medRχiv
THE PREPRINT SERVER FOR HEALTH SCIENCES

Human genetics implicates a BACH2-NRF2 axis in fetal hemoglobin activation

Chun-Jie Guo[1,2,3,4,*], Uma P. Arora[1,2,3,4,*], Wanying Xu[1,2,3,4], Xiaoheng Cheng[1,2,3,4], Liam D. Cato[1,2,3,4], Rick Li[1,2,3,4], Henry Y. Lu[1,2,3,4], Andrew J. Lee[1,2,3,4], Fulong Yu[1,2,3,4], Gaurav Agarwal[1,2,3,4], Peng Lyu[1,2,3,4], Tianyi Ye[1,2,3,4], Mateusz Antoszewski[1,2,3,4], Mariel Wissmann[1,2,3,4], Baraka S. Mkumbe[5,6,7], Supachai Ekwattanakit[8], Patrick Deelen[9,10], Liberata Mwita[11], Raphael Sangeda[5,11], Thidarat Suksangpleng[8], Suchada Riolueang[8], Paola G. Bronson[12], Dirk S. Paul[13,14], Emily Kawabata[13], William J. Astle[13,15,16,17], Francois Aguet[4], Kristin Ardlie[4], Aitzkoa Lopez de Lapuente Portilla[18,19], Guolian Kang[20], Yingze Zhang[21], Seyed Mehdi Nouraie[21], Victor R. Gordeuk[22], Mark T. Gladwin[23], Melanie E. Garrett[24], Allison Ashley-Koch[24], Marilyn J. Telen[24], Brian Custer[25,26], Shannon Kelly[25,27], Carla Luana Dinardo[28,29], Ester C. Sabino[29], Paula Loureiro[30], Anna Bárbara Carneiro-Proietti[31], Cláudia Maximo[32], NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium[33], Adriana Méndez[34], Angelika Hammerer-Lercher[34], Julie Makani[5,6], Vivien A. Sheehan[35], Mitchell J. Weiss[20], Lude Franke[10,36], Björn Nilsson[4,18,19], Adam S. Butterworth[13,14,15,37,38], Vip Viprakasit[8,39], Siana Nkya[5,6,40], Vijay G. Sankaran[1,2,3,4,41,†]

[1]Division of Hematology/Oncology, Boston Children's Hospital, Harvard Medical School, Boston, Massachusetts, USA, [2]Department of Pediatric Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts, USA, [3]Howard Hughes Medical Institute, Boston, MA 02115, USA,[4]Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA, [5]Sickle Cell Program, Department of Hematology and Blood Transfusion, Muhimbili University of Health and Allied Sciences, Dar es Salaam, Tanzania, [6]Department of Biochemistry, Muhimbili University of Health and Allied Science, Dar es Salaam, Tanzania, [7]Department of Artificial Intelligence and Innovative Medicine, Graduate School of Medicine, Tohoku University, Sendai, Japan, [8]Siriraj Thalassemia Center, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand, [9]Department of Genetics, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands, [10]Oncode Institute, Amsterdam, the Netherlands, [11]Department of Pharmaceutical Microbiology, Muhimbili University of Health and Allied Sciences, Dar es Salaam, Tanzania, [12]R&D Translational Biology, Biogen, Cambridge, Massachusetts, USA, [13]British Heart Foundation Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK, [14]British Heart Foundation Centre of Research Excellence, University of Cambridge, Cambridge, UK, [15]National Institute for Health and Care Research Blood and Transplant Research Unit in Donor Health and Behaviour, University of Cambridge, Cambridge, UK, [16]MRC Biostatistics Unit, University of Cambridge, Cambridge, UK, [17]NHS Blood and Transplant, Cambridge, UK, [18]Lund Stem Cell Center, Lund University, 221 84 Lund, Sweden, [19]Department of Laboratory Medicine, Lund University, 221 84 Lund, Sweden, [20]St. Jude Children's Research Hospital, Memphis, Tennessee, USA, [21]Department of Medicine, School of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania, USA, [22]Division of Hematology and Oncology, Department of Medicine, Comprehensive Sickle Cell Center, University of Illinois at Chicago, Chicago, Illinois, USA, [23]Department of Medicine, University of Maryland School of Medicine, Baltimore, Maryland, USA, [24]Department of Medicine, Duke University Medical Center, Durham, North Carolina, USA, [25]Vitalant Research Institute, San Francisco, California, USA, [26]Department of Laboratory Medicine, UCSF, San Francisco, California, USA, [27]Division of Pediatric Hematology, UCSF Benioff Children's Hospital, Oakland, California, USA, [28]Fundacao Pro-Sangue Hemocentro de Sao Paulo, Sao Paulo, Brazil, [29]Institute of Tropical Medicine, Faculdade de Medicina da Universidade de Sao Paulo, Sao Paulo, Brazil, [30]Fundacao Hemope, Recife, Pernambuco, Brazil, [31]Fundacao Hemominas, Belo Horizonte, Brazil, [32]Fundacao Hemorio, Rio de Janeiro, Brazil, [33]The TOPMed Consortium is detailed in Supplemental Acknowledgments, [34]Institute of Laboratory Medicine, Cantonal Hospital Aarau, 5000 Aarau, Switzerland, [35]Aflac Cancer & Blood Disorders Center, Children's Healthcare of Atlanta & Department of Pediatrics, Emory University School of Medicine, Atlanta, Georgia, USA, [36]Department of Genetics, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands, [37]Health Data Research UK Cambridge, Wellcome Genome Campus and University of Cambridge, Cambridge, UK, [38]Heart and Lung Research Institute, University of Cambridge, Cambridge, UK, [39]Department of Pediatrics, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand, [40]Tanzania Human Genetics Organisation, Tanzania, [41]Harvard Stem Cell Institute, Cambridge, Massachusetts, USA
*These authors contributed equally to this work
†Corresponding author, sankaran@broadinstitute.org

Human genetic studies have identified key regulators of fetal hemoglobin (HbF) expression, including BCL11A, resulting in therapeutic advances. Yet the mechanisms by which HbF expression is activated remain incompletely understood. Here, we conduct a multi-ancestry genome-wide association study of HbF levels in 28,237 individuals that identifies 131 conditionally-independent associations across 12 genomic regions. In one previously uncharacterized associated region, the high-HbF-linked causal variant, rs1010474-C, reduces BACH2 expression and elevates HbF levels. Direct perturbation or inhibition of BACH2 elevates HbF expression. Mechanistically, BACH2 restrains activation of the HbF-encoding γ-globin genes, while loss of BACH2 enhances NRF2 chromatin occupancy and promotes activation foci formation at the γ-globin genes. BACH2 and

NRF2 binding motifs in the γ-globin promoters, while overlapping, can be selectively edited to activate or repress γ-globin, respectively, and function independently of BCL11A. These findings illustrate how human genetic variation continues to advance our knowledge of therapeutically-relevant regulatory mechanisms underlying HbF expression.

Introduction

During human development, hemoglobin transitions shortly after birth from the fetal form (HbF), which contains γ-globin, a β-like globin encoded by the *HBG1/2* genes, to the adult form (HbA), which contains β-globin encoded by the *HBB* gene - a process referred to as the fetal-to-adult hemoglobin switch[1]. Persistently increased production of HbF after infancy can ameliorate clinical symptoms in common and life-threatening disorders arising from *HBB* mutations, including sickle cell disease and β-thalassemia[2,3]. While the ameliorating effects of HbF in these hemoglobin disorders have been known for decades, the mechanisms regulating HbF and strategies to target them have only recently been elucidated. Initial genome-wide association studies (GWAS) revealed three regions of association with HbF levels - the *HBB* locus on chromosome 11, the *HBS1L-MYB* locus on chromosome 6, and the *BCL11A* locus on chromosome 2[4–6]. These studies prompted functional follow-up that revealed BCL11A as a key and direct repressor of *HBG1/2* transcription[7]. Subsequently, additional insights have emerged from the study of rare loss-of-function mutations impacting *BCL11A*[8–10]; transcriptional regulatory elements necessary for erythroid expression of *BCL11A*[11,12]; upstream regulators of *BCL11A* expression[13–16]; and analysis of the mechanisms by which *BCL11A* alters transcription[17–19]. Suppression of *BCL11A* or its erythroid-specific enhancer using genome editing or gene therapy approaches has emerged as an effective treatment strategy for sickle cell disease and β-thalassemia[20,21]. In addition, functional studies in cellular models of other HbF regulators have provided new insights[22–26], such as the identification of ZBTB7A as another direct repressor of the γ-globin genes[19,27].

Despite substantial progress in understanding HbF regulation, largely derived from studies in cell lines and mouse models, many fundamental questions about how HbF is regulated in humans in vivo remain unanswered. The majority of genetically-validated regulators of HbF expression that have been identified to date are repressors. However, the precise mechanisms by which the γ-globin genes get activated remain incompletely understood. Despite the identification of new putative regulators of HbF expression, many bona fide regulators likely remain undefined.

To address these existing limitations, we conducted the largest multi-ancestry GWAS of HbF levels to date, involving 28,237 participants from a range of global populations. Through this study, we identified BACH2 as a previously unrecognized, genetically-nominated regulator of HbF expression. The high-HbF-associated causal variant, rs1010474-C, reduces BACH2 expression. Direct perturbation or pharmacological inhibition of BACH2 also elevates HbF levels. Mechanistically, BACH2 prevents activation of the γ-globin genes via NRF2, while loss of BACH2 enhances NRF2 chromatin occupancy and promotes NRF2 activation foci formation at the γ-globin genes. Direct base editing of the BACH2 binding motif shifts cis-element affinity toward NRF2, resulting in elevated HbF expression, whereas disruption of NRF2 binding at the promoters prevents HbF activation. Together, these findings provide new insights into the mechanisms underlying HbF activation and pave the way for future therapeutic strategies inspired by human genetic variation.

Results

HbF GWAS meta-analysis

To gain insights into the regulation of HbF expression, we collected cohorts from around the world and conducted a large-scale multi-ancestry GWAS meta-analysis, which included 28,237 individuals from 11 distinct cohorts spanning African, Asian, and European ancestries (**Figure 1a and Supplementary Table 1**). In the GWAS meta-analysis we observed no inflation in test statistics (genomic inflation factor (λ_{gc})=0.99, linkage disequilibrium score regression (LDSC) intercept=0.98), indicating that the included covariates sufficiently captured ancestry differences across cohorts (**Figure S1a-d**). We identified 131 conditionally independent signals associated with HbF levels across 12 windows (9 windows at the genome-wide significance threshold, $P < 5 \times 10^{-8}$, and 3 windows at the suggestive threshold, $P < 1 \times 10^{-6}$) (**Figure 1b and Supplementary Tables 2, 3, 4**). We annotated these windows with genes that satisfied a combination of criteria - (i) distance from a significantly associated variant, (ii) long-range interaction data linking regulatory elements to genes in erythroid cells via promoter capture Hi-C (PCHi-C), (iii) correlations between gene expression and chromatin accessibility (RNA- and ATAC-seq correlations) in hematopoietic cells, and (iv) expression quantitative trait loci (eQTLs) from whole blood (**Supplementary Table 5**). In addition to the previously characterized regions, we identified regions near known HbF regulators that had not been previously identified by other population-based genetic studies, including *ZBTB7A* and *KLF1*, as well as other regions that did not harbor genes previously implicated in HbF regulation.

We next conducted ancestry-specific analyses (**Figure 1c-e, Supplementary Table 6**), including for individuals with African (AFR, n=3,963), European (EUR, n=22,882), and Thai (n=1,392) ancestry and found largely conserved association windows at the major loci identified, including at the *BCL11A*, *HBS1L-MYB*, *HBB*, and *CTC1* loci. Windows nominating *BACH2* were found to be significantly associated with HbF in the AFR and EUR populations, while *PSME4* and *ABCC1* were found in the EUR populations alone. For SNPs present across all populations, effect sizes from the multi-ancestry GWAS

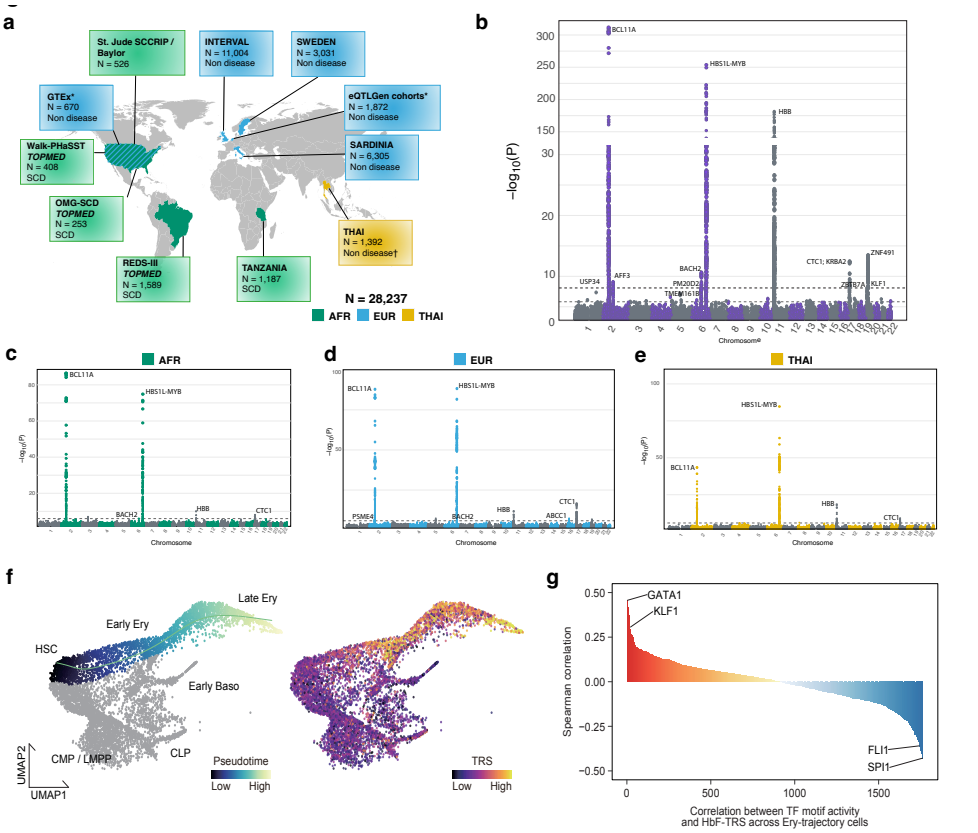

Figure 1: GWAS of HbF across global populations with enrichment of fine-mapped variants in erythroid cells. (a) Population geography of included studies and associated sample numbers. †, the Thai population had a proportion of individuals selected for elevated HbF; *, select cohorts were included from the eQTLGen consortium (see Methods). **(b)** Combined meta-analysis of fetal hemoglobin identifies several unexplored loci. Gene symbols shown are the most likely impacted gene nominated using several approaches. Window boxes are drawn over significant and suggestive signals identified via conditional analysis (see methods). **(c-e)** Ancestry-specific analyses using multi-ancestry meta-analysis (MAMA; see Methods) for African (AFR), European (EUR), and THAI ancestry backgrounds, respectively. Areas of differential signal indicate potential ancestry-specific effects on HbF, y-axis was limited to $P>1 \times 10^{-100}$. **(f)** SCAVENGE analysis using scATAC-seq data of hematopoietic differentiation, showing strong enrichment of the trait relevance score (TRS) in the mid-late erythroid population. **(g)** Spearman correlations between SCAVENGE TRS and chromVAR TF motif enrichment scores across erythroid trajectory cells (y-axis) ranked in order of highest to lowest correlation along the x-axis to identify co-regulated transcription factor motifs.

meta-analysis (MAMA) were highly correlated (R = 0.61–0.74, **Figure S1e-g**), supporting our assumption of fixed effect in the joint meta-analysis. Some of the observed ancestral heterogeneity might arise from limited power as a consequence of small cohort size and/or reduced genetic variation within specific populations.

For the whole-cohort analysis, the linkage disequilibrium adjusted kinships (LDAK) model estimated heritability to be 0.143 (SD 0.016). When we investigated SNP heritability by ancestry group, the EUR population revealed a heritability of 0.20 (SD 0.038), AFR at 0.33 (SD 0.22), and the Thai population at 0.40 (SD 0.32). The higher trait heritability in the AFR populations could be attributed to the fact that the AFR population cohorts used in this study were exclusively individuals with sickle cell disease, leading to a higher heritability for HbF, as suggested previously[28]. The Thai population was also enriched for individuals with very high HbF levels, which may bias these estimates. Upon analysis of heritability enrichments for different histone modifications or genomic regions, the major enrichments were seen in putative enhancer elements, suggesting that SNPs residing in and potentially altering regulatory elements contribute most to the currently observed heritability in HbF levels (**Figure S2 a**).

Given the well-powered insights from genetic analysis of blood cell phenotypes across populations[29,30], we examined the extent to which genetic variation impacting HbF levels might also have genetic overlap with these phenotypes (**Figure S2 b and Supplementary Table 7**). We observed positive genetic correlations between HbF levels and multiple white blood cell phenotypes, including white blood cell (WBC, p=0.007) basophil (BAS, p=0.015), lymphocyte (LYM, p=0.019), monocyte (MON, p=0.026), and neutrophil (NEU, p=0.007) counts. Positive correlation trends were also observed for red cell and platelet-related traits, although these did not reach statistical significance.

Cellular contexts of HbF-associated variants

We next sought to gain insights into the cellular contexts underlying HbF-associated variation. We employed the Single Cell Analysis of Variant Enrichment through Network propagation of GEnomic data (SCAVENGE)[31] approach to identify relevant cell states where the fine-mapped variants showed significant co-localization with accessible chromatin across human hematopoiesis. The cell type enrichment is represented by a trait relevance score (TRS), a scaled and normalized probability distribution representing a quantitative metric for the relevance of a cell for a phenotype, while accounting for cell-to-cell similarity using network propagation. We fine-mapped variants in each locus window to identify a credible set of potential causal variants. Half of the locus windows had 95% credible sets containing 10 or fewer variants (**Supplementary Table 8**), which were primarily localized to introns (**Figure S2d**). We identified a strong enrichment in erythroid cells compared to other hematopoietic cell types (**Figure S2 e-f**). At single-cell resolution, using a pseudotime projection of human erythropoiesis[32], we found a strong enrichment at the mid-

maturation stages of erythropoiesis, peaking around the proerythroblast to basophilic erythroblast stages (**Figure 1f**). Given these enrichments, we sought to define co-regulated transcription factor (TF) motifs. Spearman correlations between the SCAVENGE TRS and chromVAR[33] TF motif enrichment scores across erythroid cells were calculated (**Figure 1g, Supplementary Table 9**). Notably, KLF1 and GATA1 motifs were highlighted and both of these transcription factors are critical in HbF regulation[34]. Collectively, these results highlight key differentiation stages and transcriptional regulatory networks involved in HbF-associated genetic variation.

High-HbF variant rs1010474-C reduces BACH2 expression

Having shown globally that much of the genetic variation impacting HbF levels mapped to intermediate stages of human erythropoiesis and to transcriptional regulatory elements, we next sought to identify previously unappreciated regulators of this process from the associations. Of the regions identified from the GWAS meta-analysis, a notable region contained a lead variant (rs2325259) tagging a haplotype within the *BACH2* gene. By fine-mapping, we identified two candidate causal variants (rs1010473 and rs1010474) in tight linkage disequilibrium with the lead variant (D`>0.98, R^2>0.97 in both EUR and AFR populations), which overlapped a region of accessible chromatin in human hematopoietic stem and progenitor cells (HSPCs) and erythroid progenitors (**Figure 2a and Figure S3a**). To study how these naturally-occurring variants influence regulatory activity, we generated a lentiviral reporter containing the putative causal variants in this regulatory element positioned upstream of a minimal promoter. We demonstrated that this regulatory element functions as an enhancer during human erythroid differentiation and the high-HbF-associated allele (C) at rs1010474 reduces enhancer activity (**Figure 2b and Figure S3b**).

To decipher the role of this regulatory element in an endogenous setting, we performed CRISPR/Cas9-mediated deletion with two pairs of guide RNAs (ENH del1 and ENH del2; **Figure S3c, d**). We achieved ~40% deletion at all alleles and observed resultant reductions in *BACH2* mRNA and protein levels (**Figure S3e-g**), as well as an induction of HbF, without any disruption of erythroid differentiation (**Figure S3h-j**). Notably, no other genes in the topologically associating domain containing this regulatory element were impacted (**Figure S3k**). Next, to examine the impact of allelic variation

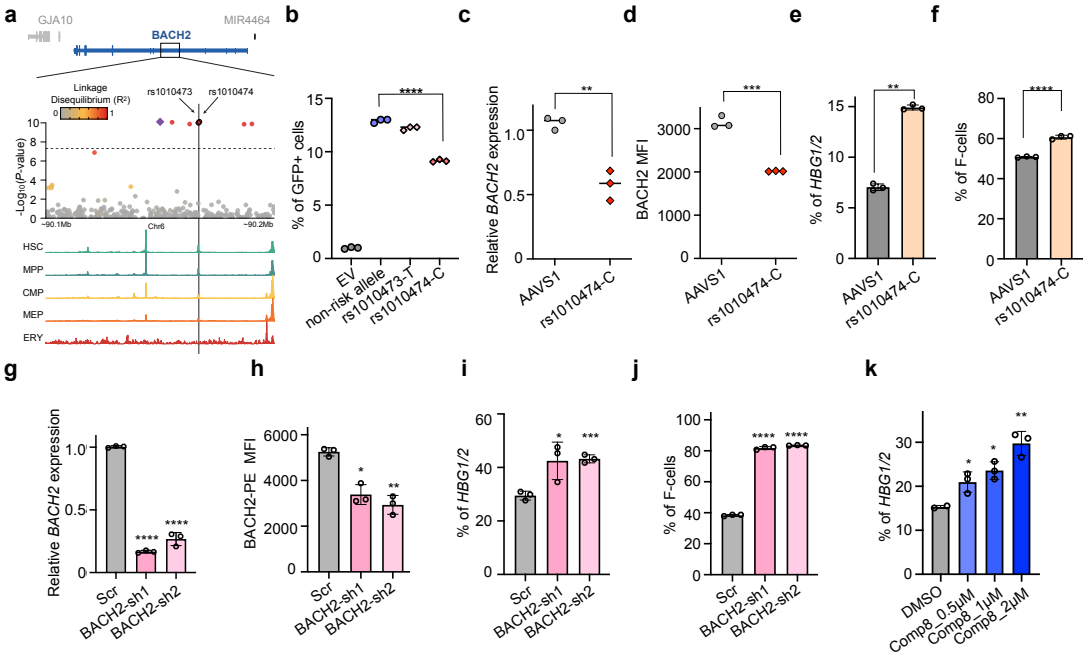

Figure 2: Genetic and functional dissection of BACH2 as a regulator of HbF. (a) The *BACH2* locus, with sentinel variant rs* shown as a purple diamond, and LD R^2 colored from red (high) to yellow (low). Two variants rs1010473 and rs1010474 (highlight in black line) display high LD (both in EUR and AFR populations) with the sentinel and are positioned in a peak of accessible chromatin in HSC cells, tracks of bulk ATAC-seq for erythroid relevant trajectories are shown below. **(b)** *BACH2* putative causal variants were found to reduce enhancer activity in lentiviral reporter assays. Percentage of GFP positive cells were measured by flow cytometry. **(c)** Relative *BACH2* expression in bulk human primary CD34⁺ HSPCs three days after editing of the high-HbF linked variant rs1010474-C by adenine base editor (ABE), compared to *AAVS1* control. **(d)** Quantification of intracellular BACH2 Mean Fluorescent Intensity (MFI) by flow cytometry in bulk human primary CD34⁺ HSPCs three days after *AAVS1* editing or BACH2 rs1010474-C editing using ABE. **(e)** Proportion of *HBG1/2* expression on day 13 of erythroid differentiation in *AAVS1* and rs1010474-C edited HSPCs, which was calculated by compared to overall *HBG1/2* + *HBB* expression here and throughout this study. **(f)** Frequency of F-cells on day 13 of erythroid differentiation in *AAVS1* and rs1010474-C edited HSPCs. **(g)** Relative *BACH2* expression in bulk human primary CD34⁺ HSPCs four days after two individual BACH2 shRNA transduced, compared to Scr (scramble) control. **(h)** Quantification of intracellular BACH2 MFI by flow cytometry in bulk human primary CD34⁺ HSPCs four days after shRNA transduced. **(i)** Proportion of *HBG1/2* expression on day 13 of erythroid differentiation in BACH2 shRNA and Scr control transduced HSPCs. **(j)** Frequency of F-cells on day 13 of erythroid differentiation in BACH2 shRNA and Scr control transduced HSPCs. **(k)** Proportion of *HBG1/2* expression on day 13 of erythroid differentiation in different concentrates (0.5-2 μM) of BACH1/2 inhibitor Compound 8 (Comp8) treated HSPCs. All data are presented as the mean ± the standard deviation; significance is indicated as $*p < 0.05$, $**p<0.01$, $***p < 0.001$, $****p < 0.0001$.

at this regulatory element, we used an adenine base editor (ABE8E) to introduce an T>C conversion at rs1010474 with ~40% of editing efficiency (**Figure S3l**). We observed consequently reduced levels of *BACH2* mRNA and protein with this single nucleotide change (**Figure 2c, d**), accompanied by an induction of HbF (**Figure 2e, f**), and without any obvious perturbation of erythropoiesis (**Figure S3m**). These findings demonstrate that the high-HbF allele, rs1010474-C, is likely causal and reduces BACH2 expression within erythroid progenitors, thereby altering HbF levels.

BACH2 represses fetal hemoglobin activation

Having demonstrated that altered expression of BACH2 appears to be the key functional target of the high-HbF-associated variant rs1010474-C, we next sought to define the mechanisms by which BACH2 might play a role in globin gene regulation. While prior work had implicated the related transcription factor, BACH1, in globin gene regulation[35,36], no prior work has assessed a role for BACH2 in this process. To examine the role of BACH2 in fetal hemoglobin regulation, we first employed shRNA knockdown to reduce *BACH2* mRNA by up to 80% and protein by up to 50% in primary human HSPCs and throughout erythroid differentiation (**Figure 2g, h and Figure S4a-c**). Notably, knockdown of BACH2 resulted in very minor alterations in erythroid differentiation, but induced HbF levels during erythroid differentiation, as assessed by both measurement of *HBG1/2* mRNA levels and the production of HbF-containing erythroid cells (F-cells; **Figure 2i, j and Figure S4d-h**). We next employed different doses of a BACH1/2[37] inhibitor, compound 8, during erythroid differentiation and observed very minor impacts on erythroid differentiation (**Figure S4i, j**). Similar to shRNA-mediated BACH2 knockdown, we observed a dose-dependent induction of HbF (**Figure 2k and Figure S4k**), further reinforcing the role of BACH2 in HbF regulation.

Finally, we sought to reciprocally increase BACH2 expression and assess its role in globin gene expression. We achieved different levels of increased BACH2 expression by sorting the top (BACH2-GFP^{hi}) and bottom (BACH2-GFP^{lo}) 30% of GFP^{+} transduced cells (**Figure S5a, b**). By segregating cells that either had a low or high levels of GFP expression, which is linked on the same transcript to the human *BACH2* cDNA through an internal ribosomal entry site, we observed a dose-dependent suppression of HbF levels, as assessed at both the mRNA and F-cell levels (**Figure S5c, d**). There was some perturbation of erythroid differentiation in this setting, which was most notable with high (~400-fold) overexpression of BACH2 (**Figure S5e**). Together, these complementary perturbations establish BACH2 as a direct and dosage-sensitive regulator of HbF.

Loss of BACH2 enhances NRF2 occupancy at the β-globin locus

BACH2 encodes a transcription factor that competes with NFE2-like proteins for binding to small Maf proteins at Maf recognition elements, thereby modulating gene expression across multiple loci[38]. We prioritized NRF2 and NFE2, since both have been shown to modulate transcription of the β-globin genes and therefore might alter HbF expression[35,39–44]. To investigate the mechanisms modulating HbF/ γ -globin activation, we mapped the chromatin occupancy of BACH2, NRF2, and NFE2 by CUT&RUN in human CD34^{+} HSPCs and differentiated erythroid cells. Genome-wide analysis of chromatin occupancy by all the factors in these cells revealed that their peaks were broadly distributed across promoter, distal intergenic, and intronic regions (**Figure S6a**). Within the β-globin loci, we observed strong BACH2 occupancy at the γ-globin promoter and also its intronic regions before differentiation, suggesting that BACH2 might restrict γ-globin activation by binding *cis*-regulatory elements of these genes. Upon erythroid differentiation, BACH2 binding further increased at the γ-globin promoter and extended to multiple DNase I hypersensitive sites within the local control regions (LCR) (**Figure 3a, b and Figure S6b**), suggesting a role in restricting activation of γ -globin through long-range chromatin interactions across the β-globin locus. NRF2 also showed robust occupancy at the γ-globin (*HBG1/2*) promoters, which was further enhanced at hypersensitive sites (HS) HS1–HS4 within the LCR following erythroid differentiation (**Figure 3c, d**). Of note, these regions overlapped with those bound by BACH2. By contrast, NFE2 showed strong occupancy at the LCR, but was only weakly bound at the γ-globin promoters (*HBG1/2*), with minimal changes in this occupancy pattern observed during differentiation (**Figure S6c, d**). Together, these results suggest that BACH2 and NRF2 may compete for binding at overlapping regulatory elements, including at the γ -globin promoters; whereas NFE2 likely plays a distinct role in globin gene regulation.

We therefore examined whether BACH2 modulates HbF expression by altering NRF2 binding events that are relevant for γ-globin activation. We performed NRF2 CUT&RUN upon BACH2 knockdown in HSPCs and differentiating erythroid cells. Consistent with this expectation, loss of BACH2 markedly reduced its occupancy at the γ-globin promoter in human CD34^{+} HSPCs, with further diminished binding at the γ-globin promoters and multiple hypersensitive sites within the LCR upon erythroid differentiation (**Figure S7a, b**). Importantly, loss of BACH2 increased NRF2 occupancy at multiple hypersensitive sites within the LCR and the γ-globin promoter in human CD34^{+} HSPCs, and this increase persisted during erythroid differentiation (**Figure 3e, f**). NRF2 ChIP-qPCR analysis confirmed enhanced NRF2 chromatin occupancy at the γ -globin (*HBG1/2*) promoter regions in differentiated erythroid cells with a BACH2 knockdown (**Figure S7c**). These findings indicate that BACH2 normally antagonizes NRF2 binding and its depletion allows enhanced NRF2 occupancy at key regulatory elements that might enable the activation of HbF expression.

BACH2 knockdown promotes NRF2 foci formation at the γ-globin locus

Having observed increased NRF2 chromatin occupancy at the γ-globin promoter and multiple HSs within the LCR upon BACH2 knockdown, we next asked whether this activation reflects a locus-specific effect at the β-globin cluster. Transcriptional activation of NRF2 targets primarily depends on its stabilization, nuclear accumulation, and translocation[45]. We therefore examined NRF2 subcellular localization by immunofluorescence in cells with BACH2 perturbation. Knockdown of BACH2 led to the appearance of discrete NRF2 nuclear foci, with an average of approximately two foci per cell in human CD34$^+$ HSPCs (**Figure 3g and Figure S7d, e**). Knockdown of BACH2 in K562 cells displayed an average of six to seven NRF2 foci per nucleus (**Figure S7f, g**). BACH2 knockdown likely relieves its repressive constraint on NRF2 binding, allowing NRF2 to accumulate at active chromatin regions and form discrete nuclear foci. The limited number of foci suggests that NRF2 recruitment might occur only at a subset of loci. We hypothesized that NRF2 might accumulate near the γ-globin genes to promote transcriptional activation. To test this, we performed combined NRF2 immunofluorescence and concomitant primary γ -globin (*HBG1/2*) RNA FISH in human CD34+ HSPCs and differentiated erythroid cells, which demonstrated that NRF2 nuclear foci colocalize with primary *HBG1/2* transcripts (**Figure 3h and Figure S7h, i**). Quantitative analysis further revealed that BACH2 knockdown cells with higher *HBG1/2* expression exhibited larger NRF2 foci and stronger colocalization (**Figure 3i**). Notably, analysis of oxidative stress-related genes in erythroid cells with BACH2 knockdown revealed no significant enrichment, and expression of canonical NRF2 targets remained largely unchanged (**Figure S7j, k**). Collectively, these findings support a model in which BACH2 loss allows NRF2 to occupy chromatin and

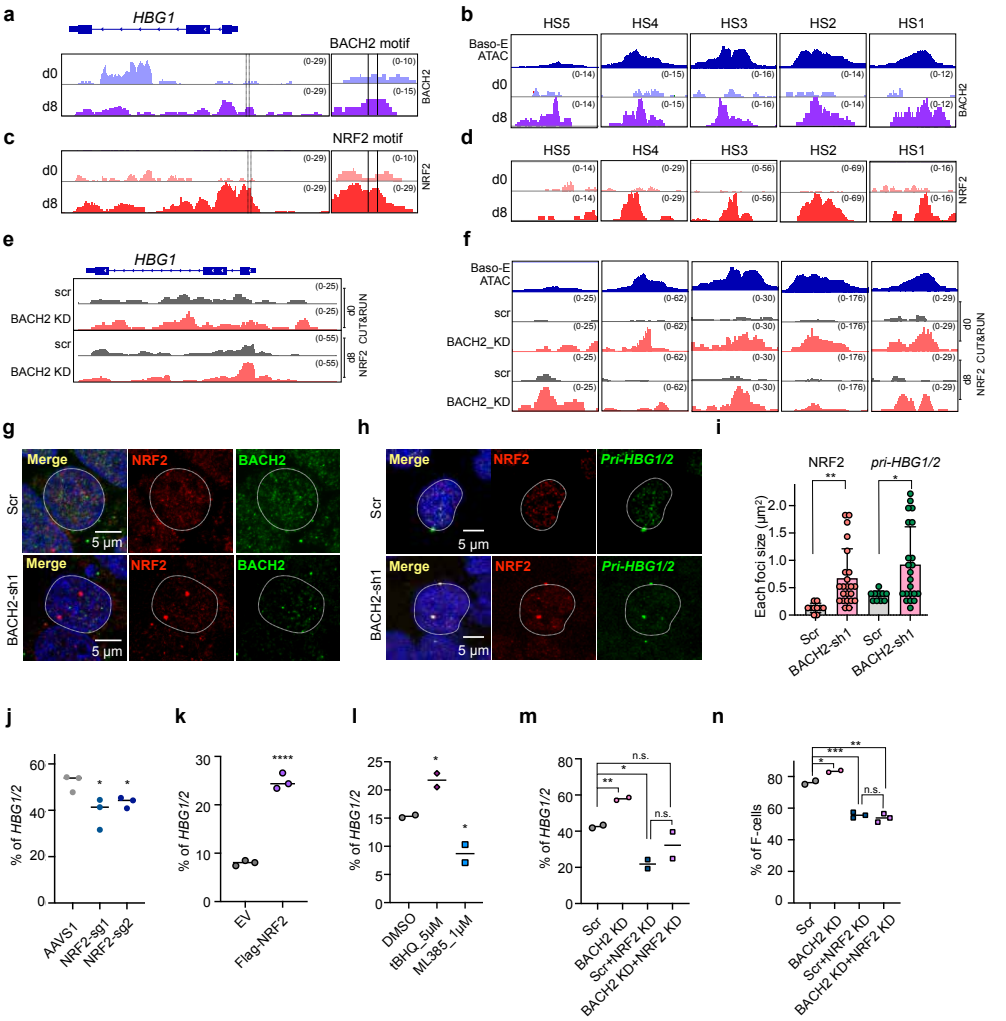

Figure 3: BACH2 knockdown activates NRF2 accumulation at the γ-globin locus. (a) BACH2 binding at the *HBG1* promoter on day 0 and day 8 of erythroid differentiation, with a zoomed-in view highlighting the BACH2 consensus motif at *HBG1* promoter, as revealed by CUT&RUN (used throughout this figure). **(b)** BACH2 binding at multiple local control regions (LCR) of β-globin clusters at d0 and d8 of erythroid differentiation. **(c)** NRF2 binding at the *HBG1* promoter at d0 and d8 of erythroid differentiation, with a zoomed-in view highlighting the NRF2 consensus motif at *HBG1* promote. **(d)** NRF2 binding at multiple LCR of β-globin cluster at d0 and d8 of erythroid differentiation. **(e)** NRF2 binding at *HBG1* promoter at d0 and d8 of erythroid differentiation upon BACH2 knockdown. **(f)** NRF2 binding at multiple LCR of β-globin cluster at d0 and d8 of erythroid differentiation upon BACH2 knockdown. **(g)** Co-staining of BACH2 and NRF2 in Scr control and BACH2 shRNA knockdown HSPCs by immunofluorescence. **(h)** Co-staining of NRF2 and *HBG1/2* primary transcript (*pri-HBG1/2*) in Scr control and BACH2 shRNA knockdown erythroid cells at day 8 of differentiation. **(i)** Quantification of NRF2 foci size and *pri-HBG1/2* transcript size in panel **h**, each dot represents one focus, n>=20 cells were quantified. **(j)** Proportion of *HBG1/2* expression on the day 7 of erythroid differentiated in *AAVS1* control and NRF2 CRISPR/Cas9 edited HSPCs. **(k)** Proportion of *HBG1/2* expression on day 11 of erythroid differentiation in empty vector (EV) and Flag tagged NRF2 overexpressed HSPCs. **(l)** Proportion of *HBG1/2* expression on day 11 of erythroid differentiation in DMSO control, NRF2 activator tBHQ or inhibitor ML385 treated HSPCs. **(m)** Proportion of *HBG1/2* expression on day 11 of erythroid differentiation in BACH2, NRF2 knockdown and combined BACH2 and NRF2 knockdown HSPCs. **(n)** Frequency of F-cells on day 11 of erythroid differentiation in Scr control, BACH2, NRF2 knockdown and combined BACH2 and NRF2 knockdown HSPCs. All data are presented as the mean ± the standard deviation; significance is indicated as $*p < 0.05$, $**p<0.01$, $***p < 0.001$, $****p < 0.0001$ or n.s. (not significant).

concentrate at γ -globin regulatory elements, thereby facilitating localized transcriptional activation.

NRF2 is functionally required for BACH2-mediated regulation of HbF

Since our prior data suggests a key role of NRF2 in enabling BACH2-mediated HbF regulation, we sought to directly perturb NRF2 expression and activity in primary human HSPCs undergoing erythroid differentiation (**Figure S8a, b**). CRISPR/Cas9-mediated depletion of NRF2 reduced HbF levels, as assessed by measurement of γ -globin (*HBG1/2*) mRNA abundance and the frequency of F-cells (**Figure 3j and Figure S8c-e**). Reciprocally, increased expression of NRF2 elevated HbF levels (**Figure 3k and Figure S8f, g**). Consistent with this observation and recent studies[43,44,46], we found that treatment of cells with the NRF2 inhibitor ML385[47] suppressed HbF, whereas activation of NRF2 by tBHQ induced HbF (**Figure 3l and Figure S8h-j**). Importantly, combined BACH2 and NRF2 depletion failed to restore HbF induction (**Figure 3m, n**), indicating that BACH2 primarily exerts its effects on HbF expression by antagonizing NRF2-mediated HbF activation. Together, these findings demonstrate that BACH2 constrains HbF expression by limiting NRF2-mediated activation of the γ-globin genes.

To assess whether BACH2 and NRF2 directly interact, we performed co-immunoprecipitation assays in multiple cell types. BACH2 was detected in NRF2 immunoprecipitants in HEK293T cells in which these proteins were exogenously expressed, and this association was independent of DNA (**Figure S9a-c**). Consistently, endogenous BACH2 co-precipitates with NRF2 in NRF2-overexpressing K562 cells (**Figure S9d**). Native PAGE analysis of recombinant BACH2 and NRF2 proteins showed that the two factors form a complex both in the presence and absence of the small Maf protein, MAFK (**Figure S9e**). Domain mapping revealed that BACH2 associates with NRF2 through its bZip domain (**Figure S9f**). Supporting the specificity of this interaction, we did not observe any association between BACH2 and NFE2 (**Figure S9g, h**). The observed association between NRF2 and BACH2 likely reflects their co-engagement within a shared transcriptional complex. Together, these findings support a model in which BACH2 constrains NRF2 activity within a common regulatory network, and the dissociation of BACH2

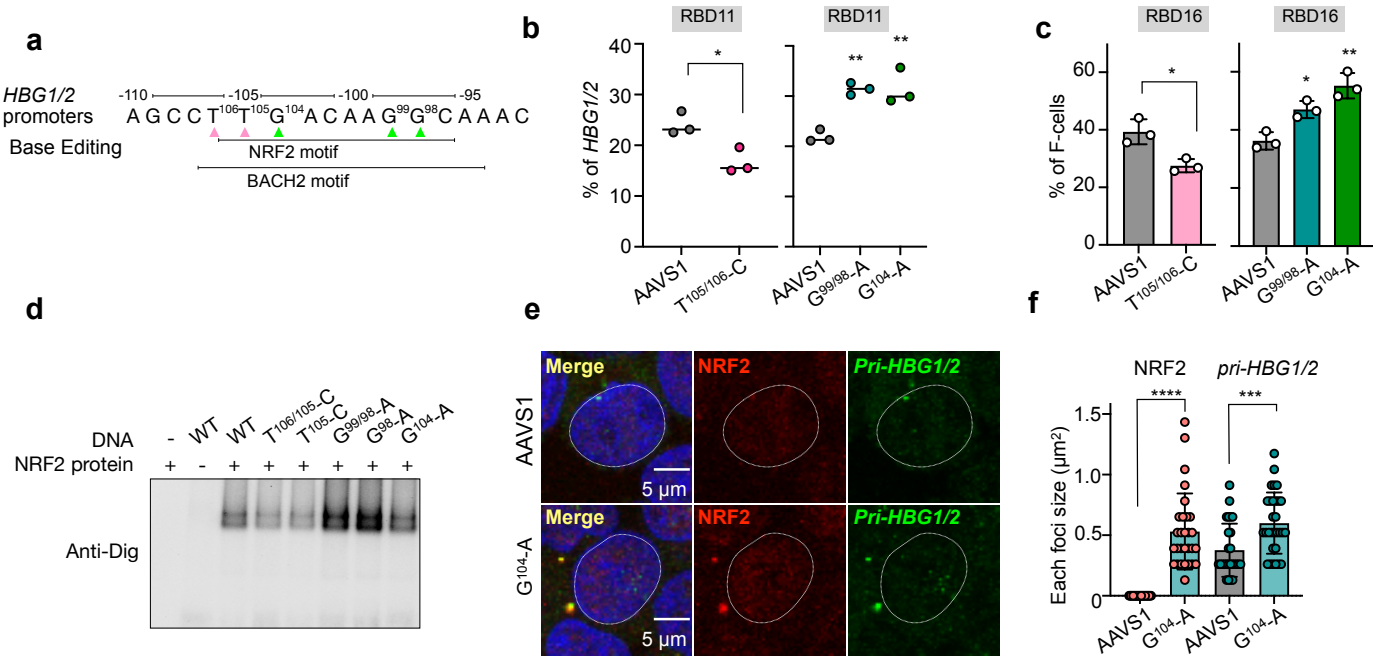

Figure 4: Editing of BACH2/NRF2 motif modulates HbF expression. (a) Schematic showing the positions of BACH2 and NRF2 motifs in the *HBG1/2* promoters. Green triangles indicate cytosine base editing (CBE) sites, and pink triangles indicate adenosine base editing (ABE) sites. Superscript numbers indicate nucleotide positions within the *HBG1/2* promoter. **(b)** Proportion of *HBG1/2* expression on day 11 of erythroid differentiated (RBD11) in *AAVS1* control and BACH2/NRF2 motif-edited HSPCs, edited using ABE and CBE respectively. **(c)** Frequency of F-cells on day 16 of erythroid differentiated (RBD16) in *AAVS1* control and BACH2/NRF2 motif-edited HSPCs, edited using ABE and CBE respectively. **(d)** EMSA examining NRF2 binding to DNA probes containing wild-type or base-edited NRF2 motifs within the HBG1/2 promoter. **(e)** Co-staining of NRF2 and *pri-HBG1/2* in *AAVS1* control and G³>A motif edited cells on day 8 of erythroid differentiated HSPCs. **(f)** Quantification of NRF2 foci size and *pri-HBG1/2* transcript size, each dot represents one focus, n>=30 cells were quantified. All data are presented as the mean ± the standard deviation; significance is indicated as $*p < 0.05$, $**p<0.01$, $***p < 0.001$, $****p < 0.0001$.

enables NRF2-mediated activation of γ -globin (*HBG1/2*) transcription.

Modulating HbF expression by editing BACH2/NRF2 motifs at the γ-globin promoters

Given the robust chromatin occupancy data, we analyzed the γ -globin promoters and identified highly overlapping, but functionally distinct motifs for BACH2 and NRF2 binding. Specifically, BACH2 has a consensus motif centered around the -94 to -106 region, and NRF2 has a motif centered around the -96 to -105 region (**Figure 4a**). To investigate how these *cis*-elements regulate HbF, we next perturbed the potential binding sites within the -95 to -106 region using either adenosine base editors (ABE) or cytosine base editors (CBE) recruited by single guide RNAs, followed by assessment of fetal hemoglobin expression (**Figure S10a**). We achieved >90% −105 T>C and >50% −106 T>C substitutions using ABE, and ~50% −98 G>A, −99 G>A, and −104 G>A substitutions using CBE (**Figure S10b**). Interestingly, the −105 and −106 T>C substitutions reduced HbF expression, whereas the −98, −99, and −104 G>A substitutions induced HbF expression (**Figure 4b, c and Figure S10c-e**). Importantly, perturbation of these *cis*-elements in the γ -globin promoter did not impact erythroid differentiation (**Figure S10f**). Perturbing these potential binding sites exerts differential effects on HbF expression, suggesting that they may differentially modulate the binding preference of BACH2 or NRF2 within this region.

To assess DNA binding by NRF2 or BACH2, we first carried out *in silico* predictions of transcription factor binding affinities at the edited *cis*-elements, which revealed that the −105 and −106 T>C substitutions were predicted to markedly reduce NRF2 binding, whereas the −98, −99, and −104 G>A substitutions were predicted to reduce BACH2 binding (**Figure S10g**). We experimentally validated variation in NRF2 binding affinity to the edited *cis*-elements by electrophoretic mobility shift assays (EMSA), which revealed that the −105 and −106 T>C substitutions indeed reduce NRF2-DNA interaction and the −98 and −99 G>A and −104 G>A substitutions increased NRF2-DNA binding (**Figure 4d**). In parallel, we also observed NRF2 accumulation at primary *HBG1/2* transcripts upon cytosine base editing (**Figure 4e, f and and Figure S10h**), supporting the functional relevance of NRF2 binding in γ-globin activation and the associated increase in HbF expression.

BACH2 regulates HbF independently of BCL11A

BCL11A is a key repressor of HbF and binds two motifs within the γ-globin promoters, a distal site (−113 to −118) and a proximal site (−86 to −91)[18,19]. The NRF2/BACH2 motifs (−94 to −106) are positioned between these two BCL11A sites (**Figure 5a**). We therefore investigated whether BACH2 regulates HbF expression in a BCL11A-dependent manner. We inhibited BCL11A expression by genome editing of the +58 erythroid-specific enhancer of BCL11A concomitant with BACH2 perturbation. Knockdown of BACH2 or BCL11A individually promotes HbF expression upon erythroid differentiation, whereas simultaneous depletion of both factors resulted in an additive effect, producing higher HbF levels than either single knockdown alone (**Figure 5b**). To examine whether these proteins physically interact, we performed co-immunoprecipitation assays by overexpressing BCL11A with

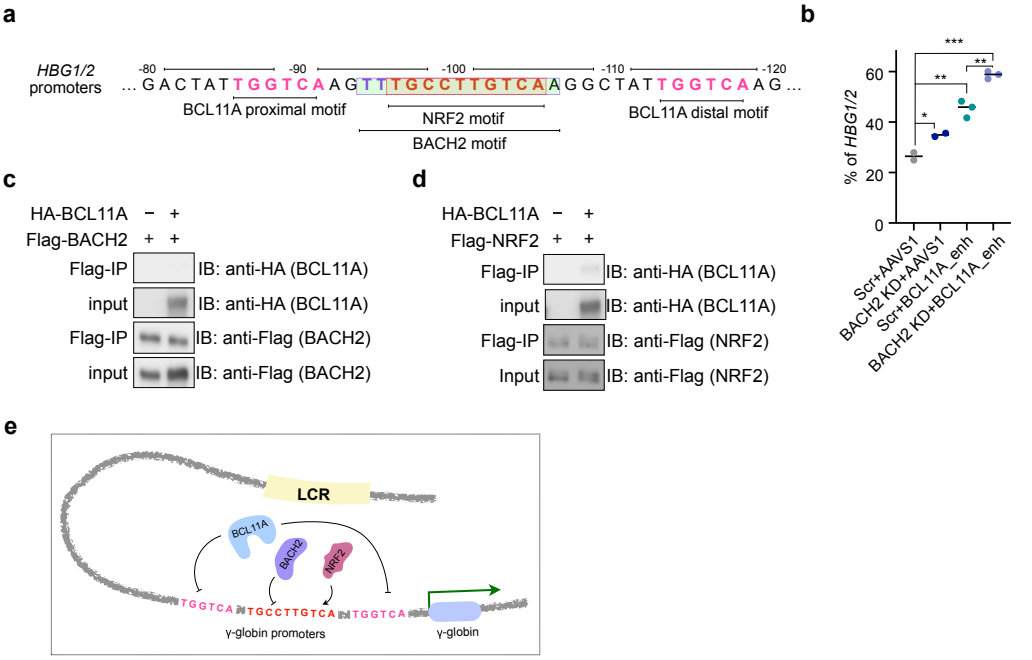

Figure 5: BACH2 regulates HbF independent of BCL11A. (a) Schematic of the positions of BACH2 motif, NRF2 motif and BCL11A distal and proximal motif in *HBG1/2* promoters. **(b)** Proportion of *HBG1/2* expression on day 11 of erythroid differentiated in HSPCs with BACH2 combined BCL11A knockdown. **(c)** BCL11A is not associated with BACH2, assays were performed by anti-Flag tagged BACH2 in HEK293 overexpressed cells, followed by WB with anti-HA tagged BCL11A. **(d)** BCL11A is not associated with NRF2, assays were performed by anti-Flag tagged NRF2 in HEK293 overexpressed cells, followed by WB with anti-HA tagged BCL11A. **(e)** Schematic of the working model illustrating how BACH2, NRF2, and BCL11A binding at the γ-globin promoters regulate HbF expression. BCL11A binds to the proximal and distal motifs to repress HbF expression. BACH2 and NRF2 competitively bind to the motif located between the proximal and distal BCL11A motifs to repress or activate HbF expression. Arrows indicate activation of HbF, and blocked arrows indicate repression. All data are presented as the mean ± the standard deviation; significance is indicated as $*p<0.05$, $**p<0.01$, $***p<0.001$.

BACH2 or NRF2. Neither BCL11A-BACH2 nor BCL11A-NRF2 interactions were detected (**Figure 5c, d**). These results indicate that BACH2-mediated regulation of HbF occurs independently of BCL11A and that the two pathways act in parallel (**Figure 5e**).

Discussion

Tremendous progress in our understanding of HbF regulation has emerged through studies of human genetic variation, enabling therapeutic advances. However, the studies of genetic variation to date have been limited in size to several thousands of individuals at most and have typically been focused on specific ancestry groups[48–50]. Concomitantly, studies of individuals with exceptionally elevated HbF levels have revealed rare genetic variation at the β-globin gene locus and in other genes, including *BCL11A*, *KLF1*, and *ZBTB7A*, which leads to larger increases in HbF levels[8,17,51,52]. Here, by conducting the biggest multi-ancestry GWAS of HbF levels to date we have uncovered a number of previously undescribed loci underlying variation in HbF levels. Leveraging the multi-ancestry cohorts further increased the power and robustness of our discoveries by capturing more diverse LD structures and genetic backgrounds. Cross-examining populations of distinct ancestries not only highlights authentic variants that are functionally consequential across all cohorts, but also provides reassurance that there is no confounding resulting from population stratification. These new insights have led to the identification of BACH2 as a genetically-nominated regulator of HbF.

To date, genetic variation in *BACH2* has primarily been associated with autoimmune and allergic diseases[53,54], with little attention given to its potential roles in HbF regulation. Here, by following up on insights from our human genetic studies, we provide functional evidence that BACH2 restrains NRF2-mediated activation of HbF. Critically, we disentangle the mechanisms through which these factors modulate activation of the γ-globin promoters. These findings establish the BACH2-NRF2 axis as a tractable and potentially therapeutically targetable regulatory node involved in HbF activation. Importantly, we demonstrate that this axis is independent of the well-studied and targeted BCL11A-dependent repressive mechanisms, suggesting synergistic opportunities to modulate HbF expression by targeting both pathways in tandem.

In summary, by studying the genetic basis of variation in HbF levels across diverse populations, we uncovered previously unidentified genetic variation and new biological insights, particularly defining the BACH2-NRF2 axis in enabling HbF activation. Many further biological mechanisms are likely to be discovered through the study of additional loci identified here and through further expansion of such HbF GWAS. Indeed, a recent study has similarly employed more recent genetic association findings to define a distinct regulator of HbF, ABCC1[55]. Our study provides an example of how increasingly larger GWAS can enable further insights into even seemingly well-understood and therapeutically-important human phenotypes like HbF.

Reference

1. Sankaran, V. G. & Orkin, S. H. The switch from fetal to adult hemoglobin. *Cold Spring Harb. Perspect. Med.* **3**, a011643 (2013).
2. Akinsheye, I. *et al.* Fetal hemoglobin in sickle cell anemia. *Blood* **118**, 19–27 (2011).
3. Weatherall, D. J. Phenotype-genotype relationships in monogenic disease: lessons from the thalassaemias. *Nat Rev Genet* **2**, 245–255 (2001).
4. Uda, M. *et al.* Genome-wide association study shows BCL11A associated with persistent fetal hemoglobin and amelioration of the phenotype of beta-thalassemia. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 1620–1625 (2008).
5. 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. U. S. A.* **105**, 11869–11874 (2008).
6. 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).
7. Sankaran, V. G. *et al.* Human fetal hemoglobin expression is regulated by the developmental stage-specific repressor BCL11A. *Science* **322**, 1839–1842 (2008).
8. Basak, A. *et al.* BCL11A deletions result in fetal hemoglobin persistence and neurodevelopmental alterations. *J. Clin. Invest.* **125**, 2363–2368 (2015).
9. Dias, C. *et al.* BCL11A Haploinsufficiency Causes an Intellectual Disability Syndrome and Dysregulates Transcription. *Am. J. Hum. Genet.* **99**, 253–274 (2016).
10. Shen, Y. *et al.* Pathogenic BCL11A variants provide insights into the mechanisms of human fetal hemoglobin silencing. *PLoS Genet.* **17**, e1009835 (2021).
11. Bauer, D. E. *et al.* An erythroid enhancer of BCL11A subject to genetic variation determines fetal hemoglobin level. *Science* **342**, 253–257 (2013).
12. Canver, M. C. *et al.* BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. *Nature* **527**, 192–197 (2015).
13. Basak, A. *et al.* Control of human hemoglobin switching by LIN28B-mediated regulation of BCL11A translation. *Nat. Genet.* **52**, 138–145 (2020).
14. de Vasconcellos, J. F. *et al.* IGF2BP1 overexpression causes fetal-like hemoglobin expression patterns in cultured human adult erythroblasts. *Proc. Natl. Acad. Sci. U. S. A.* **114**, (2017).
15. Lee, Y. T. *et al.* LIN28B-mediated expression of fetal hemoglobin and production of fetal-like erythrocytes from adult human erythroblasts ex vivo. *Blood* **122**, (2013).
16. Huang, P. *et al.* HIC2 controls developmental hemoglobin switching by repressing BCL11A transcription. *Nat. Genet.* **54**, (2022).
17. Shen, Y. *et al.* A unified model of human hemoglobin switching through single-cell genome editing. *Nat. Commun.* **12**, 4991 (2021).
18. Liu, N. *et al.* Direct Promoter Repression by BCL11A Controls the Fetal to Adult Hemoglobin Switch. *Cell* **173**, 430–442.e17 (2018).
19. Martyn, G. E. *et al.* Natural regulatory mutations elevate the fetal globin gene via disruption of BCL11A or ZBTB7A binding. *Nat. Genet.* **50**, 498–503 (2018).
20. Esrick, E. B. *et al.* Post-Transcriptional Genetic Silencing of BCL11A to Treat Sickle Cell Disease. *N. Engl. J. Med.* **384**, 205–215 (2021).
21. Frangoul, H. *et al.* CRISPR-Cas9 Gene Editing for Sickle Cell Disease and β-Thalassemia. *N. Engl. J. Med.* **384**, 252–260 (2021).
22. Qin, K. *et al.* Dual function NFI factors control fetal hemoglobin silencing in adult erythroid cells. *Nat Genet* **54**, 874–884 (2022).
23. Zhou, D., Liu, K., Sun, C.-W., Pawlik, K. M. & Townes, T. M. KLF1 regulates BCL11A expression and gamma- to beta-globin gene switching. *Nat Genet* **42**, 742–744 (2010).
24. Borg, J. *et al.* Haploinsufficiency for the erythroid transcription factor KLF1 causes hereditary persistence of fetal hemoglobin. *Nat Genet* **42**, 801–805 (2010).
25. Huang, P. *et al.* The HRI-regulated transcription factor ATF4 activates BCL11A transcription to silence fetal hemoglobin expression. *Blood* **135**,

2121–2132 (2020).
26. Feng, R. *et al.* Activation of γ-globin expression by hypoxia-inducible factor 1α. *Nature* **610**, 783–790 (2022).
27. Masuda, T. *et al.* Transcription factors LRF and BCL11A independently repress expression of fetal hemoglobin. *Science* **351**, 285–289 (2016).
28. Bao, E. L. *et al.* Heritability of fetal hemoglobin, white cell count, and other clinical traits from a sickle cell disease family cohort. *Am. J. Hematol.* **94**, 522–527 (2019).
29. Vuckovic, D. *et al.* The Polygenic and Monogenic Basis of Blood Traits and Diseases. *Cell* **182**, 1214–1231.e11 (2020).
30. Chen, M.-H. *et al.* Trans-ethnic and Ancestry-Specific Blood-Cell Genetics in 746,667 Individuals from 5 Global Populations. *Cell* **182**, 1198–1213.e14 (2020).
31. Yu, F. *et al.* Variant to function mapping at single-cell resolution through network propagation. *Nat. Biotechnol.* **40**, 1644–1653 (2022).
32. Caulier, A. L. & Sankaran, V. G. Molecular and cellular mechanisms that regulate human erythropoiesis. *Blood* **139**, 2450–2459 (2022).
33. Schep, A. N., Wu, B., Buenrostro, J. D. & Greenleaf, W. J. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat Methods* **14**, 975–978 (2017).
34. Barbarani, G., Fugazza, C., Strouboulis, J. & Ronchi, A. E. The Pleiotropic Effects of GATA1 and KLF1 in Physiological Erythropoiesis and in Dyserythropoietic Disorders. *Front. Physiol.* **10**, 91 (2019).
35. Brand, M. *et al.* Dynamic changes in transcription factor complexes during erythroid differentiation revealed by quantitative proteomics. *Nat. Struct. Mol. Biol.* **11**, 73–80 (2004).
36. Palani, C. D., Zhu, X., Alagar, M., Attucks, O. C. & Pace, B. S. Bach1 inhibitor HPP-D mediates γ-globin gene activation in sickle erythroid progenitors. *Blood Cells Mol. Dis.* **104**, (2024).
37. Son, J. *et al.* BACH2 inhibition reverses β cell failure in type 2 diabetes models. *J. Clin. Invest.* **131**, (2021).
38. Oyake, T. *et al.* Bach proteins belong to a novel family of BTB-basic leucine zipper transcription factors that interact with MafK and regulate transcription through the NF-E2 site. *Mol. Cell. Biol.* **16**, 6083–6095 (1996).
39. Sun, J. *et al.* Heme regulates the dynamic exchange of Bach1 and NF-E2-related factors in the Maf transcription factor network. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 1461–1466 (2004).
40. Andrews, N. C. *et al.* The ubiquitous subunit of erythroid transcription factor NF-E2 is a small basic-leucine zipper protein related to the v-maf oncogene. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 11488–11492 (1993).
41. Sawado, T., Igarashi, K. & Groudine, M. Activation of beta-major globin gene transcription is associated with recruitment of NF-E2 to the beta-globin LCR and gene promoter. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 10226–10231 (2001).
42. Zhu, X., Li, B. & Pace, B. S. NRF2 mediates γ-globin gene regulation and fetal hemoglobin induction in human erythroid progenitors. *Haematologica* **102**, e285–e288 (2017).
43. Zhu, X., Oseghale, A. R., Nicole, L. H., Li, B. & Pace, B. S. Mechanisms of NRF2 activation to mediate fetal hemoglobin induction and protection against oxidative stress in sickle cell disease. *Exp. Biol. Med.* **244**, 171–182 (2019).
44. Krishnamoorthy, S. *et al.* Dimethyl fumarate increases fetal hemoglobin, provides heme detoxification, and corrects anemia in sickle cell disease. *JCI Insight* **2**, (2017).
45. Zhang, D. D. Thirty years of NRF2: advances and therapeutic challenges. *Nat Rev Drug Discov* **24**, 421–444 (2025).
46. Zhu, X. *et al.* NRF2 mediates γ-globin gene regulation through epigenetic modifications in a β-YAC transgenic mouse model. *Exp. Biol. Med.* **245**, 1308 (2020).
47. Singh, A. *et al.* Small Molecule Inhibitor of NRF2 Selectively Intervenes Therapeutic Resistance in KEAP1-Deficient NSCLC Tumors. *ACS Chem. Biol.* **11**, (2016).
48. Danjou, F. *et al.* Genome-wide association analyses based on whole-genome sequencing in Sardinia provide insights into regulation of hemoglobin levels. *Nat. Genet.* **47**, 1264–1271 (2015).
49. Mtatiro, S. N. *et al.* Genome wide association study of fetal hemoglobin in sickle cell anemia in Tanzania. *PLoS One* **9**, e111464 (2014).
50. Liu, L. *et al.* Original Research: A case-control genome-wide association study identifies genetic modifiers of fetal hemoglobin in sickle cell disease. *Exp. Biol. Med.* **241**, 706–718 (2016).
51. Perkins, A. *et al.* Krüppeling erythropoiesis: an unexpected broad spectrum of human red blood cell disorders due to KLF1 variants. *Blood* **127**, 1856–1862 (2016).
52. von der Lippe, C. *et al.* Heterozygous variants in ZBTB7A cause a neurodevelopmental disorder associated with symptomatic overgrowth of pharyngeal lymphoid tissue, macrocephaly, and elevated fetal hemoglobin. *Am. J. Med. Genet. A* **188**, 272–282 (2022).
53. Ferreira, M. A. R. *et al.* Identification of IL6R and chromosome 11q13.5 as risk loci for asthma. *Lancet* **378**, 1006–1014 (2011).
54. Jin, Y. *et al.* Genome-wide association analyses identify 13 new susceptibility loci for generalized vitiligo. *Nat Genet* **44**, 676–680 (2012).
55. Hara, Y. *et al.* Genomic discovery and functional validation of MRP1 as a novel fetal hemoglobin modulator and potential therapeutic target in sickle cell disease. *medRxiv* 2023.03.14.23287244 (2023) doi:10.1101/2023.03.14.23287244.
56. Di Angelantonio, E. *et al.* Efficiency and safety of varying the frequency of whole blood donation (INTERVAL): a randomised trial of 45 000 donors. *Lancet* **390**, 2360–2371 (2017).
57. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).
58. Võsa, U. *et al.* Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat. Genet.* **53**, 1300–1310 (2021).
59. Gladwin, M. T. *et al.* Risk factors for death in 632 patients with sickle cell disease in the United States and United Kingdom. *PLoS One* **9**, e99489 (2014).
60. Ashley-Koch, A. E. *et al.* Identification of genetic polymorphisms associated with risk for pulmonary hypertension in sickle cell disease. *Blood* **111**, 5721–5726 (2008).
61. Kleinman, S. *et al.* The National Heart, Lung, and Blood Institute Recipient Epidemiology and Donor Evaluation Study (REDS-III): a research program striving to improve blood donor and transfusion recipient outcomes. *Transfusion* **54**, 942–955 (2014).
62. Hankins, J. S. *et al.* Sickle Cell Clinical Research and Intervention Program (SCCRIP): A lifespan cohort study for sickle cell disease progression from the pediatric stage into adulthood. *Pediatr. Blood Cancer* **65**, e27228 (2018).
63. Rampersaud, E. *et al.* A polygenic score for acute vaso-occlusive pain in pediatric sickle cell disease. *Blood Adv* **5**, 2839–2851 (2021).
64. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
65. Turley, P. *et al.* Multi-Ancestry Meta-Analysis yields novel genetic discoveries and ancestry-specific associations. *bioRxiv* 2021.04.23.441003 (2021) doi:10.1101/2021.04.23.441003.
66. 1000 Genomes Project Consortium *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
67. All of Us Research Program Genomics Investigators. Genomic data in the All of Us Research Program. *Nature* **627**, 340–346 (2024).
68. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
69. Benner, C. *et al.* FINEMAP: efficient variable selection using summary data from genome-wide association studies. *Bioinformatics* **32**, 1493–1501 (2016).
70. Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
71. Speed, D., Hemani, G., Johnson, M. R. & Balding, D. J. Improved heritability estimation from genome-wide SNPs. *Am. J. Hum. Genet.* **91**, 1011–1021 (2012).
72. Granja, J. M. *et al.* Single-cell multiomic analysis identifies regulatory programs in mixed-phenotype acute leukemia. *Nat. Biotechnol.* **37**, 1458–1465 (2019).
73. Street, K. *et al.* Slingshot: cell lineage and pseudotime inference for single-

cell transcriptomics. *BMC Genomics* **19**, 477 (2018).
74. Ludwig, L. S. *et al.* Transcriptional States and Chromatin Accessibility Underlying Human Erythropoiesis. *Cell Rep.* **27**, 3228–3240.e7 (2019).
75. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137 (2008).
76. Javierre, B. M. *et al.* Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. *Cell* **167**, 1369–1384.e19 (2016).
77. Bao, E. L. *et al.* Inherited myeloproliferative neoplasm risk affects haematopoietic stem cells. *Nature* **586**, 769–775 (2020).
78. Voit, R. A. *et al.* A genetic disorder reveals a hematopoietic stem cell regulatory network co-opted in leukemia. *Nat. Immunol.* **24**, 69–83 (2023).
79. Giani, F. C. *et al.* Targeted Application of Human Genetic Variation Can Improve Red Blood Cell Production from Stem Cells. *Cell Stem Cell* **18**, 73–78 (2016).
80. Nandakumar, S. K. *et al.* Gene-centric functional dissection of human genetic variation uncovers regulators of hematopoiesis. *Elife* **8**, (2019).
81. Martin-Rufino, J. D. *et al.* Massively parallel base editing to map variant effects in human hematopoiesis. *Cell* **186**, (2023).
82. Neugebauer, M. E. *et al.* Evolution of an adenine base editor into a small, efficient cytosine base editor with low off-target activity. *Nat Biotechnol* **41**, 673–685 (2023).
83. de Smith, A. J. *et al.* A noncoding regulatory variant in IKZF1 increases acute lymphoblastic leukemia risk in Hispanic/Latino children. *Cell Genom* **4**, 100526 (2024).
84. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
85. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
86. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).
87. Skene, P. J. & Henikoff, S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *Elife* **6**, (2017).
88. Zhu, Q., Liu, N., Orkin, S. H. & Yuan, G.-C. CUT&RUNTools: a flexible pipeline for CUT&RUN processing and footprint analysis. *Genome Biol* **20**, 192 (2019).
89. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
90. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357–359 (2012).
91. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**, (2021).
92. Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44**, W160–5 (2016).
93. Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017–1018 (2011).

Acknowledgements

We are grateful to members of the Sankaran laboratory for valuable comments and discussion, as well as S. Orkin and D. Nathan for their inspiration and guidance. We acknowledge early work on these cohorts by J. Verboon, A. Cheng, and C. Fiorini. We thank V. Kuchroo, A. Schnell, and Y. Hou for providing reagents. This work was supported by the Howard Hughes Medical Institute (to V.G.S), the New York Stem Cell Foundation (to V.G.S.), and National Institutes of Health (NIH) grants R01 DK103794 and R01 HL146500 (to V.G.S). V.G.S. is an investigator of the Howard Hughes Medical Institute.

Author contributions

V.G.S., C.-J.G, U.P.A, L.D.C.conceptualized and designed the study. C.-J.G, U.P.A, L.D.C., R.L., H.Y.L., F.Y., M.W., B.M., S.E., P.D., L.M., R.S., T.S., S.R., P.G.B., D.S.P., E.K., W.J.A., F.A., K.A., A.L.L.P, G.K., Y.Z., S.N., V.R.G., M.T.G, M.E.G, A.A., M.J.T., B.C., S.K., C.D., E.C.S., P.L., A.C., C.M., T.M., A.M., A.H., G.T., K.S., U.T., V.A.S., M.J.W., L.F., B.N., A.S.B., V.V., S.N., V.G.S. obtained and provided cohort data. C.-J.G, U.P.A, W.X, L.D.C., R.L., H.Y.L., F.Y., M.W., B.M., P.D., S.E., F.A. performed functional studies and computational analyses. V.G.S., C.-J.G, U.P.A., W.X, L.D.C., R.L., H.Y.L. wrote the original draft with input from all authors. V.G.S. provided overall study oversight. All authors were involved in reviewing and editing the manuscript. All authors read and approved the final version of the manuscript.

Competing interest statement

During the drafting of the manuscript, D.S.P. became a full-time employee of AstraZeneca. F.A. is an employee and shareholder of Illumina, Inc. M.T.G. serves as a consultant for Actelion, Bayer Healthcare, Pfizer, Forma, and Fulcrum Therapeutics. A.H.L. reports speakers honoraria from Siemens Healthineers and Beckman Diagnostics, as well as participation on an advisory board of Roche Diagnostics, all unrelated to the present work. A.S.B. reports institutional grants from AstraZeneca, Bayer, Biogen, BioMarin, Bioverativ, Novartis, Regeneron and Sanofi. V.G.S. serves as an advisor to Ensoma, Cellarity, and Beam Therapeutics, all unrelated to the present work.

Data availability

Raw and processed data have been deposited to public databases, and will be available at the time of publication. GWAS summary statistics will also be made available at the time of publication.

Materials and Methods

Individual GWAS study methods and quality control

Details including study design, genotyping, and imputation methods and quality control for the included studies is in **Supplementary Table 1**. The cohorts relied upon different selection strategies, including unselected individuals from the population (Swedish, SardiNIA[4], INTERVAL[56], GTEx[57], BIOS[58]), individuals with sickle cell disease (Tanzania[49], Walk-PHaSST[59], OMG-SCD[60], REDS-III Brazil[61], St. Jude Sickle Cell Clinical Research & Intervention Program (SCCRIP)/Baylor[62,63]), or individuals selected from a screened population (Thai). We examined the relatedness among individuals in TOPMed (including Walk-PHaSST, OMG-SCD, and REDS-III Brazil), Tanzania, and Thai cohorts, and did not find outliers to be excluded. Data from the other cohorts do not include individual genotypes, but we assumed the relatedness has been accounted for in their cohort-specific association studies. Most included samples had HbF measured in the traditional way using high performance liquid chromatography, however two of the included cohorts (BIOS and GTEx) derived the HbF phenotype from expression data (in TPM units) as a ratio of gene expression of (*HBG1* + *HBG2*) / *HBB*. We found this to faithfully replicate expected results from the traditionally measured HbF cohorts. In addition, we found the ratio approach to traditionally measured approach in a EUR subset to be genetically correlated $r_g = 0.59$ (SE 0.317). The included Swedish and Thai populations are previously undescribed cohorts, and were analyzed specifically for this study. All participants provided informed consent and all studies obtained ethical approvals from local ethics review boards.

All GWAS summary statistics were lifted-over from their respective genome builds to reference genome hg38. Alleles were flipped according to the hg38 build reference allele, and if neither allele was present the variant was removed. Strand ambiguous and non-biallelic SNPs were removed. Minor allele frequency was filtered to >=0.1%. RSIDs were assigned using dbSNP version 144. All models included adjustment for age, sex, and the top 10 principal components of ancestry. In addition, for the SCD cohort analyses there was appropriate adjustment for SCD genotype and hydroxyurea use. In all cohorts, the same transformation (inverse-normalization) was performed on HbF to produce a normalized response variable.

The Thai cohort has a unique study design: from a large general population of ~86,000 individuals, we intentionally screened for HbF and sampled individuals with HbF levels higher than 2% of total hemoglobin. We then randomly selected other individuals from this same general population to ensure an appropriate distribution of HbF levels within the sampled cohort. DNA extraction, sequencing, and genotyping were performed after the sample selection. HbF levels for the entire cohort (enriched for individuals with high HbF) were transformed using an inverse normalization transformation and checked to confirm normality prior to GWAS analysis.

Meta-analysis

Fixed effects meta-analysis (FEMA) was performed using METAL r2020-05-05 (github.com/statgen/METAL)[64]. Multi-ancestry Meta analysis (MAMA)

provides improved power in meta-analysis of different populations with low type 1 error rates[65]. We used MAMA per population using LD reference panels derived from a combination of 1000 genomes[66] data and Thai population whole genome sequenced (WGS) samples.

Identifying conditionally independent loci and fine-mapping

External LD reference panels were created from AllOfUs (v5) data[67] for European (EUR, n=51125), African (AFR, n=22837) and a combination of East-Asian, South Asian and Thai population-specific WGS data (EAS-SAS-THAI, n=3788). EUR and AFR analyses used panels limited to 10,000 individuals for computational efficiency. GCTA-COJO v1.94.0[68] was used to identify conditionally independent loci that became our LD-sentinel markers. Each LD-sentinel marker was expanded by 500kb on either side and treated as a 1MBp window which was labeled with the nearest or, if known, biologically relevant gene. Once each region was determined, fine-mapping was performed using FINEMAP[69] with the following parameters: **--**non-funct **--**allow-missing **--**max-num-causal 1.

Heritability and genetic correlation analyses

LD scores were established from AllofUs (v5) LD panels as described above, using maximum 1cM window positions. LDSC[70] was performed on summary statistics restricted to high quality, HapMap 3 variants. LDAK[71] was also used to estimate heritability using the thin and BLD models appropriate for ancestry. Genetic correlations with blood cell traits were estimated using LDAK, using the thin model. Summary statistics for blood cell traits were obtained from the published BCX2 consortium summary statistics available at http://www.mhi-humangenetics.org/en/resources/.

SCAVENGE

SCAVENGE[31] was performed following the tutorial (https://sankaranlab.github.io/SCAVENGE/articles/SCAVENGE). We first obtained fine-mapped variants with posterior probabilities of causality as described above. We then collected single-cell ATAC-seq data from 10 individuals, comprising 33,819 cells across 23 annotated cell populations[72]. Slingshot[73] was used to reconstruct the developmental progression of erythroid lineage and to assign pseudotime to each cell along the trajectory. Using the fine-mapped variants and single-cell chromatin accessibility data as input, we calculated trait relevance scores (TRSs) for individual cells, which indicate their association with the trait of interest. In parallel, we generated pseudo-bulk profiles based on cell-type annotations to construct pseudo-bulk ATAC-seq datasets as previously described[74], and used these data as input for peaking calling with MACS2[75].

Promoter Capture Hi-C (PCHi-C)

For annotating genomic windows with genes, we used a published promoter capture Hi-C (PCHi-C)[76] dataset spanning 15 haematopoietic cell types. We retained chromatin looping interactions with a CHiCAGO score > 5.

Human primary HSPC culture and in vitro erythroid differentiation

Human CD34$^+$ HSPCs from mobilized peripheral blood of healthy adults were obtained from the Cooperative Center of Excellence in Hematology at the Fred Hutchinson Cancer Research Center. The primary HSPCs were thawed into a maintenance medium consisting of a StemSpan II base (StemCell Technologies), CC100 (StemCell Technologies), 50 ng/mL human TPO (Pepro Tech) and 1% penicillin/streptomycin (Life Technologies) and 1% of L-Glutamine (Life Technologies)[77,78].

After the maintenance phase, primary human HSPCs were differentiated using the three-phase culture system previously described[79,80]. First, a base erythroid medium was created by supplementing IMDM (Gibco) with 2% human AB plasma (SeraCare), 3% human AB serum (Life Technologies), 3 U/mL heparin, 10 μg/mL insulin, 200 μg/mL holo-transferrin, and 1% penicillin/streptomycin. From days 1-7 in erythroid media, this base medium was further supplemented with 3 U/mL EPO, 10 ng/mL human SCF, and 1 ng/mL IL-3. From days 7-12, this base medium was further supplemented with 3 U/mL EPO and 10 ng/mL human SCF. After day 12, the base medium was supplemented with 1 mg/mL total of holo-transferrin and 3 U/mL of EPO.

Cell line culture

HEK293T cells were cultured in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% FBS and 1% penicillin/streptomycin. K562 cells were cultured in Iscove's Modified Dulbecco's Medium (IMDM) supplemented with 10% FBS and 1% Penicillin-Streptomycin.

Genome editing of human primary HSPCs

Genome editing was performed in human primary CD34$^+$ HSPCs using the 4D-Nucleofector system (Lonza) with the P3 Primary Cell 4D-Nucleofector X Kit S (Lonza). For CRISPR-Cas9-mediated enhancer deletion, ribonucleoprotein (RNP) complexes were assembled by combining 100 pmol Cas9 protein (IDT) with 100 pmol chemically modified sgRNAs (Synthego) and incubating at room temperature for 15 min. RNPs were delivered into CD34$^+$ HSPCs with the P3 nucleofection reagent supplemented with nucleofection enhancer (IDT) at a 20:1 ratio, using program DZ-100 on the 4D-Nucleofector system. To generate the high-HbF–associated variant rs1010474, 20 μg of adenosine base editor (ABE8e)[81] protein was incubated with 100 pmol chemically modified sgRNA (Synthego) at room temperature for 20 min. The assembled ABE-RNP complexes were delivered into CD34$^+$ HSPCs using P3 reagent and program DZ-100. For editing NRF2/BACH2 motifs within the *HBG1/2* promoters, 2 μg of *in vitro* transcribed ABE8e or TadCBEb[82] were mixed with 100 pmol chemically modified sgRNAs (Synthego) and nucleofected into CD34$^+$ HSPCs using program DS-130. Electroporated cells were harvested 3 days after nucleofection for genomic DNA and RNA extraction using the AllPrep DNA/RNA Micro Kit (Qiagen) following the manufacturer's instructions. All guide RNA sequences used for genome editing are listed in **Supplementary Table 11**.

BACH2 shRNA knockdown and overexpression cloning and lentivirus packing

To knock down BACH2, DNA sequences for shRNAs were designed by the GPP Web Portal online tool (https://portals.broadinstitute.org/gpp/public/). DNA oligos of shRNA sequences for targeted RNAs or scramble shRNAs were individually cloned into pLKO.1-GFP vector. For overexpression, the human *BACH2* coding sequence was synthesized by Azenta and inserted into the HMD lentiviral vector. All the plasmids were validated by whole plasmid sequencing by plasmidsaurus. Lentiviral particles were produced as previously described. Briefly, 5×10^6 HEK293T cells in a 10 cm dish were co-transfected with 10 μg pLKO.1 shRNA constructs or overexpression construct, 7.5 μg of psPAX2 and 3 μg pMD2.G plasmids. The supernatant containing viral particles was harvested twice at 48h and 72h after transfection, then filtered through Millex-GP Filter Unit (0.45 μm pore size, Millipore). and concentrated by ultracentrifugation (24,000 rpm, 2 h, 4 °C). Concentrated virus was used to transduce HSPCs in the presence of 8 μg/mL polybrene (Millipore) by spinfection (2,000 rpm, 1.5 h, RT). Transduced cells were sorted based on GFP expression by fluorescence activated cell sorting (FACS) and subsequent to erythroid differentiation and functional analyses.

RNA isolation and RT-qPCR

RNA was collected from cultured cells using the Total RNA Purification Micro Kit (Norgen Biotek) with DNase I treatment according to the manufacturer's protocol. The cDNA synthesis was carried out using PrimeScript RT Master Mix (TaKaRa) according to the manufacturer's protocol. RT-qPCR was run on the CFX96 or CFX384 Real Time System (BioRad) using iQ SYBR Green Supermix (BioRad) following kit instructions. The relative expression of different sets of genes was quantified to *ACTB* mRNA, with control samples serving as the reference. Primer pairs for RT-qPCR are listed by gene in **Supplementary Table 11**.

Cell staining for Flow cytometry analysis

To assess erythroid differentiation, transduced or genome-edited HSPCs were collected at the indicated time points. Following a DPBS wash, cells were stained on ice for 30 min in FACS buffer (DPBS + 0.1% BSA) with anti-CD71_BV421 (BioLegend) and anti-CD235a_APC-Cy7 (BioLegend). Cells were then washed and resuspended in FACS buffer for flow cytometry analysis. For BACH2 intracellular staining, 3×10^5 human primary HSPCs were collected and washed with DPBS, then fixed in 4% paraformaldehyde (PFA) in PBS for 15 min. Cells were permeabilized with 0.2% Tween-20 in PBS for 10 min and stained with BACH2-PE (BioLegend) in 0.2% Tween-20 in PBS for 1 h at room temperature. Cells were subsequently washed and resuspended in FACS buffer for flow cytometry analysis. To quantify F-cell frequency in transduced or genome edited HSPCs undergoing erythroid differentiation, cells were fixed in 4% paraformaldehyde (PFA) in PBS for 15 min, permeabilized with 0.2% Tween-20 in PBS for 10 min, and stained with anti-HbF-PE (BD Biosciences) or anti-HbF-APC (invitrogen) in PBS containing 0.1% BSA for 1 h at room temperature. Cells were subsequently washed and

resuspended in FACS buffer for flow cytometry analysis. For each sample, 30,000-50,000 events were acquired on a BD LSRFortessa (BD Biosciences), and analyzed using FlowJo software (v10.8.1, BD Biosciences).

Lentiviral reporter assays

The lentiviral reporter constructs were used to deliver enhancer elements upstream of a minimal promoter (TATA box) driving a reporter eGFP[83]. Here we cloned non-risk allele and risk factor rs1010473 or rs1010474 upstream of minimal promoter drivened GFP. Lentiviral supernatants produced by transient transfection in HEK293T were concentrated and titrated by qPCR. Provirus copy numbers per cell were quantified using the Lenti Integration Site Analysis Kit (Takara). Titered lentivirus was then used to transduce primary human HSPCs and subjected to in vitro erythroid differentiation. GFP expression was measured by flow cytometry.

Co-immunoprecipitation and western blotting

HEK293T or K562 cells (2×10^7) expressing Flag-NRF2, HA-BACH2, or their truncation constructs were harvested and lysed in 1 mL lysis buffer (50 mM Tris pH 7.4, 150 mM NaCl, 0.05% Igepal, 0.5% NP-40, 0.5 mM PMSF, and protease inhibitor cocktail), followed by two 10-s pulses of sonication. Lysates were pre-cleared with 15 μL protein G beads (Invitrogen) for 30 min at 4 °C. Pre-cleared lysates were incubated with 5 μg anti-Flag M2 antibody (Sigma) or isotype IgG control (Santa Cruz) for 3-4 h at 4 °C, and then incubated with protein G beads for an additional 1 h. Beads were washed four times with wash buffer (50 mM Tris pH 7.4, 300 mM NaCl, 0.5% NP-40, 0.5 mM PMSF, and protease inhibitor cocktail). Protein complexes were eluted by boiling in 50 μL 1× Laemmli Sample Buffer (Bio-Rad) at 100 °C for 10 min and chilled on ice for 5 min. For western blotting, denatured samples were resolved on 4–15% Mini-PROTEAN® TGX™ precast gels (Bio-Rad) and transferred to PVDF membranes using the Trans-Blot Turbo system (Bio-Rad). Membranes were blocked in 3% BSA in TBST (TBS + 0.1% Tween-20) for 30 min at room temperature, then incubated with primary antibodies overnight at 4 °C. After washing, membranes were incubated with HRP-conjugated secondary antibodies and developed using ECL substrate (Bio-Rad). Signals were visualized using a ChemiDoc imaging system (Bio-Rad).

Native PAGE

For *in vitro* analysis of the NRF2-BACH2 complex, recombinant BACH2, NRF2, and MafK proteins were mixed under the indicated conditions in binding buffer (100 mM HEPES, pH 7.5; 200 mM KCl; 10 mM $MgCl_2$; 10 mM DTT) and incubated at room temperature for 25 min. Samples were then combined with 6× Native Sample Buffer (600 mM Tris-HCl, 50% glycerol, 0.02% bromophenol blue) and directly loaded onto 4-20% Mini-PROTEAN® TGX™ precast gels (Bio-Rad) and run in Tris/Glycine electrophoresis buffer (Bio-Rad). Gels were subjected to standard western blotting procedures, and signals were detected using a ChemiDoc imaging system (Bio-Rad).

Chromatin Immunoprecipitation (ChIP)

ChIP was performed using chromatin prepared from 5×10^6 differentiated erythroid progenitor cells derived from primary human $CD34^+$ cells. On day 8 of erythroid differentiation, cells were cross-linked with 1% formaldehyde (Pierce Life Technologies, 28906) and quenched with glycine. Chromatin was processed using the truChIP Chromatin Shearing Kit (Covaris, 520127) according to the manufacturer's protocol. Lysates were sonicated with an E220 sonicator (Covaris, 500239) to generate DNA fragments of 300–500 bp. Sheared chromatin was precleared with 15 μL Dynabeads Protein G (Invitrogen) supplemented with 100 μg BSA and 100 μg ssDNA. Precleared lysates were incubated overnight at 4 °C with 5 μg anti-NRF2 antibody (Abcam, ab62352) or normal rabbit IgG control (CST, 2729S). Beads were sequentially washed with 600 μL lysis buffer, 600 μL high-salt wash buffer (1% Triton X-100, 0.1% sodium deoxycholate, 50 mM Tris-HCl pH 8.0, 0.5 M NaCl, 5 mM EDTA), 600 μL LiCl immune complex wash buffer (0.25 M LiCl, 0.5% Igepal, 0.5% sodium deoxycholate, 10 mM Tris-HCl pH 8.0, 1 mM EDTA), followed by two washes with 600 μL TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA) at 4 °C. The complexes were eluted in 200 μL freshly prepared elution buffer (1% SDS, 0.1 M $NaHCO_3$) with rotation at room temperature for 15 min. Reverse cross-linking was performed by adding 8 μL 5 M NaCl and incubating at 65 °C for 4 h, followed by the addition of 4 μL 0.5 M EDTA and 10 μL proteinase K (10 mg/mL) and incubation at 55 °C for 2 h. DNA was purified by phenol/chloroform extraction and ethanol precipitation with 20 μg yeast tRNA as a carrier. The DNA pellets were dissolved in 50 μL ddH_2O for qRT-PCR analysis. Primer sequences are provided in **Supplementary Table 11**.

Electrophoretic Mobility Shift Assay (EMSA)

DIG-labeled DNA probes were annealed to antisense oligonucleotides by heating at 100 °C for 5 min followed by slow cooling to room temperature. His-tagged NRF2 protein was expressed in K562 cells and purified using the His60 Ni Gravity Column Purification Kit (TaKaRa) according to the manufacturer's instructions. Equal amounts of purified NRF2 protein were incubated with 1uM the indicated DIG-labeled DNA probes in binding buffer (100 mM HEPES pH 7.5, 200 mM KCl, 10 mM $MgCl_2$, 10 mM DTT) for 25 min at room temperature. Binding reactions were resolved on a 6% TBE gel (invitrogen) in 0.5X TBE buffer, transferred to a nylon membrane, and incubated with anti-digoxigenin antibody following the DIG Northern Starter Kit (Roche) protocol. Signals were visualized using a ChemiDoc imaging system (Bio-Rad). DNA Probe sequences are provided in **Supplementary Table 11**.

Immunofluorescence and RNA In Situ hybridization

Immunofluorescence was performed as described previously with minor modifications. Briefly, 3×10^5 cells were collected, washed once with DPBS, and resuspended in 0.1% BSA in DPBS. Cells were cytospun onto poly-D-lysine–coated German coverslips (Neuvitro Corporation) at 300 rpm for 4 min and fixed with 4% paraformaldehyde (Electron Microscopy Sciences, 15714) for 15 min at room temperature. Fixed cells were permeabilized with 0.5% Triton X-100 for 5 min on ice and blocked with 1% BSA in PBS for 30 min at room temperature. Cells were then incubated overnight at 4 °C with mouse anti-NRF2 (Santa Cruz, sc-365949) and rabbit anti-BACH2 (Cell Signaling Technology, 80775) antibodies (1:50 in 1% BSA). After primary incubation, AF647-conjugated goat anti-mouse IgG and AF594-conjugated goat anti-rabbit IgG secondary antibodies (1:1000 in 1% BSA) were applied for 1 h at room temperature in the dark. Nuclei were counterstained with DAPI for 1 min at room temperature. Coverslips were mounted on glass slides and sealed. Images were acquired using a Leica TCS SP8 confocal microscope.

For co-staining of the primary transcripts of *HBG1/2* (*Pri-HBG1/2*) and NRF2, RNA FISH was performed first, followed by immunofluorescence. Briefly, cells were fixed with 4% paraformaldehyde for 15 min, permeabilized in 0.5% Triton X-100 containing 2 mM Ribonucleoside Vanadyl Complex (NEB), and dehydrated in 75% ethanol overnight. Cells were then incubated with denatured Dig-labeled FISH probes in hybridization buffer (50% formamide in 2× SSC) at 50 °C overnight. After hybridization, cells were washed twice with Sal I buffer (50% formamide, 0.1× SSC, 0.1% SDS) and incubated with sheep anti-digoxigenin (Roche, 11333089001) for 1 h at room temperature. Following two washes with Sal II buffer (2× SSC containing 8% formamide), cells were incubated with mouse anti-NRF2 (Santa Cruz, sc-365949) overnight at 4 °C. After two additional washes with Sal II buffer, cells were incubated with Cy3-conjugated donkey anti-sheep IgG and AF647-conjugated goat anti-mouse IgG secondary antibodies for 1 h at room temperature in the dark. Nuclei were counterstained with DAPI. Coverslips were mounted on glass slides and sealed. Images were acquired using a Leica TCS SP8 confocal microscope.

DIG-labeled *Pri-HBG1/2* RNA probes were synthesized using the DIG RNA Labeling Mix (Roche, 11277073910) and T7 RNA Polymerase (Promega) according to the manufacturer's instructions. DNA templates containing a T7 promoter were generated by PCR using *HBG1/2* intron 2 primers (Table S13). In vitro transcription (IVT) was performed at 37 °C for 2 h using T7 RNA polymerase in the presence of the DIG RNA Labeling Mix, followed by DNase I treatment to remove the DNA template. RNA probes were purified using Monarch RNA Cleanup Columns (NEB) and analyzed by RNA gel electrophoresis to assess probe integrity and fragment size.

Measurement of NRF2 and pri-HBG1/2 transcription foci size

For quantifying NRF2 or *pri-HBG1/2* transcription foci in K562 cells, HSPCs, and differentiated erythroid cells, cells were subjected to IF or IF-FISH as described above. Images were acquired on a Leica TCS SP8 confocal microscope with z-stack acquisition. Raw images were imported into Fiji/ImageJ for analysis. Each channel was separated, and a maximum-intensity z-projection was generated. When necessary, background subtraction and linear contrast adjustment were applied uniformly across samples. NRF2 and *pri-HBG1/2* signals were then segmented using an optimal threshold. Regions of interest (ROIs) were defined based on the DAPI signal.

Individual cells were analyzed using the "Analyze Particles" tool in Fiji/ImageJ (default settings: size 0-infinity, circularity 0.00-1.00). A total of 20-30 cells were quantified per condition, and the results were plotted in Prism 10.

Colocalization analysis of NRF2 and *pri-HBG1/2*

Colocalization between NRF2 and *pri-HBG1/2* was quantified using Manders' overlap coefficients (M1 and M2) implemented in the JaCoP plugin in Fiji/ImageJ. Images were acquired on a Leica TCS SP8 confocal microscope with z-stack acquisition. Raw images were imported into Fiji/ImageJ, channels were separated, and a maximum-intensity z-projection was generated for analysis. When necessary, background subtraction and linear contrast adjustment were applied uniformly across samples. Individual cells were segmented based on DAPI. Manders' coefficients M1 and M2 were calculated following the optimal threshold to determine the fraction of signal from channel A overlapping with channel B, and vice versa. A total of 20–30 cells were quantified per condition, and the results were plotted in Prism 10.

RNA-seq data processing

All RNA-seq data were aligned to the human reference genome hg38 using STAR[84] with the parameters --runMode alignReads --chimSegmentMin 20. The resulting alignments were sorted, and unmapped reads were removed using SAMtools with option -F 516. Gene-level read counts were then generated by featureCounts[85] using the human reference gene annotation. Differentially expressed genes (DEGs) were identified with DESeq2[86] using the Wald test, applying thresholds of $p < 0.05$ and fold change > 1.4 (**Supplementary Table 10**)

CUT&RUN and library preparation and sequencing

Cleavage Under Targets and Release Using Nuclease (CUT&RUN) experiments were carried out using CUTANA™ ChIC/CUT&RUN Kit, with slight modifications as described previously[87]. Briefly, ~500,000 cells were washed in a buffer containing digitonin (0.01% digitonin) and bound to 10 μl of activated Concanavalin A beads. Bead-bound cells were incubated with either BACH2 primary antibody and NRF2 primary antibody (sigma, SAB5700720) overnight at 4C. After washing, cells were incubated with pA/G-MNase for 1h at 4C and targeted chromatin digestion initiated by the addition of 100 mM $CaCl_2$ and allowed to proceed for 30 min at 0°C, at which time stop buffer containing E. coli-Spike-in DNA was added. Incubate the reaction at 37 for 30min, to release chromatin fragments and then purified by phenol-chloroform extraction followed by ethanol precipitation.

CUT&RUN DNA library preparation were carried out using The NEBNext Ultra II DNA Library Prep Kit as described previously[18]. Briefly, 15ng of CUT&RUN DNA were treated with endprep module at 20 °C for 30 min and 50 °C for 1 h to reduce the melting of short DNA. Ligation was performed by adding 1 pmol of NEB adapter and ligation mix and incubated at 20 °C for 15 min. To clean up the reaction, add 1.75× volume of SPRIselect beads (Beckman Coulter) to capture short ligation products. PCR amplification was performed for 12 cycles. The resulting libraries were purified with 1.2× volume of SPRIselect beads then analyzed and quantified by Qubit and Tapestation.Libraries with different indexes were pooled, and **Illumina** paired-end sequencing was performed using Nextseq platform with NextSeq-1000 P2 Kit (75 cycles) (2 × 50bp, 6-bp index).

CUT&RUN data processing and peak calling

All CUT&RUN sequencing data in this study were processed using CUT&RUNTool (https://bitbucket.org/qzhudfci/cutruntools/src/default/)[88], which performs sequential steps of read trimming, alignment, peak calling, and motif analysis. First, paired-end sequencing reads were trimmed with Trimmomatic[89] to remove adapter sequences from the 3′ end using the parameters: ILLUMINACLIP:2:15:4:4: true LEADING:20 TRAILING:20 SLIDINGWINDOW:4:15. To further eliminate residual adapter sequences (up to 6 bp) not removed by Trimmomatic, an additional trimming step was performed using Kseq. The cleaned reads were then aligned to the human reference genome hg38 by Bowtie2[90] with parameters --dovetail and --phred33. The resulting BAM files were sorted, indexed, and duplicate reads were marked using Picard MarkDuplicates. Finally, unmapped, unmated, and duplicate reads were removed with SAMtools[91], generating high-quality alignments for downstream analyses. MACS[75] version 2.2 was used to call peaks from the BAM file with narrowPeak setting, with *q*-value cutoff 0.01.

CUN&RUN normalizing to E.coli Spike-in DNA

To ensure comparability across CUT&RUN data, we normalized the data using *E. coli* spike-in DNA. In addition to aligning reads to the human reference genome hg38, sequencing reads were separately aligned to the *E. coli* K12 MG1655 reference, and non-uniquely mapped reads were removed. For each CUT&RUN dataset, the number of uniquely aligned *E. coli* reads was quantified and normalized to the total number of uniquely aligned human hg38 reads. A normalization factor was then calculated so that the *E. coli* spike-in signal was equalized across all CUT&RUN data. This single scalar normalization ratio was applied using the --scaleFactor option in the deepTools[92] bamCoverage function to generate normalized bigWig files for visualization and comparative analyses.

In silico base-perturbation analysis of transcription factor binding

To evaluate the contribution of individual nucleotides to the binding affinity of BACH2, NRF2, and NFE2, we generated a total of 5 variant sequences, each containing a single in silico base substitution relative to the reference sequence GTTTGCCTTGTCAAGGCTAT. These sequences were analyzed using FIMO[93] for motif scanning with a significance threshold of $p < 0.05$. For each sequence, we extracted the log-likelihood ratio score reported by FIMO for potential motif matches at each position. Comparison of these scores allowed us to estimate the effect of single-base perturbations on motif binding affinity.

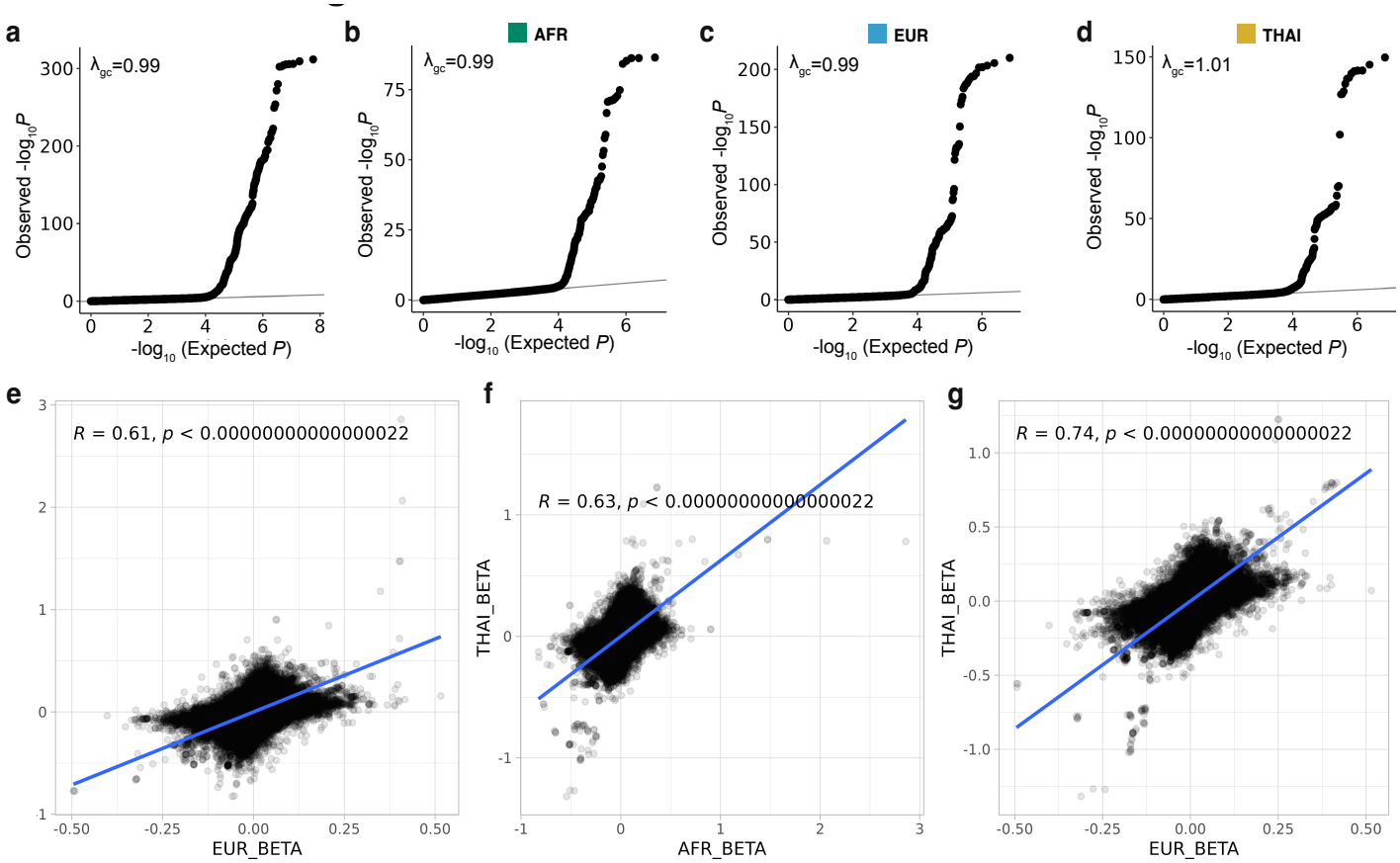

Figure S1: META GWAS and MAMA QC and concordance across ancestry. (a-d) Quantile–Quantile (QQ) plots for meta-analysis and ancestry-specific analyses from MAMA (Multi-Ancestry Meta-Analysis). LambdaGC values (λ_{gc}) are annotated in each plot. **(e-f)** Concordance of SNP effect sizes across ancestry groups in the multi-ancestry meta-analysis (MAMA). Correlations (R value) are indicated on each plot.

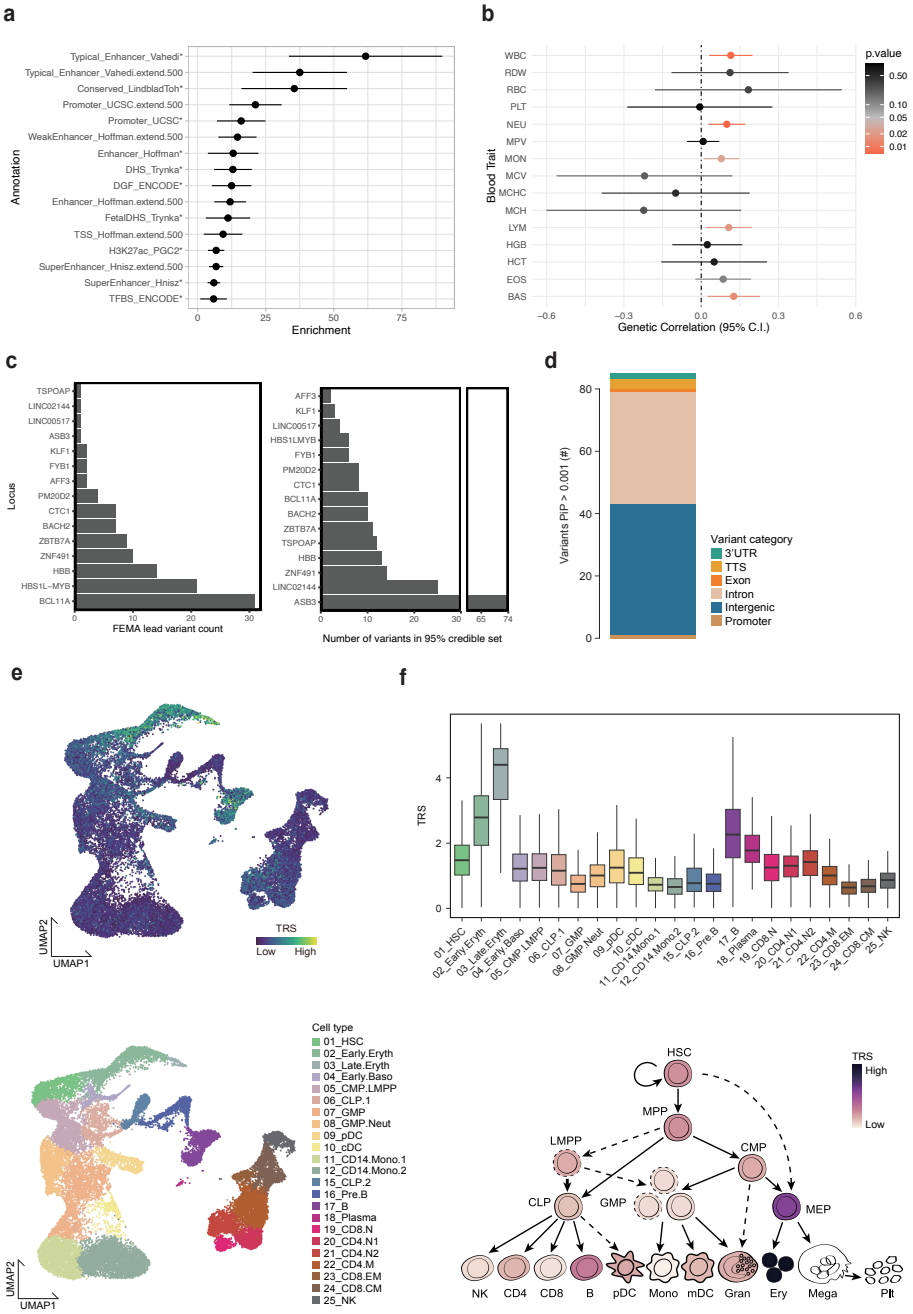

Figure S2: Heritability, genetic correlation, conditional and fine-mapping analysis. (a) Heritability enrichments from LDAK for 64 functional categories in the BLD model. **(b)** Genetic correlations between HbF with various blood cell traits. **(c)** Conditionally independent analysis revealed a number of potential lead variants per locus, and after fine-mapping, 95% credible sets are shown. **(d)** Functional consequences of the fine-mapped variants are shown. **(e)** UMAP projection SCAVENGE cell-stage enrichment results from single cell ATAC-seq data. **(f)** Cell type specific representation of SCAVENGE cell-stage enrichment results from single cell ATAC-seq data.

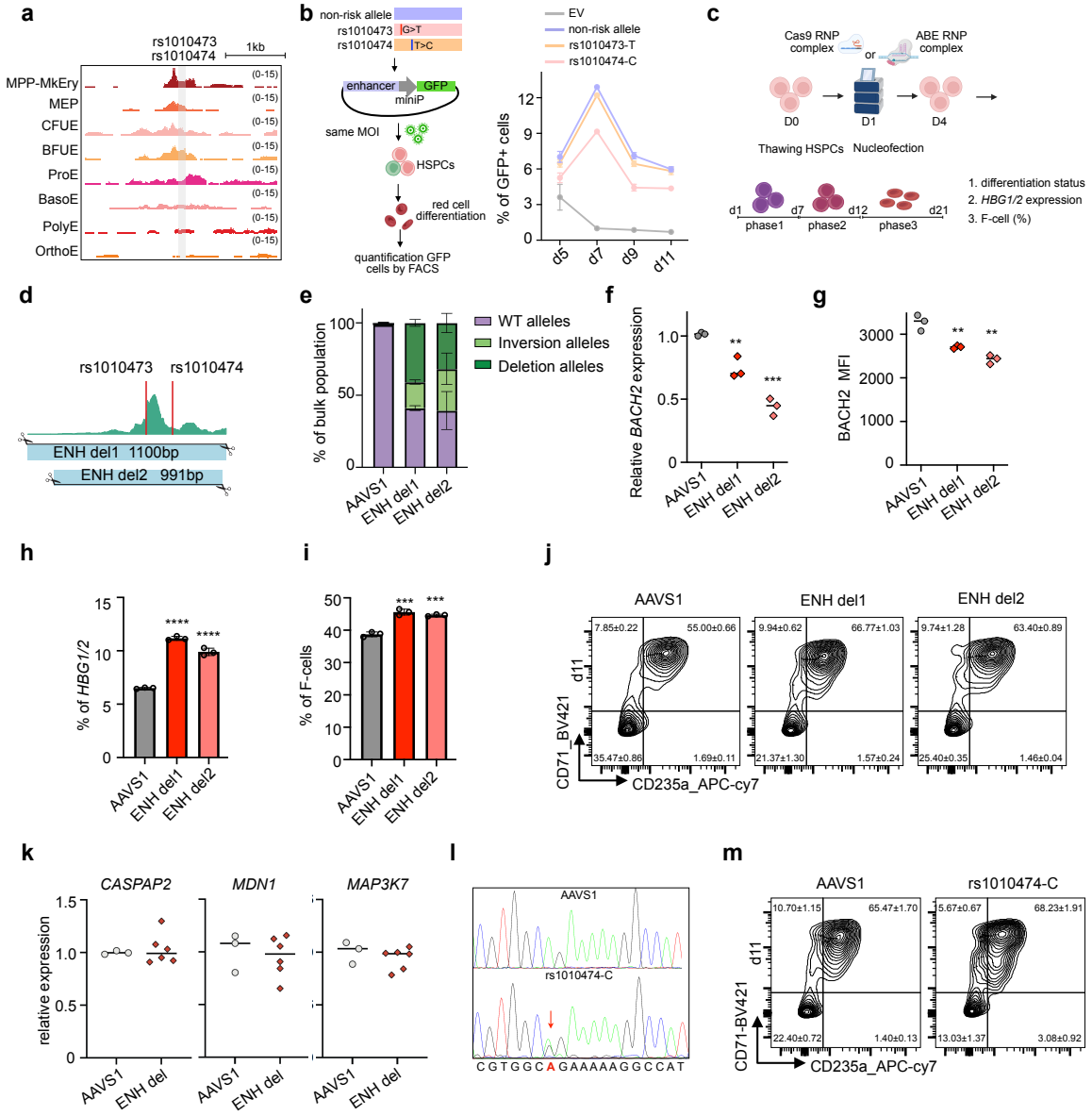

Figure S3: Variant rs1010474 and its harboring regulatory elements regulate BACH2 and HbF expression. (a) ATAC-seq profiles at the *BACH2* locus during human erythroid differentiation. Accessible chromatin (0–15 scale) are shown across erythroid stages from progenitors (MPP-MkEry, MEP, CFUE, BFUE) to differentiated cells (ProE, BasoE, PolyE, OrthoE). HbF-associated variants rs1010473 and rs1010474 are indicated in gray shading. **(b)** Lentiviral reporter assay for *BACH2* enhancer variants during human erythroid differentiation. *BACH2* regulatory elements harboring the non-risk allele or single variants (rs1010473-T, rs1010474-C) were cloned upstream of a minimal promoter driving GFP. Lentiviral constructs were delivered into HSPCs at equal multiplicity of infection (MOI), followed by erythroid differentiation (left). Reporter activity was quantified as the percentage of GFP⁺ cells by flow cytometry during erythroid differentiation (days 5–11) (right). **(c)** Schematic of experimental design for *BACH2* regulatory element depletion by CRISPR/Cas9 and rs1010474-C editing in human CD34⁺ HSPCs and then subjected to erythroid differentiation and functional evaluation. **(d)** Schematic of CRISPR/Cas9-mediated deletion of the *BACH2* regulatory element. Two pairs of guide RNAs (indicated by scissors) were designed to remove the open chromatin region (~0.6 kb, green peak). ENH del1 deleted 1,100 bp and ENH del2 deleted 991 bp within this region. The red line marks the positions of the variants. **(e)** Enhancer perturbations were quantified by qPCR and shown as a percentage of total alleles in the bulk population. Plotted are wild-type alleles, inversions, and deletions in *AAVS1* control and deletion of the *BACH2* enhancer by two pairs CRISPR/Cas9 guide RNAs (ENH del1 and ENH del2). **(f)** Relative *BACH2* expression in bulk human primary CD34⁺ HSPCs three days after deletion of *BACH2* enhancer, compared to *AAVS1* control. **(g)** Quantification of intracellular BACH2 MFI by flow cytometry in bulk human primary CD34⁺ HSPCs three days after *AAVS1* editing or *BACH2* enhancer deletion. **(h)** Proportion of *HBG1/2* expression on day 13 of erythroid differentiation in *AAVS1* or *BACH2* ENH-deleted HSPCs. **(i)** Frequency of F-cells on day 13 of erythroid differentiation in *AAVS1* or *BACH2* ENH-deleted HSPCs. **(j)** Erythroid differentiation status on days 11 of erythroid differentiation in *AAVS1* control or *BACH2* ENH-deleted HSPCs, as assessed by surface expression of CD71 and CD235a. **(k)** Relative expression of *MDN1*, *CASP8AP2*, and *MAP3K7* transcript in *AAVS1* control and *BACH2* Enhancer deleted (ENH del) HSPCs, as measured by RT-qPCR. **(l)** Editing efficiency of rs1010474-C on day 11 of erythroid differentiation, as determined by Sanger sequencing. **(m)** Erythroid differentiation status of *AAVS1* and rs1010474-C samples on day 11 of erythroid differentiation. All data are presented as the mean ± the standard deviation; significance is indicated as $*p < 0.05$, $**p<0.01$, $***p < 0.001$, $****p < 0.0001$ or n.s. (not significant).

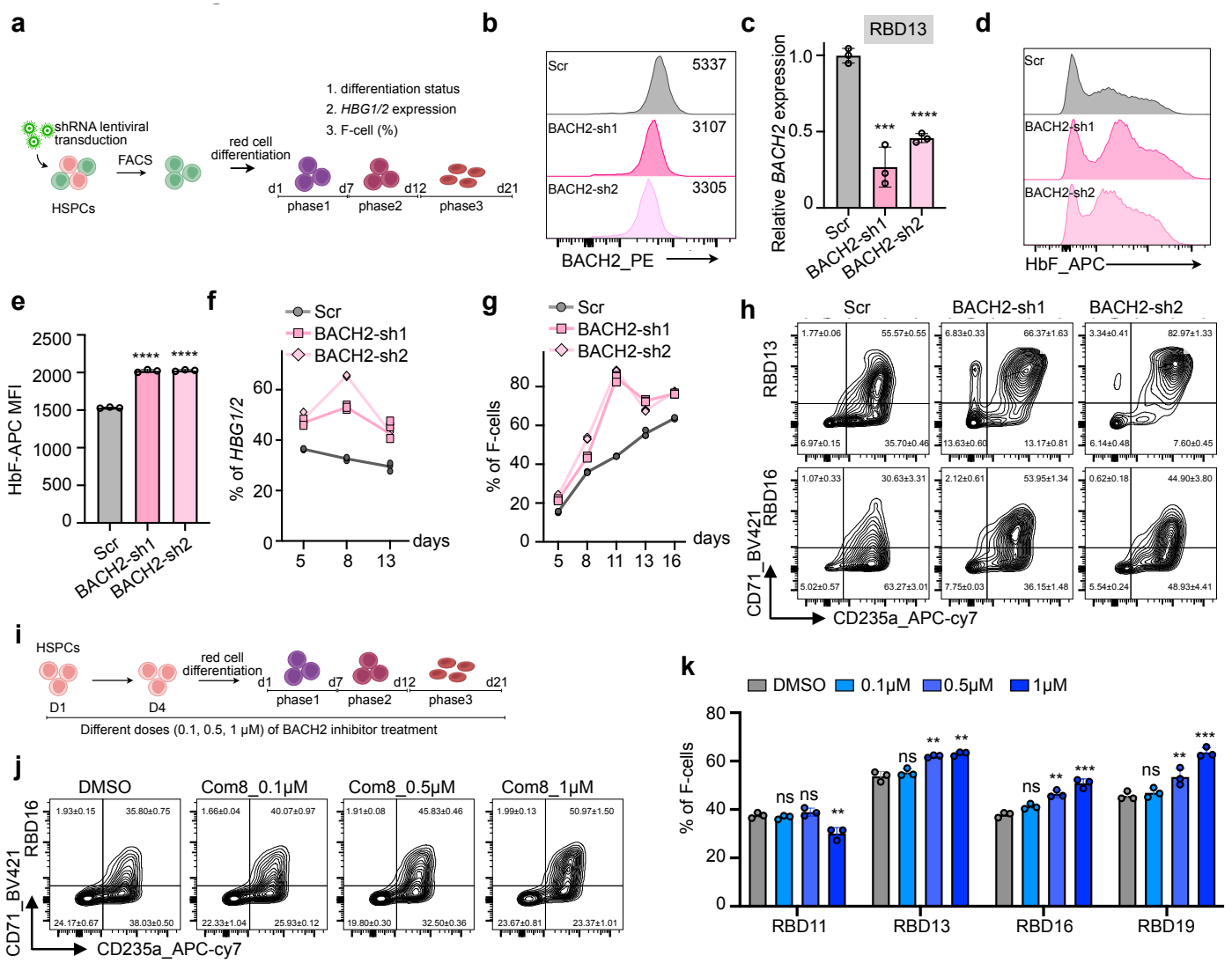

Figure S4: Knockdown or pharmacological inhibition of BACH2 promotes HbF expression. (a) Schematic of experimental design for BACH2 knockdown in human CD34⁺ HSPCs and erythroid differentiation. Human CD34⁺ HSPCs were transduced with lentiviruses expressing shRNA targeting BACH2 or scrambled (Scr) control. GFP⁺ cells were enriched by FACS and subjected to *in vitro* erythroid differentiation. Differentiation status, *HBG1/2* expression, and F-cell percentage were evaluated at the indicated time points. **(b)** Representative histogram of BACH2 intracellular flow cytometric in bulk human primary CD34⁺ HSPCs four days after BACH2 knockdown. **(c)** Relative *BACH2* expression on day 13 of erythroid differentiation (RBD13) in shRNA transduced HSPCs, compared to Scr control. **(d)** Representative histogram of fetal hemoglobin (HbF) intracellular flow cytometric on day 13 of erythroid differentiation in shRNA transduced HSPCs. **(e)** Quantification of MFI of HbF expression on day 13 of erythroid differentiation in shRNA transduced HSPCs. **(f)** Proportion of *HBG1/2* expression on indicated time (d5, d8 and d13) of erythroid differentiation in Scr control and *BACH2* shRNA transduce HSPCs. **(g)** Frequency of F-cells on indicated time (d5, d8, d11, d13 and d16) of erythroid differentiation in Scr control and *BACH2* shRNA transduce HSPCs. **(h)** Erythroid differentiation status on indicated time (d13 and d16) of erythroid differentiation in Scr control and *BACH2* shRNA transduce HSPCs. **(i)** Schematic of experimental design for BACH2 inhibitor (Comp8) treatment in HSPCs and during erythroid differentiation. **(j)** Erythroid differentiation status on day 16 of erythroid differentiation in DMSO control and BACH2 inhibitor treated HSPCs. **(k)** Frequency of F-cells at indicated time points (days 11, 13, 16, and 19) of erythroid differentiation in HSPCs treated with different concentrations of Comp8. All data are presented as the mean ± the standard deviation; significance is indicated as $**p<0.01$, $***p<0.001$, $****p<0.0001$ or n.s. (not significant).

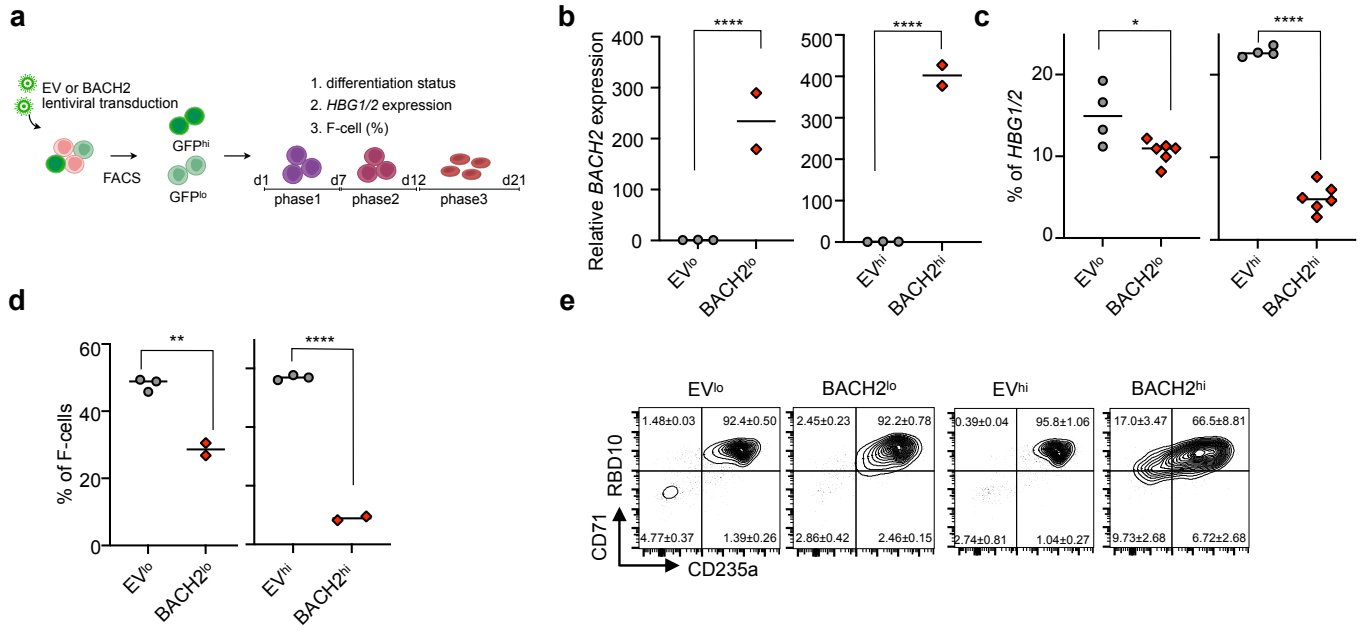

Figure S5: Overexpression of BACH2 inhibits HbF expression. (a) Schematic of experimental design for BACH2 overexpression in human CD34$^+$ HSPCs and erythroid differentiation. Human CD34$^+$ HSPCs were transduced with lentiviruses expressing either empty vector (EV) or BACH2, followed by sorting into GFP^{hi} and GFP^{lo} fractions. Enriched cells were subjected to *in vitro* erythroid cell differentiation. Differentiation status, *HBG1/2* expression, and F-cell percentage were evaluated at the indicated time points. **(b)** Relative *BACH2* expression on days 11 of erythroid differentiation in BACH2^{lo} and BACH2^{hi} overexpressing HSPCs, compared with the corresponding EV control. **(c)** Proportion of *HBG1/2* expression on day 11 of erythroid differentiation in BACH2^{lo} and BACH2^{hi} overexpressing HSPCs. **(d)** Frequency of F-cells on day 11 of erythroid differentiation in BACH2^{lo} and BACH2^{hi} overexpressing HSPCs, quantified by intracellular staining of fetal hemoglobin (HbF). **(e)** Erythroid differentiation status on day 10 of erythroid differentiation in EV control and BACH2^{lo} and BACH2^{hi} overexpressing HSPCs. All data are presented as the mean ± the standard deviation; significance is indicated as $*p < 0.05$, $**p<0.01$, $****p < 0.0001$.

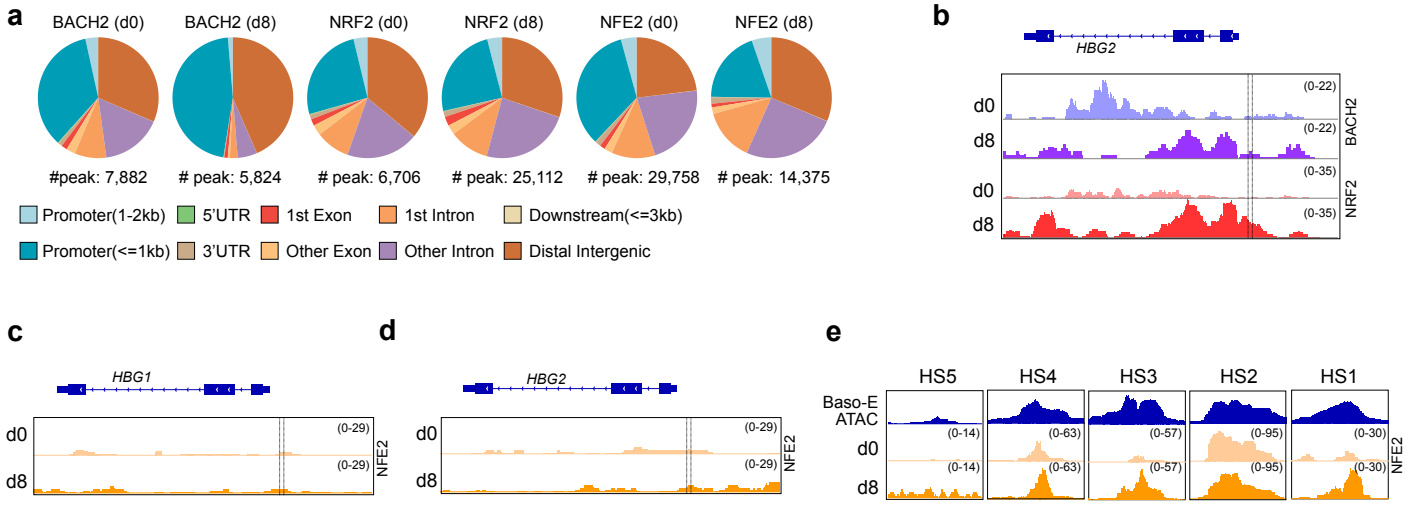

Figure S6: CUT&RUN profiling of BACH2 and NRF2 in human primary HSPCs and differentiated erythroid cells. (a) Genome-wide distribution of BACH2 and NRF2, NFE2 CUT&RUN peaks in human CD34⁺ HSPC (d0) and day 8 (d8) of erythroid differentiation. The number of peaks identified from CUT&RUN is indicated at the bottom. **(b)** BACH2 occupancy at the *HBG2* promoter at human CD34⁺ HSPC (d0) and d8 of erythroid differentiation (top), NRF2 occupancy at the *HBG2* promoter at human CD34⁺ HSPC (d0) and d8 of erythroid differentiation (bottom), as revealed by CUT&RUN (used throughout this figure). **(c)** NFE2 occupancy at the *HBG1* promoter at human CD34⁺ HSPC (d0) and d8 of erythroid differentiation. **(d)** NFE2 occupancy at the *HBG2* promoter at HSPC stage (d0) and d8 of erythroid differentiation. **(e)** NFE2 occupancy at multiple local control regions (LCR) of β-globin clusters at human CD34⁺ HSPC (d0) and d8 of erythroid differentiation.

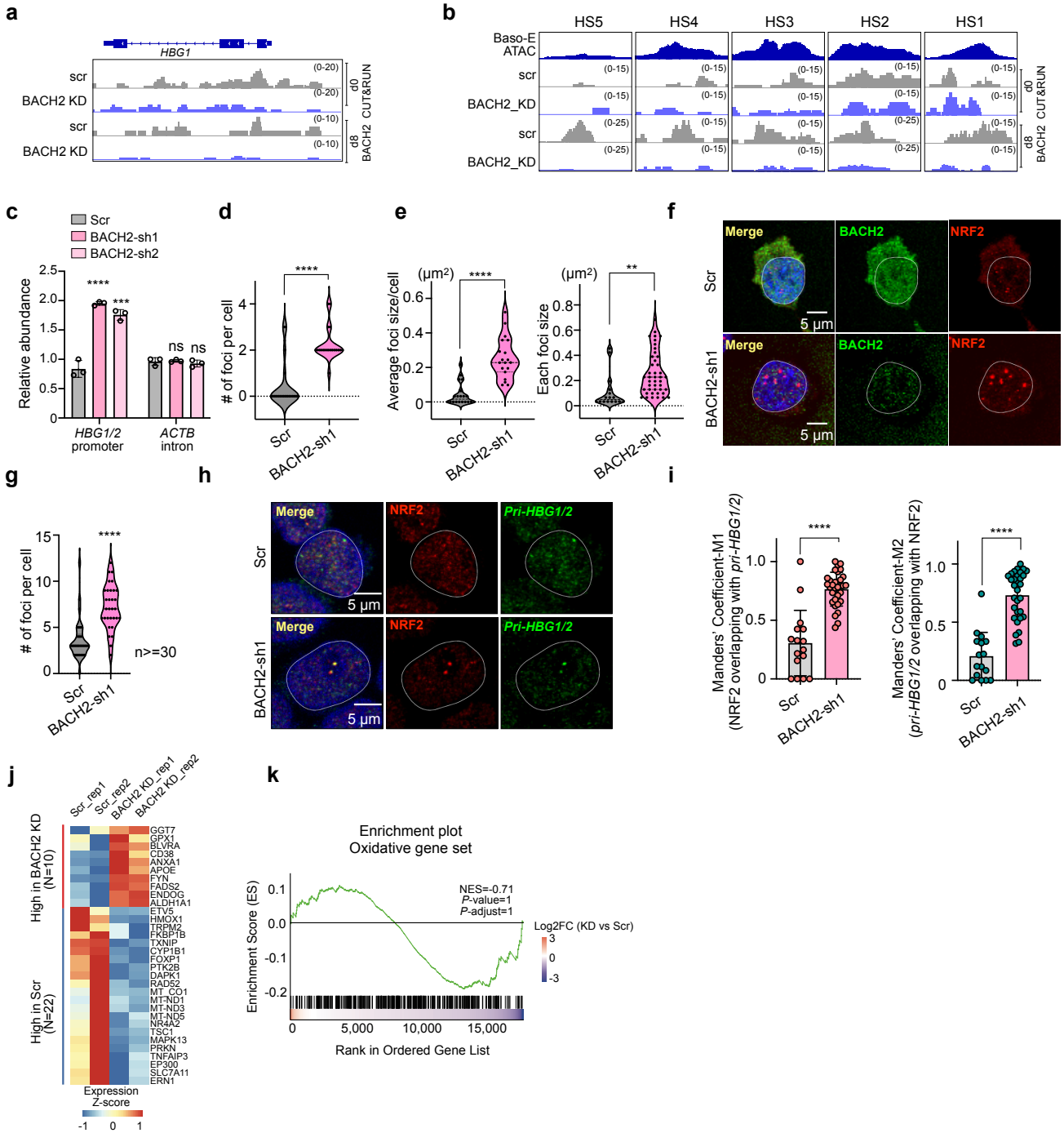

Figure S7: Knockdown of BACH2 enhances NRF2 accumulation at the γ-globin locus. (a) BACH2 occupancy at the *HBG1* promoter on the day 0 and day 8 of erythroid differentiation in Scr control and BACH2 knockdown HSPCs, as revealed by CUT&RUN. **(b)** BACH2 occupancy at multiple local control regions (LCR) of β-globin clusters on the day 0 and day 8 of erythroid differentiation in Scr control and BACH2 knockdown HSPCs, as revealed by CUT&RUN. **(c)** Relative NRF2 occupancy on *HBG1/2* promoter on day 8 of erythroid differentiation in Scr control and BACH2 shRNA transduced HSPCs, revealed by ChIP–qPCR. *ACTB* intron serves as a negative control. **(d)** Quantification of NRF2 foci numbers in Scr control and BACH2 knockdown HSPCs. Each dot represents one cell; n>=30 cells were analyzed. **(e)** Quantification of average NRF2 foci size per cell (left) and individual NRF2 foci size (right) in Scr control and BACH2 knockdown HSPCs. **(f)** Co-staining of BACH2 and NRF2 in Scr control and BACH2 knockdown K562 cells. **(g)** Quantification of NRF2 foci numbers in Scr control and BACH2 knockdown K562 cells. Each dot represents one cell; n ≥ 30 cells were analyzed. **(h)** Co-staining of NRF2 and HBG1/2 primary transcript (pri-HBG1/2) in Scr control and BACH2 knockdown HSPCs. **(i)** Quantification of colocalization efficiency between NRF2 and *pri-HBG1/2* transcript by Manders' coefficient. Each dot represents one cell. **(j)** Differential expression of oxidative stress-related genes in Scr control and BACH2 knockdown erythroid cells. **(k)** Gene enrichment analysis of oxidative stress-related genes in Scr control and BACH2 knockdown erythroid cells. All data are presented as the mean ± the standard deviation; significance is indicated as $*p < 0.05$, $**p<0.01$, $***p < 0.001$, $****p < 0.0001$ or n.s. (not significant).

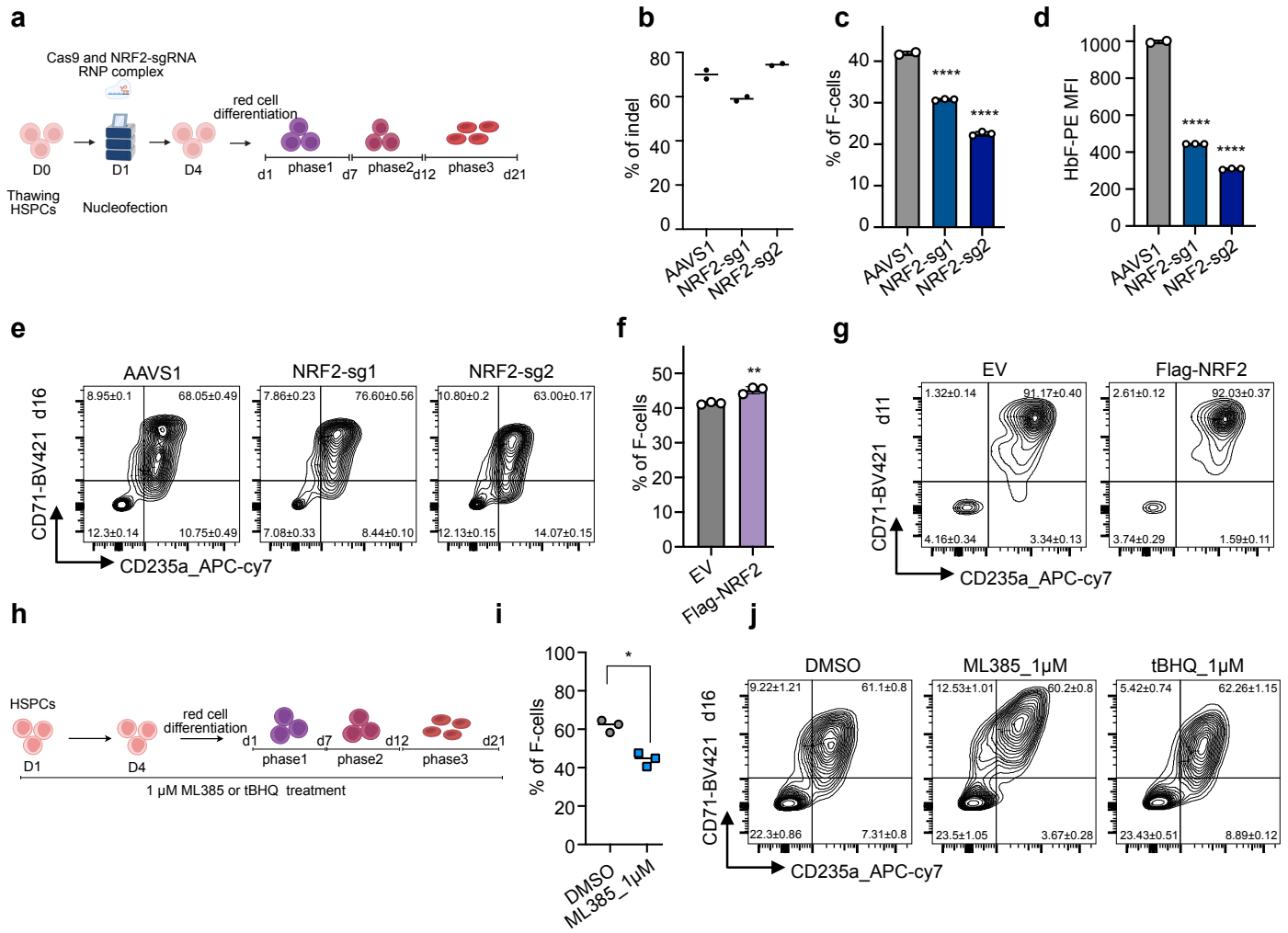

Figure S8: NRF2 functions as an activator to regulate HbF expression. (a) Schematic of experimental design for CRISPR/Cas9-mediated NRF2 knockdown in human CD34⁺ HSPCs followed by erythroid differentiation. **(b)** ICE analysis of CRISPR/Cas9-mediated NRF2 knockdown efficiency on day 11 of erythroid differentiation. **(c)** Frequency of F-cells on day 11 of erythroid differentiation in *AAVS1* control and NRF2 knockdown HSPCs. **(d)** Quantification of MFI of HbF on day 11 of erythroid differentiation in *AAVS1* control and NRF2 knockdown HSPCs. **(e)** Erythroid differentiation status on day 16 of erythroid differentiation in *AAVS1* control and NRF2 knockdown HSPCs. **(f)** Frequency of F-cells on day 11 of erythroid differentiation in EV control and Falg tagged NRF2 overexpressing HSPCs. **(g)** Erythroid differentiation status on day 11 of erythroid differentiation in EV control and NRF2 overexpressing HSPCs. **(h)** Schematic of experimental design for NRF2 inhibitor (ML385) and activator (tBHQ) treatment in human CD34⁺ HSPCs and during erythroid differentiation. **(i)** Frequency of F-cells on day 11 of erythroid differentiation in DMSO control and ML385 treated HSPCs. **(j)** Erythroid differentiation status on day 11 of erythroid differentiation in DMSO control, ML385-treated and tBHQ-treated HSPCs. All data are presented as the mean ± the standard deviation; significance is indicated as $*p < 0.05$, $**p<0.01$, $****p < 0.0001$ or n.s. (not significant).

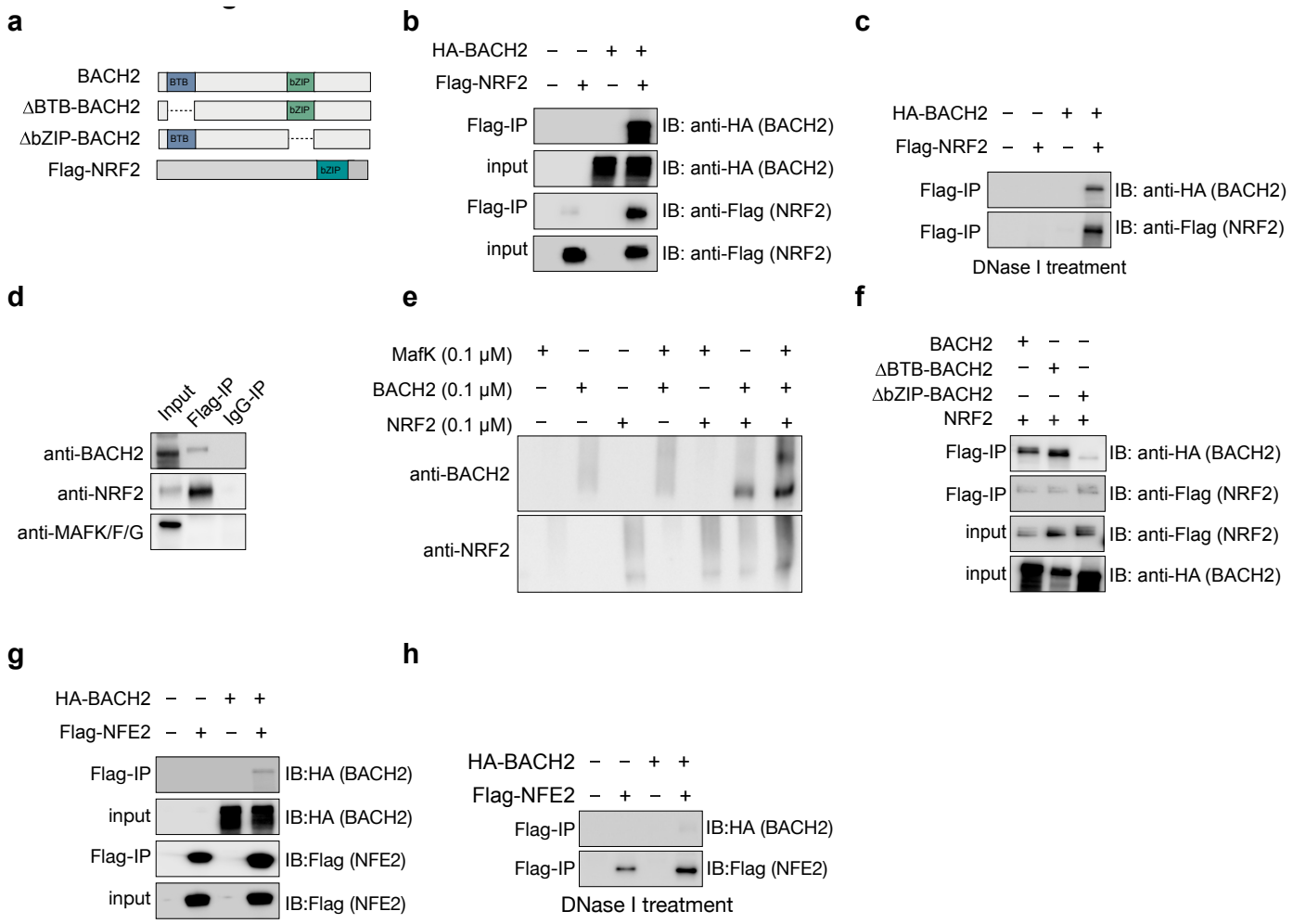

Figure S9: BACH2 specifically associates with NRF2, but not NFE2 (a) Schematic of truncated BACH2 and NRF2 proteins. All BACH2 truncations were fused to an N-terminal HA tag, and NRF2 truncations were fused to an N-terminal Flag tag. **(b)** NRF2 associates with BACH2, as shown by co-immunoprecipitation (Co-IP) in HEK293 cells overexpressing Flag-NRF2, using an anti-Flag antibody followed by immunoblotting with an anti-HA antibody against BACH2. **(c)** NRF2 associates with BACH2 upon DNase I treatment, as shown by Co-IP in HEK293 cells overexpressing Flag-NRF2 in the presence of DNase I, using an anti-Flag antibody followed by immunoblotting with an anti-HA antibody against BACH2. **(d)** NRF2 associates with endogenous BACH2 under stringent washing conditions, as shown by Co-IP in K562 cells overexpressing Flag-NRF2, using an anti-Flag antibody followed by immunoblotting with antibodies against endogenous BACH2 and MAFK/F/G. **(e)** Native PAGE analysis of recombinant BACH2 and NRF2 proteins showing complex formation with or without MAFK. **(f)** NRF2 interacts with BACH2 through the bZIP domain, as shown by Co-IP in HEK293 cells. **(g)** NFE2 does not associate with BACH2, as shown by Co-IP in HEK293 cells. **(h)** NRF2 does not associate with BACH2 after DNase I treatment, as shown by Co-IP in HEK293 cells.

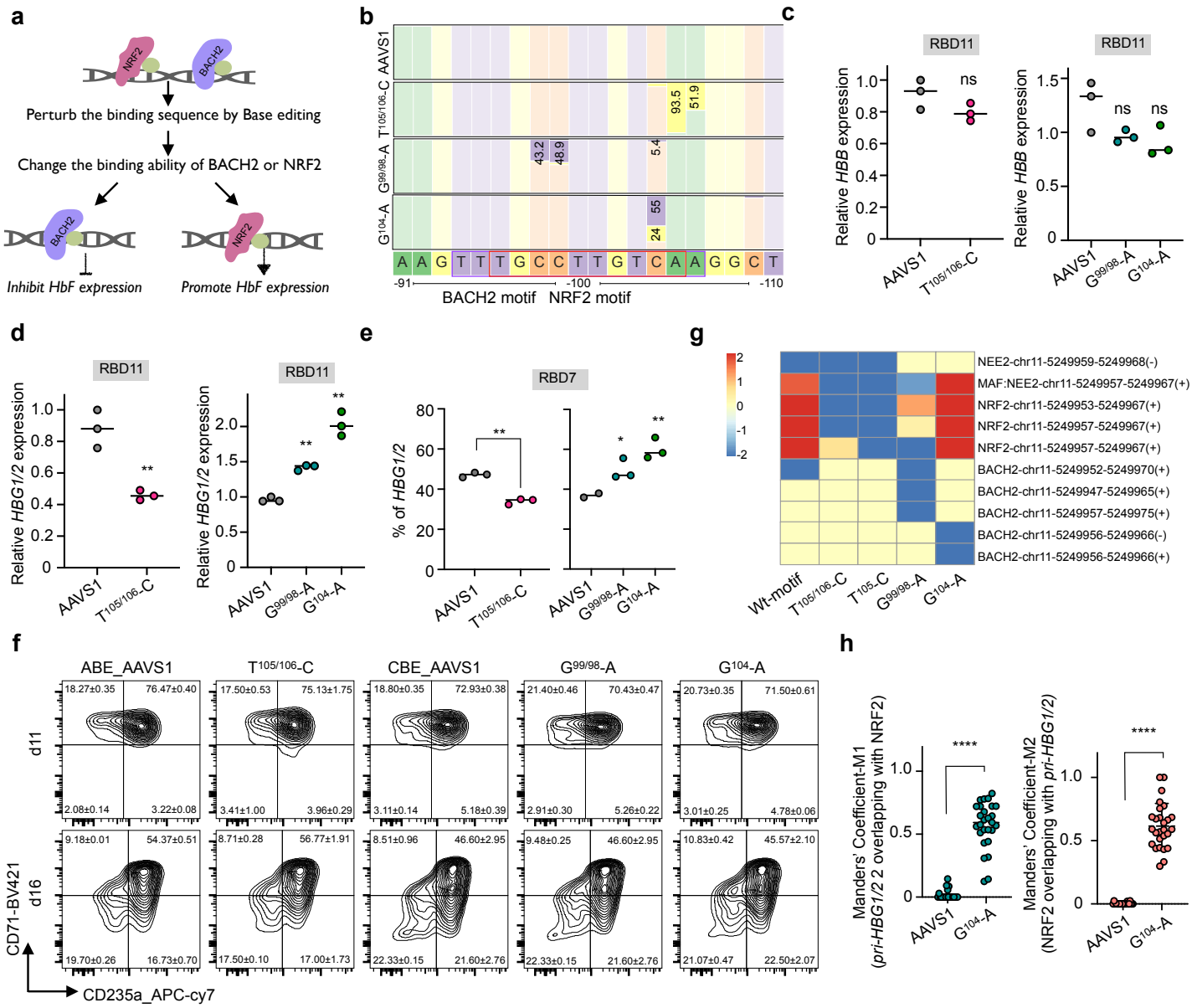

Figure S10: Editing of BACH2/NRF2 binding motifs in the γ-globin. (a) Proposed model of base editing at BACH2 and NRF2 binding motifs in the γ-globin promoter. Editing of the *cis* elements alters the binding preference of BACH2 or NRF2, leading to differential regulation of HbF expression. **(b)** Editing efficiency of BACH2 and NRF2 binding motifs in the γ-globin on day 11 of erythroid differentiation, revealed by CRISPResso analysis. **(c)** Relative *HBB* expression on day 11 of erythroid differentiation (RBD11) in BACH2 and NRF2 binding motifs edited HSPCs, compared to *AAVS1* control. **(d)** Relative *HBG1/2* expression on day 11 of erythroid differentiation (RBD11) in BACH2 and NRF2 binding motifs edited HSPCs, compared to *AAVS1* control. **(e)** Proportion of *HBG1/2* expression on day 7 of erythroid differentiated (RBD7) in BACH2 or NRF2 motif edited HSPCs, compared to *AAVS1* control. **(f)** Erythroid differentiation status of *AAVS1* control and BACH2 and NRF2 motif edited HSPCs on day 11 and 16 of erythroid differentiation. **(g)** BACH2 and NRF2 binding affinity by *in-silico* base-perturbation analysis. **(h)** Quantification of colocalization efficiency between NRF2 and pri-HBG1/2 transcript by Manders' coefficient in panel. Each dot represents one cell; n ≥ 30 cells were quantified. All data are presented as the mean ± the standard deviation; significance is indicated as $*p < 0.05$, $**p<0.01$or n.s. (not significant).