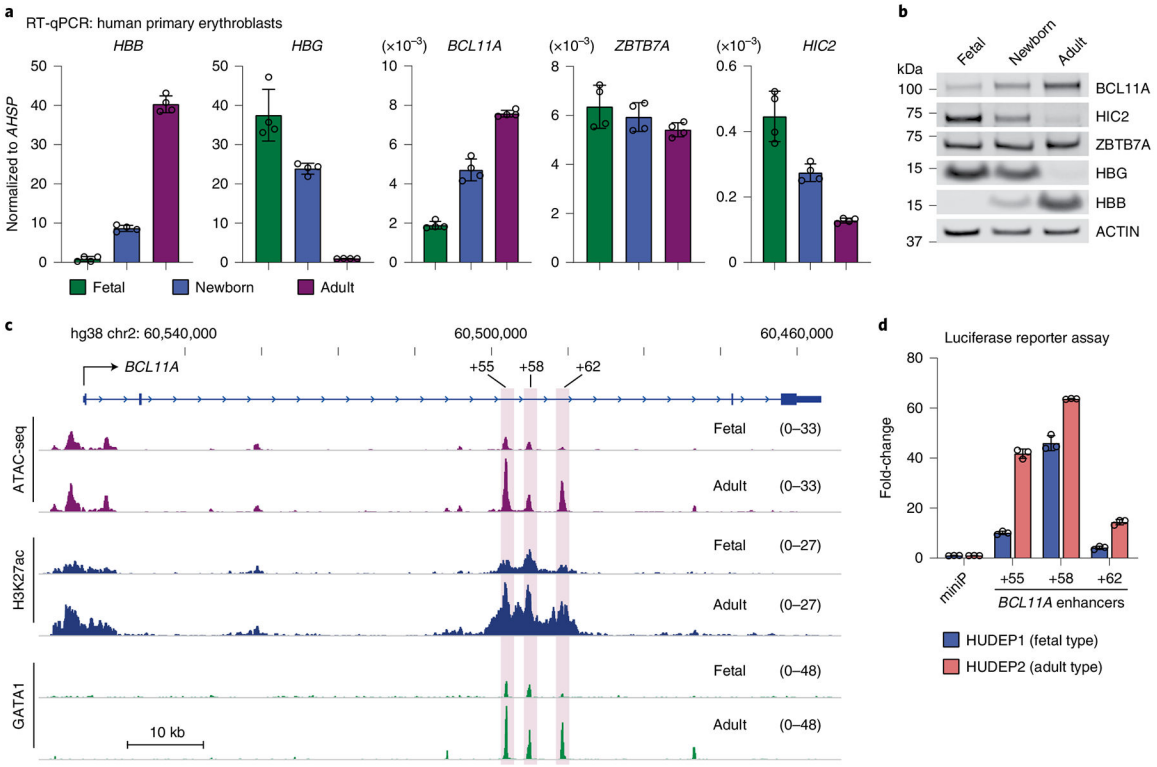



**Fig. 1 |. Developmental regulation of *BCL11A* occurs primarily at the transcriptional level.**

**a**, *HBB*, *HBG*, *BCL11A*, *ZBTB7A* and *HIC2* mRNA measured by RT–qPCR in primary human erythroblasts from three sequential developmental stages. Results are normalized to *AHSP* and shown as mean ± s.d. (two independent donors for each stage, with two technical replicates for each donor). **b**, Western blotting with indicated antibodies of extracts from primary human erythroblasts from three sequential developmental stages. Results represent two biological replicates. **c**, Representative tracks of ATAC-seq, H3K27ac and GATA1 CUT&RUN from primary human fetal and adult erythroblasts. The results were normalized to reads in peaks. The *BCL11A* erythroid enhancers +55, +58 and +62 are highlighted in pink. **d**, Luciferase reporter assay of human *BCL11A* +55, +58 and +62 enhancers in HUDEP1 and HUDEP2 cells. Results are shown as mean ± s.d. ($n = 3$). Statistical tests compare HUDEP1 and HUDEP2 samples using two-tailed Student's *t*-test; *BCL11A* +55, $P < 0.0001$; *BCL11A* +58, $P = 0.0005$; *BCL11A* +62, $P = 0.0001$.

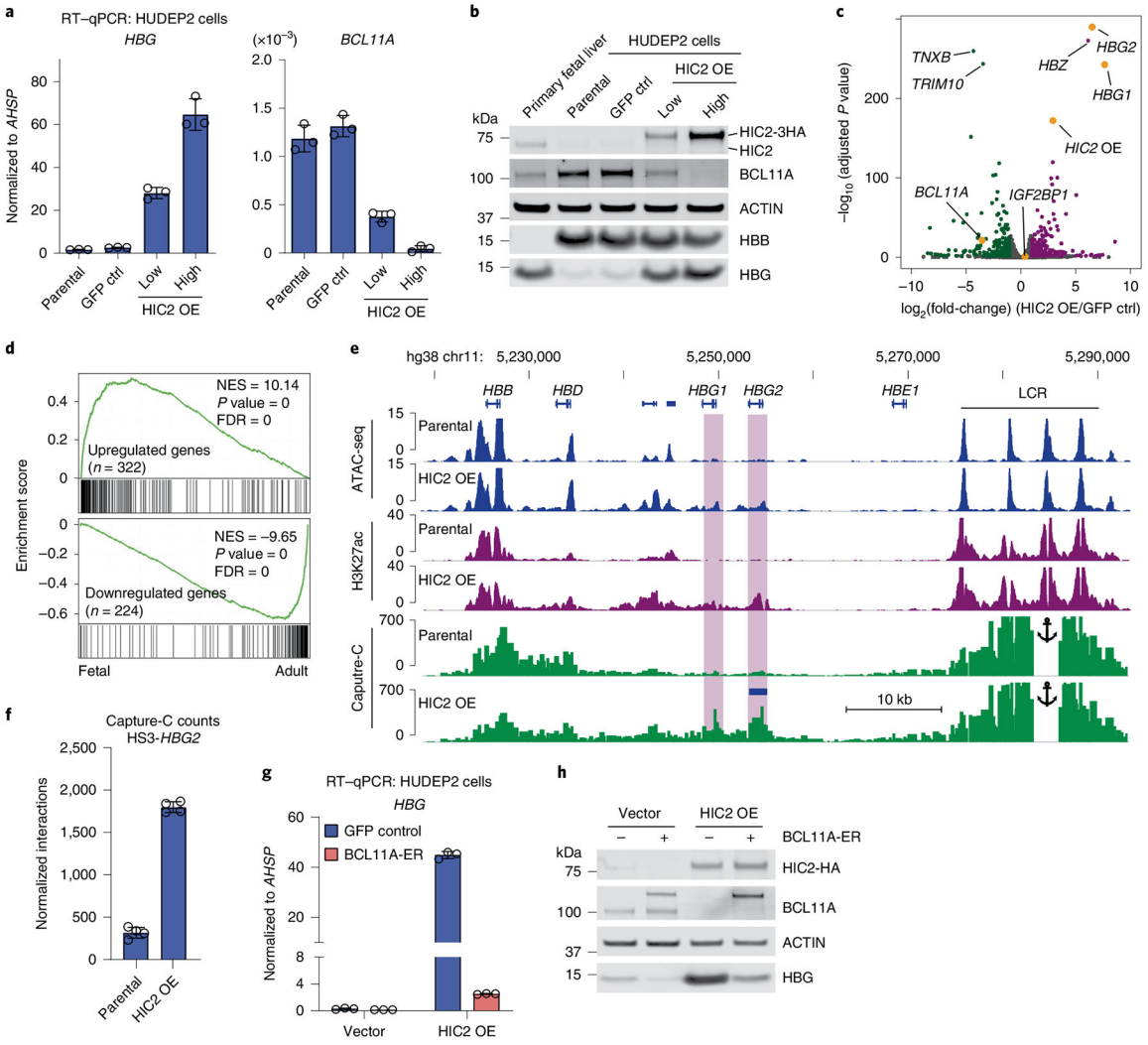



**Fig. 2 |. HIC2 is a *BCL11A* repressor and *HBG* activator.**

**a**, *HBG* (low, $P = 0.0001$; high, $P < 0.0001$) and *BCL11A* (low, $P < 0.0001$; high, $P < 0.0001$) mRNA measured by RT–qPCR in HIC2 OE HUDEP2 cells. Results are normalized to *AHSP* and shown as mean ± s.d. ($n = 3$). Statistical tests compare HIC2 OE and parental samples. *P* values were calculated by one-way ANOVA (analysis of variance). **b**, Western blot with indicated antibodies of extracts from HUDEP2 cells on HIC2 OE. The results represent three biological replicates. **c**, Volcano plot of RNA-seq results in HUDEP2 cells. Downregulated ($n = 224$) and upregulated ($n = 322$) (fold-change >2 and FDR < 0.01) genes are highlighted in green and purple, respectively. **d**, GSEA plots of differentially expressed genes in **a** compared with human fetal and adult primary erythroblasts. NES, normalized enrichment score. **e**, Representative tracks of ATAC-seq, H3K27ac ChIP–seq and Capture-C of the β-globin locus. The HS3 anchors for Capture-C are indicated by the anchor symbol. ATAC-seq and H3K27ac are normalized to reads in peaks, Capture-C is normalized to total interactions. **f**, Capture-C interaction counts of the segment shown by blue bar in **e**. Results are shown as mean ± s.d. ($n = 4$). The statistical test compares parental and HIC2 OE samples: $P < 0.0001$ by two-tailed Student's *t*-test. **g**, *HBG* mRNA measured by RT–qPCR

in HUDEP2 cells after *BCL11A*-ER and HIC2 OE. Results are normalized to *AHSP* and shown as mean ± s.d. ($n = 3$). Cells were treated with 0.1 μM tamoxifen. Statistical tests compare GFP control and BCL11A-ER samples. Vector, $P = 0.0438$; HIC2 OE, $P < 0.0001$ by two-tailed Student's *t*-test. **h**, Western blot analysis of HUDEP2 cells on BCL11A-ER and HIC2 OE. Cells were treated with 0.1 μM tamoxifen. Results represent two biological replicates.

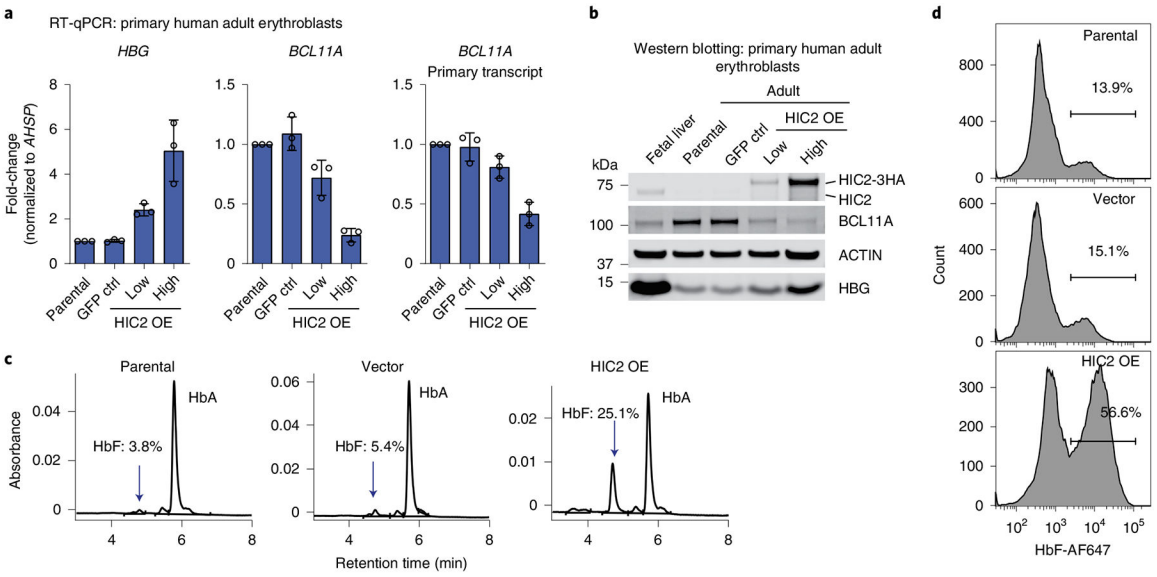



**Fig. 3 |. HIC2 OE represses *BCL11A* transcription and augments fetal globin production in primary erythroid cells.**

**a**, *HBG* (low, $P = 0.0938$; high, $P = 0.0003$), *BCL11A* mature (low, $P = 0.0290$; high, $P < 0.0001$) and primary (low, $P = 0.0762$; high, $P = 0.0001$) transcripts measured by RT–qPCR in primary human erythroblasts. Results are normalized to *AHSP* and shown as mean ± s.d. ($n = 3$). Statistical tests compare HIC2 OE and parental samples. *P* values were calculated using a one-way ANOVA test. **b**, Western blot with indicated antibodies of extracts from primary human erythroblasts on HIC2 OE. Results represent three biological replicates. **c**, Representative HPLC analysis of hemoglobin in primary human erythroblasts. HbF peaks are indicated by blue arrows and the HbF peak area is displayed as the percentage of total HbF + HbA. **d**, Representative flow cytometry of HbF-stained primary human erythroblasts.

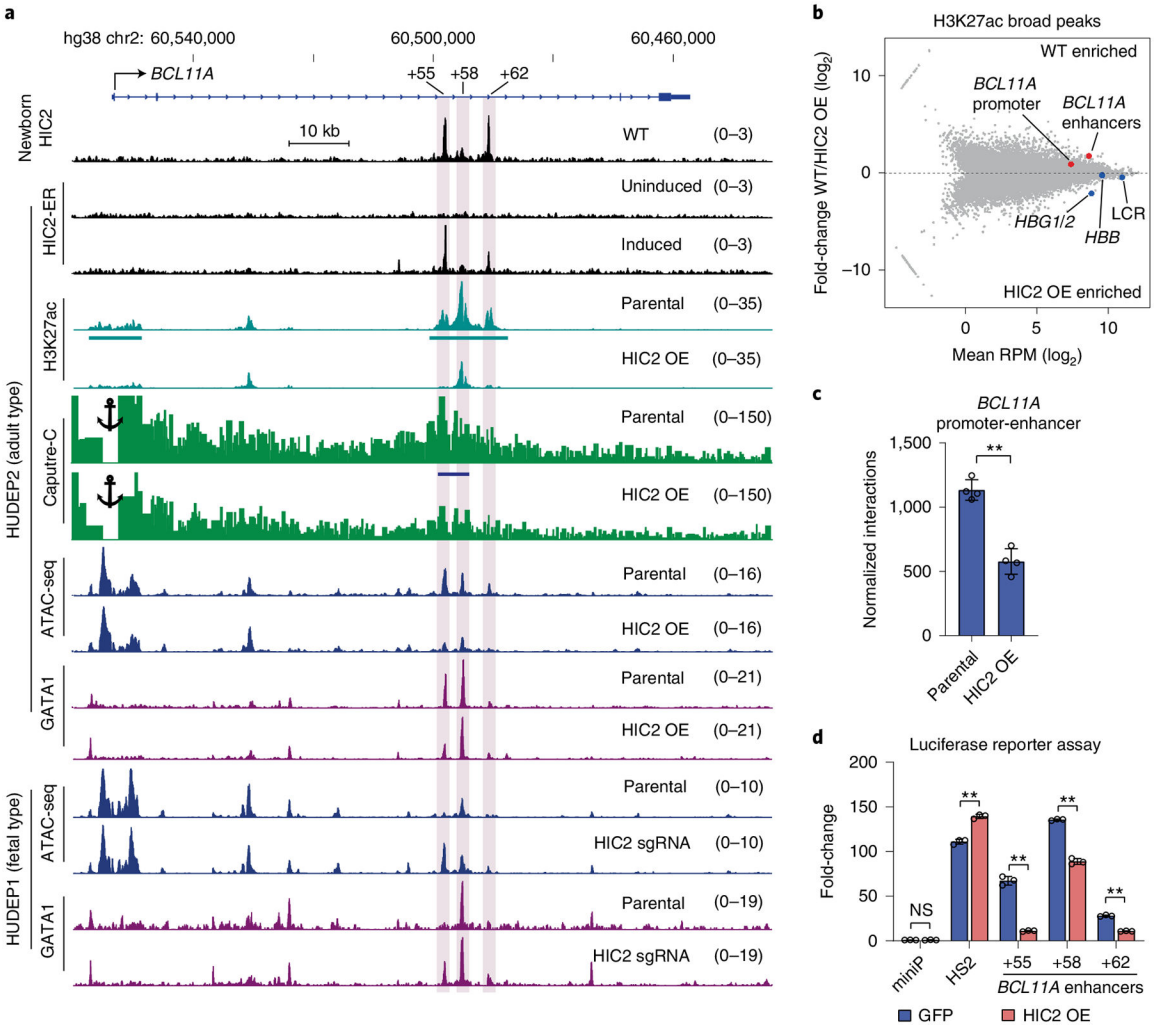



**Fig. 4 |. HIC2 controls *BCL11A* transcription through erythroid enhancers.**

**a**, Representative tracks of endogenous HIC2, HIC2-ER, H3K27ac, Capture-C, ATAC-seq and GATA1 ChIP–seq in primary newborn erythroblasts, HUDEP2 and HUDEP1 cells on HIC2 perturbations. The *BCL11A* enhancers are highlighted in purple. The *BCL11A* promoter viewpoints for Capture-C are denoted by anchor symbols. The results are normalized to reads in peaks, except for HIC2 and HIC2-ER ChIP–seq and Capture-C, which are normalized to total reads. **b**, MA plot of H3K27ac ChIP–seq read counts of parental and HIC2 OE HUDEP2 cells. *BCL11A* promoter and enhancers are highlighted in red, and the regions quantified are shown by cyan bars in **a**. *HBG1/2*, *HBB* and LCR are highlighted in blue. RPM, reads per million. **c**, Capture-C interaction counts of the segment shown by blue bar in **a**. Results are shown as mean ± s.d. ($n = 4$). Statistical test compares parental and HIC2 OE samples: $P = 0.0001$ by two-tailed Student's *t*-test. **d**, Luciferase reporter assay of enhancer activities in HUDEP2 cells. The HS2 from β-globin LCR serves as the positive control. Results are shown as mean ± s.d. ($n = 3$). Statistical tests compare GFP and HIC2 OE samples using two-tailed Student's *t*-test: HS2, $P = 0.0002$; *BCL11A* +55, $P < 0.0001$; *BCL11A* +58, $P < 0.0001$; *BCL11A* +62, $P < 0.0001$. NS, not significant.

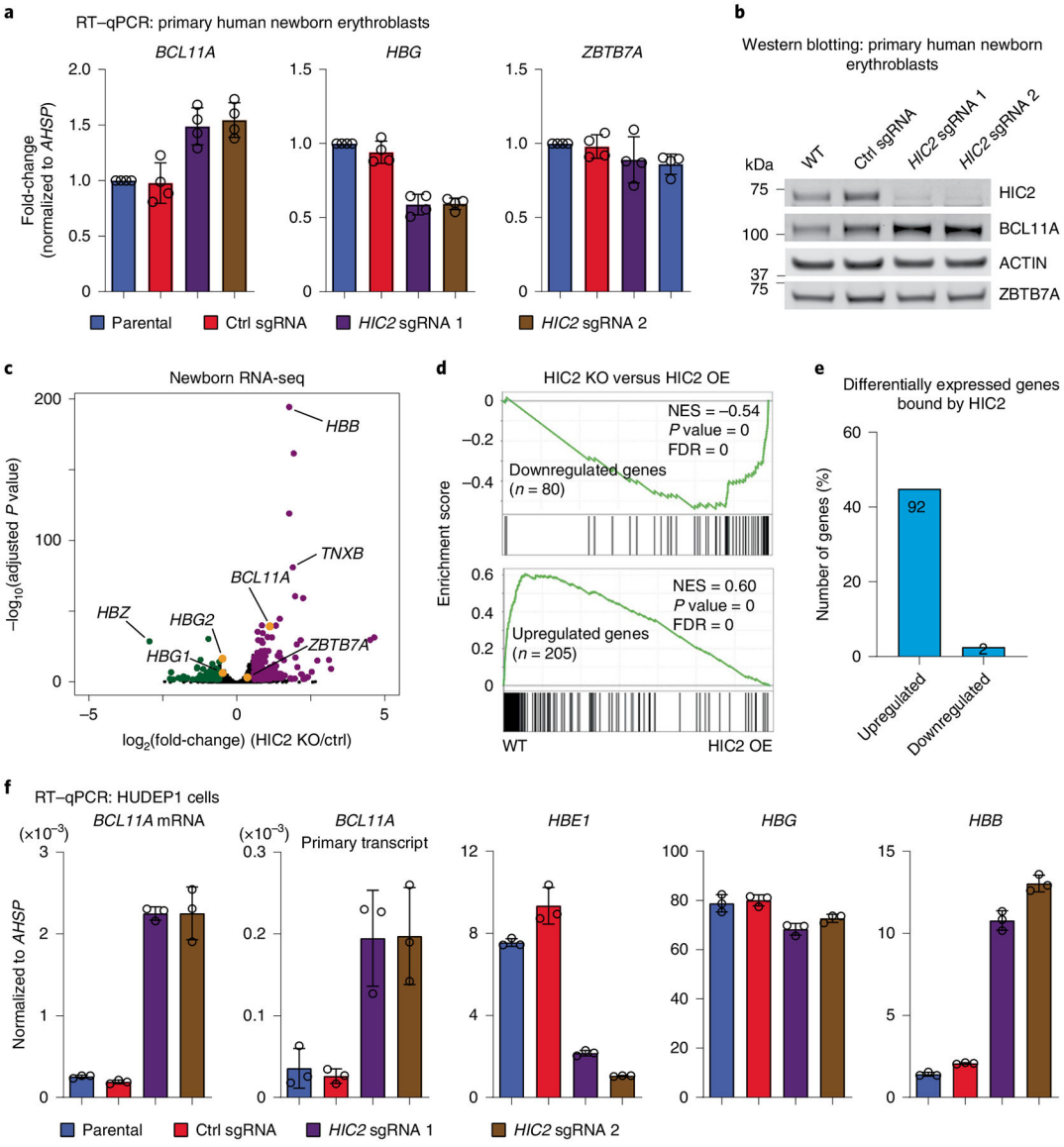



**Fig. 5 |. HIC2 controls *HBG* transcription via repression of *BCL11A*.**

**a**, *BCL11A* (sgRNA 1, $P = 0.0014$; sgRNA 2, $P = 0.0006$), *HBG* (sgRNA 1, $P < 0.0001$; sgRNA 2, $P < 0.0001$) and *ZBTB7A* (sgRNA 1, $P = 0.2708$; sgRNA 2, $P = 0.1344$) measured by RT–qPCR in primary human newborn erythroblasts. Results are normalized to *AHSP* and parental samples and shown as mean ± s.d. (two independent donors with two technical replicates for each donor). Statistical tests compare HIC2 sgRNAs and parental samples. *P* values were calculated using one-way ANOVA. **b**, Western blot with indicated antibodies of extracts from primary human newborn erythroblasts on HIC2 depletion. Results represent two biological replicates. **c**, Volcano plot of RNA-seq results in primary newborn erythroblasts. Downregulated ($n = 80$) and upregulated ($n = 205$) (fold-change >1.5 and FDR < 0.01) genes are highlighted in green and purple, respectively. KO, knockout. **d**, GSEA plots of differentially expressed genes in **c** compared with RNA-seq results of HIC2 OE HUDEP2 cells. NES, normalized enrichment score. **e**, Percentage of differentially

expressed genes in **c** bound by HIC2. HIC2-bound genes are defined by the HIC2 peaks located in gene exons/introns or <1 kb to transcription start/termination sites. **f**, *BCL11A* mature mRNA (sgRNA 1, $P < 0.0001$; sgRNA 2, $P < 0.0001$) and primary transcripts (sgRNA 1, $P = 0.0054$; sgRNA 2, $P = 0.0049$), *HBE1* (sgRNA 1, $P < 0.0001$; sgRNA 2, $P < 0.0001$), *HBG* (sgRNA 1, $P = 0.0025$; sgRNA 2, $P = 0.0460$) and *HBB* (sgRNA 1, $P < 0.0001$; sgRNA 2, $P < 0.0001$) measured by RT–qPCR in HUDEP1 cells on HIC2 depletion. Results are normalized to *AHSP* and shown as mean ± s.d. ($n = 3$). Statistical tests compare HIC2 sgRNAs and parental samples. *P* values were calculated using one-way ANOVA.

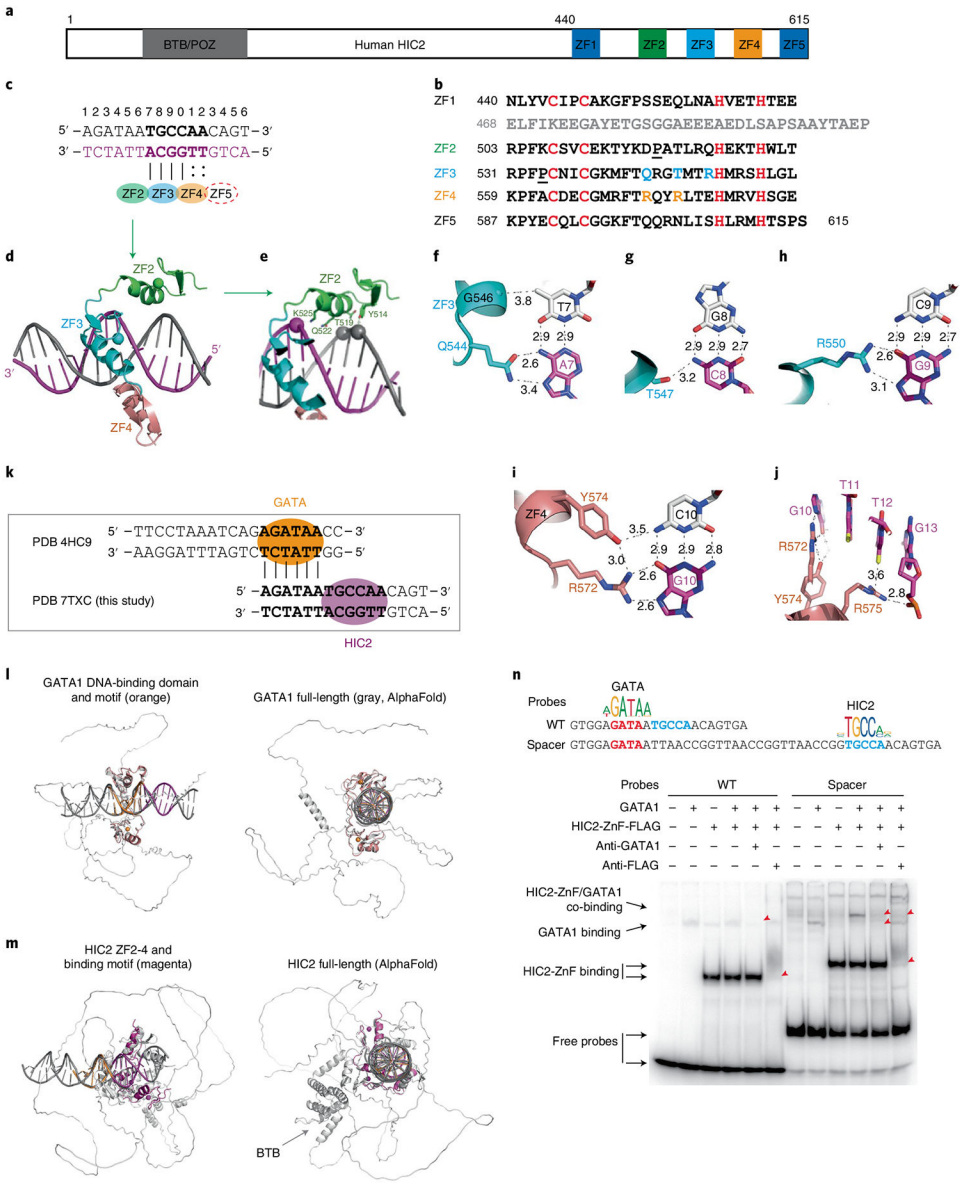



**Fig. 6 |. HIC2 and GATA1 compete for binding at the *BCL11A* +55 enhancer.**

**a**, Schematic of HIC2 full-length protein. **b**, Sequence alignment of the five ZFs of HIC2. Note the long insertion between ZF1 and ZF2 enriched with negatively charged amino acids. **c**, Oligonucleotide used for crystallization and positions of each basepair are indicated above the sequence. **d**, An overview of HIC2 ZF2–ZF4 binding to DNA. **e**, ZF2 interacting with DNA phosphate backbone of both strands across the minor groove. **f**, Gln544 and Gly546 of ZF3 interacting with T:A basepair at position 7. **g**, Thr547 of ZF3 interacting with C8. **h**, R550 of ZF3 interacts with G9. **i**, Arg572 and Tyr574 of ZF4 interacting with C:G basepair at position 10. **j**, Arg575 of ZF4 spanning three basepairs and interacting with Thy12 and Gua13. **k**, Common six basepairs of DNA sequences in the GATA DNA-binding domain (PDB accession no. 4HC9) and HIC2 ZF domain (the present study). **l**, Two orthogonal views of the superimposition of GATA DNA-binding domain on to the AlphaFold GATA1 full-length protein. **m**, Superimposition of HIC2 ZF domain on to the AlphaFold HIC2 full-

length protein. **n**, EMSA of competition binding of GATA1 and HIC2 at the *BCL11A* +55 enhancer. Red arrowheads indicate the super-shifted bands. Results represent two biological replicates.