

various donor strains to promote Mpf and T6SS activation with *P. aeruginosa*, which is similar to what we observed in our RP4 mutant analysis (Fig. 1D). *P. aeruginosa* mutants defective in the attack-sensing pathway genes *tagT* and *pppA* (10) also exhibited greater conjugation efficiency as recipient strains (Fig. 3A). Examination of mixtures of T6SS⁺ *P. aeruginosa* and *E. coli* RP4⁺ donor cells by means of fluorescence microscopy revealed rounding and blebbing of *E. coli* cells—a response that is typical of T6SS-mediated bacterial killing (Fig. 3B). Thus, inhibition of the conjugative transfer of pPSV35 was likely due to killing of *E. coli* cells through a donor-directed T6SS attack by *P. aeruginosa*.

The fact that multiple secretion systems can induce a T6SS counterattack suggested that the signal initiating this response really is a generalized disruption of the *P. aeruginosa* membrane. Accordingly, we asked whether polymyxin B—an antibiotic known to disrupt Gram-negative bacterial membranes by binding the lipid A component of lipopolysaccharides (24–26)—could induce T6SS activity in *P. aeruginosa*. We used a *P. aeruginosa* strain carrying a ClpV1-GFP and fluorescent time-lapse microscopy to monitor T6SS organelle formation and dynamics (9, 10) after exposure to polymyxin B. Cells exhibited a sixfold increase in the average number of visible ClpV1-GFP foci per cell within 90 s of being spotted onto agar pads containing 20 µg/mL of polymyxin B (Fig. 4, A and C, and movie S1). After this increase in T6SS activity, most ClpV1-GFP foci disappeared over the next 3 min, with the remaining foci becoming nondynamic (Fig. 4A and movie S1). The loss of dynamics likely reflects consumption of intracellular adenosine 5'-triphosphate pools after prolonged exposure to polymyxin B intoxication. This increase in T6SS activity was not observed when cells were spotted onto agar pads lacking polymyxin B (Fig. 4, A and D, and movie S1). Additionally, this increase in ClpV1-GFP foci was not observed in *tagT* mutants even in the presence of polymyxin B (Fig. 4, A, E, and F; and movie S2), suggesting that the same attack-sensing pathway that senses T4SS and T6SS attacks is responding to this antibiotic and mediates activation of the T6SS.

These studies support a model in which the donor-directed T6SS attack response in *P. aeruginosa* likely involves detection of perturbations in the cell envelope caused by the invasive components of the T4SS conjugation machinery. T6SS may represent a type of bacterial “innate immune system” that can detect and attack invading infectious elements not by recognizing their molecular patterns [such as nucleic acid sequences, as do the clustered regularly interspaced short palindromic repeat (CRISPR) elements (27, 28); or unmethylated DNA, as do restriction enzymes (29)] but rather by recognizing “transfer-associated patterns” (TAPs), including membrane disruptions that occur during interactions with other cells deploying T6SS and

T4SS translocation machines. Broad-host-range conjugative elements represent infectious bacterial “diseases” that can cause metabolic stress on their newly acquired hosts and thus represent a fitness burden to bacterial populations unable to combat their acquisition. The donor-directed T6SS attack paradigm may represent a strategy for suppressing the movement of horizontally transferred genetic elements in bacterial populations regardless of their signature molecular patterns (such as nucleic acid chemical structures or primary sequences).

References and Notes

- M. E. Hibbing, C. Fuqua, M. R. Parsek, S. B. Peterson, *Nat. Rev. Microbiol.* **8**, 15–25 (2010).
- S. Pukatzki et al., *Proc. Natl. Acad. Sci. U.S.A.* **103**, 1528–1533 (2006).
- F. Boyer, G. Fichant, J. Berthod, Y. Vandenbrouck, I. Attree, *BMC Genomics* **10**, 104 (2009).
- P. G. Leiman et al., *Proc. Natl. Acad. Sci. U.S.A.* **106**, 4154–4159 (2009).
- M. Basler, M. Pilhofer, G. P. Henderson, G. J. Jensen, J. J. Mekalanos, *Nature* **483**, 182–186 (2012).
- T. G. Dong, B. T. Ho, D. R. Yoder-Himes, J. J. Mekalanos, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 2623–2628 (2013).
- R. D. Hood et al., *Cell Host Microbe* **7**, 25–37 (2010).
- J. D. Mougous et al., *Science* **312**, 1526–1530 (2006).
- M. Basler, J. J. Mekalanos, *Science* **337**, 815 (2012).
- M. Basler, B. T. Ho, J. J. Mekalanos, *Cell* **152**, 884–894 (2013).
- Z. Q. Luo, R. R. Isberg, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 841–846 (2004).
- W. Pansegrau et al., *J. Mol. Biol.* **239**, 623–663 (1994).
- P. J. Christie, J. P. Vogel, *Trends Microbiol.* **8**, 354–360 (2000).
- G. Schröder, E. Lanka, *Plasmid* **54**, 1–25 (2005).
- L. A. Giebelhaus et al., *J. Bacteriol.* **178**, 6378–6381 (1996).
- Z. Li, H. Hiasa, U. Kumar, R. J. DiGate, *J. Biol. Chem.* **272**, 19582–19587 (1997).
- J. Haase, R. Lurz, A. M. Grahn, D. H. Bamford, E. Lanka, *J. Bacteriol.* **177**, 4779–4791 (1995).
- D. Fernandez, G. M. Spudich, X. R. Zhou, P. J. Christie, *J. Bacteriol.* **178**, 3168–3176 (1996).
- M. Bayer et al., *J. Bacteriol.* **177**, 4279–4288 (1995).
- P. J. Langer, W. G. Shanabruch, G. C. Walker, *J. Bacteriol.* **145**, 1310–1316 (1981).
- Z. Zhong, D. Helinski, A. Toukdarian, *Plasmid* **54**, 48–56 (2005).
- A. Rietsch, I. Vallet-Gely, S. L. Dove, J. J. Mekalanos, E. Exs, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 8006–8011 (2005).
- R. Simon, U. Priefer, A. Pühler, *Nat. Biotechnol.* **1**, 784–791 (1983).
- D. C. Morrison, D. M. Jacobs, *Immunochimistry* **13**, 813–818 (1976).
- J. B. McPhee, S. Lewenza, R. E. Hancock, *Mol. Microbiol.* **50**, 205–217 (2003).
- J. V. Hankins, J. A. Madsen, D. K. Giles, J. S. Brodbelt, M. S. Trent, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 8722–8727 (2012).
- R. Barrangou et al., *Science* **315**, 1709–1712 (2007).
- P. Horvath, R. Barrangou, *Science* **327**, 167–170 (2010).
- T. Naito, K. Kusano, I. Kobayashi, *Science* **267**, 897–899 (1995).

Acknowledgments: Supporting movies and table can be found in the supplementary materials. This work was supported by National Institute of Allergy and Infectious Diseases grants AI-018045 and AI-26289 to J.J.M.

Supplementary Materials

www.sciencemag.org/content/342/6155/250/suppl/DC1
Materials and Methods

Table S1
References (30, 31)
Movies S1 and S2

25 July 2013; accepted 6 September 2013
10.1126/science.1243745

An Erythroid Enhancer of *BCL11A* Subject to Genetic Variation Determines Fetal Hemoglobin Level

Daniel E. Bauer,^{1,2,3} Sophia C. Kamran,^{3,4} Samuel Lessard,⁵ Jian Xu,^{1,3} Yuko Fujiwara,¹ Carrie Lin,¹ Zhen Shao,¹ Matthew C. Canver,³ Elenoe C. Smith,¹ Luca Pinello,⁶ Peter J. Sabo,^{7,8} Jeff Vierstra,^{7,8} Richard A. Voit,⁹ Guo-Cheng Yuan,^{6,10} Matthew H. Porteus,⁹ John A. Stamatoyannopoulos,^{7,8} Guillaume Lettre,⁵ Stuart H. Orkin^{1,2,3,4*}

Genome-wide association studies (GWASs) have ascertained numerous trait-associated common genetic variants, frequently localized to regulatory DNA. We found that common genetic variation at *BCL11A* associated with fetal hemoglobin (HbF) level lies in noncoding sequences decorated by an erythroid enhancer chromatin signature. Fine-mapping uncovers a motif-disrupting common variant associated with reduced transcription factor (TF) binding, modestly diminished *BCL11A* expression, and elevated HbF. The surrounding sequences function in vivo as a developmental stage-specific, lineage-restricted enhancer. Genome engineering reveals the enhancer is required in erythroid but not B-lymphoid cells for *BCL11A* expression. These findings illustrate how GWASs may expose functional variants of modest impact within causal elements essential for appropriate gene expression. We propose the GWAS-marked *BCL11A* enhancer represents an attractive target for therapeutic genome engineering for the β -hemoglobinopathies.

Genome-wide association studies (GWASs) have identified numerous common single-nucleotide polymorphisms (SNPs) associated with human traits and diseases. However, advancing from genetic association to causal bi-

ologic process has been challenging (1). Recent genome-scale chromatin mapping studies have highlighted the enrichment of GWAS variants in regulatory DNA elements, suggesting many causal variants may affect gene regulation (2–6). GWASs

of HbF level have identified trait-associated variants at *BCL11A* (supplementary text) (7–12). The transcriptional repressor BCL11A has been validated as a direct regulator of HbF level (13–18). Although constitutive BCL11A deficiency results in embryonic lethality and impaired lymphocyte development (19, 20), erythroid-specific deficiency of BCL11A counteracts developmental silencing of embryonic and fetal globin genes and rescues the hematologic and pathologic features of sickle cell disease (SCD) in mouse models (17).

¹Division of Hematology/Oncology, Boston Children’s Hospital, Boston, MA 02115, USA. ²Department of Pediatric Oncology, Dana-Farber Cancer Institute, Boston, MA 02115, USA. ³Harvard Medical School, Boston, MA 02115, USA. ⁴Howard Hughes Medical Institute, Boston, MA 02115, USA. ⁵Montreal Heart Institute and Université Montréal, Montreal, Quebec H1T 1C8, Canada. ⁶Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, MA 02115, USA. ⁷Department of Genome Sciences, University of Washington, Seattle, WA 98195, USA. ⁸Department of Medicine, University of Washington, Seattle, WA 98195, USA. ⁹Department of Pediatrics, Stanford University, Palo Alto, CA 94304, USA. ¹⁰Harvard School of Public Health, Boston, MA 02115, USA.
*Corresponding author. E-mail: stuart_orkin@dfci.harvard.edu

To further understand how common genetic variation affects *BCL11A*, HbF level, and β -globin disorder severity, we compared the distribution of the HbF-associated SNPs at *BCL11A* with deoxyribonuclease I (DNase I) sensitivity, which is an indicator of chromatin state suggestive of regulatory potential. In primary human erythroblasts, three peaks of DNase I hypersensitivity were observed in intron-2, adjacent to and overlying the HbF-associated variants (Fig. 1A). We term these DNase I hypersensitive sites (DHSs) +62, +58, and +55 based on distance in kilobases from the transcription start site (TSS) of *BCL11A*. Brain and B-lymphocytes, two tissues that express high levels, and T-lymphocytes, which do not express BCL11A, showed distinct patterns of DNase I sensitivity at the *BCL11A* locus, with a paucity of hypersensitivity overlying the trait-associated SNPs (Fig. 1A and fig. S1).
Chromatin immunoprecipitation sequencing (ChIP-seq) demonstrated histone modifications with an enhancer signature overlying the trait-associated SNPs at *BCL11A* intron-2, including the presence of H3K4me1 and H3K27ac and absence of H3K4me3 and H3K27me3 marks

(Fig. 1A and fig. S1). The major erythroid TFs GATA1 and TAL1 also occupy this enhancer region. ChIP–quantitative polymerase chain reaction (PCR) confirmed three discrete peaks of GATA1 and TAL1 binding within *BCL11A* intron-2, each falling within an erythroid DHS (Fig. 1B). A common feature of distal regulatory elements is long-range interaction with cognate promoters. We evaluated the interactions between the *BCL11A* promoter and fragments across 250 kb of the *BCL11A* locus using a chromosome conformation capture assay. The greatest promoter interaction was observed within the region of intron-2 containing the trait-associated SNPs (Fig. 1C).
We hypothesized that the causal trait-associated SNPs could function by modulating critical cis-regulatory elements. Therefore, we performed extensive genotyping of SNPs within the three erythroid DHSs +62, +58, and +55 in 1263 DNA samples from the Cooperative Study of SCD (CSSCD) (21). We used 1178 individuals and 38 SNPs for association testing (fig. S2A). Analysis of common variants [minor allele frequency (MAF) > 1%] revealed that rs1427407 in DHS +62 had the strongest association to HbF

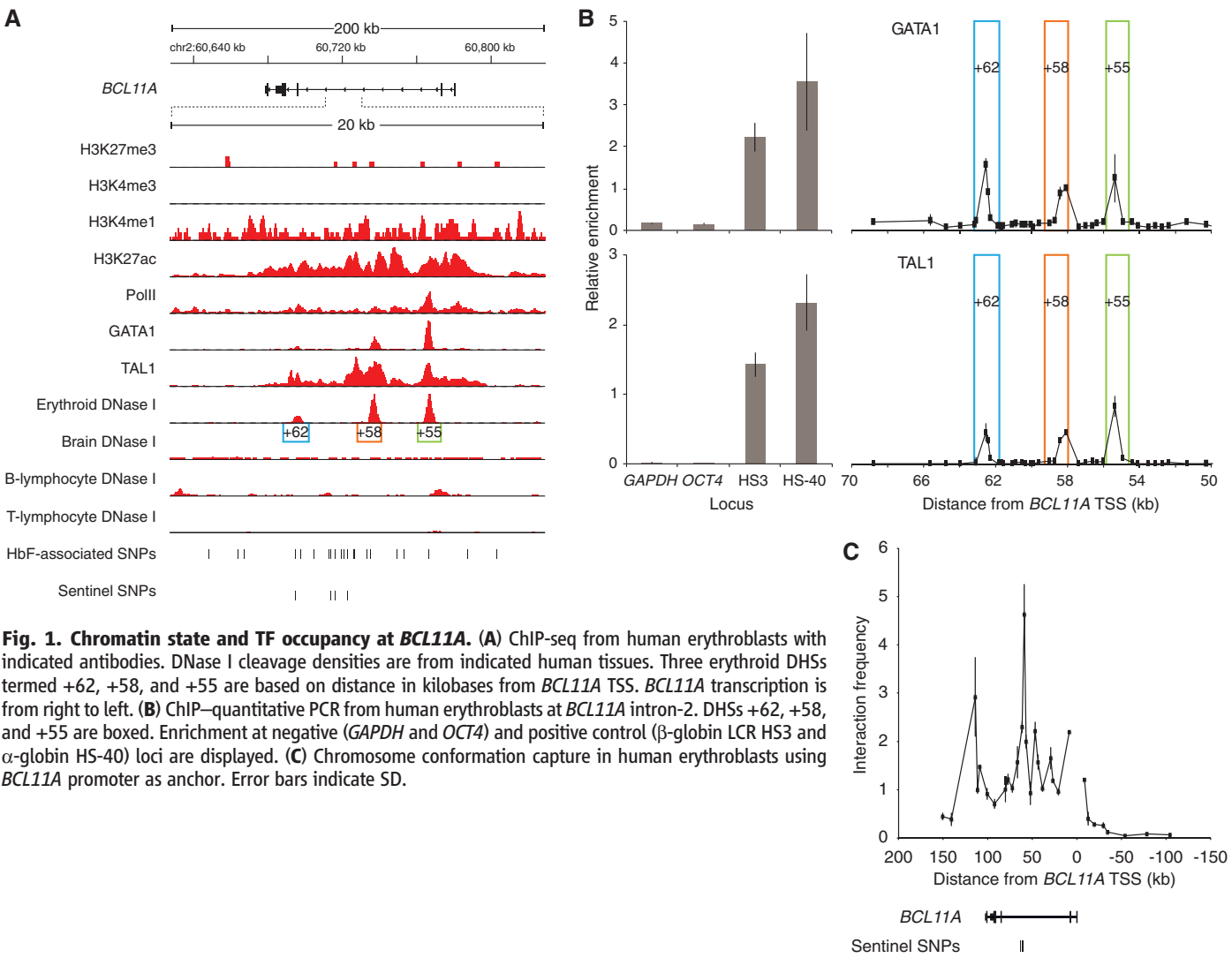


Fig. 2. Regulatory variants at *BCL11A*. (A) Genotype data obtained in 1178 individuals from CSSCD for 38 variants within *BCL11A* +62, +58, or +55 DHSs. Shown are most highly significant associations to HbF level among common (MAF > 1%) SNPs ($n = 10$ variants) before (rs1427407) or after (rs7606173) conditional analysis on rs1427407. SNP coordinates are chromosome 2, build hg19. (B) Chromatin from erythroblasts of individuals heterozygous for rs1427407, immunoprecipitated by GATA1 or TAL1 and pyrosequenced to quantify the relative abundance of the rs1427407-G allele. Composite half E-box–GATA motif previously identified (23) is shown. (C) gDNA and cDNA from erythroblasts of individuals heterozygous for rs1427407, rs7606173, and rs7569946. Haplotyping demonstrated rs7569946-G, rs1427407-G, and rs7606173-C on the same chromosome in each. Pyrosequencing was performed to quantify the relative abundance of the rs7569946-G allele.

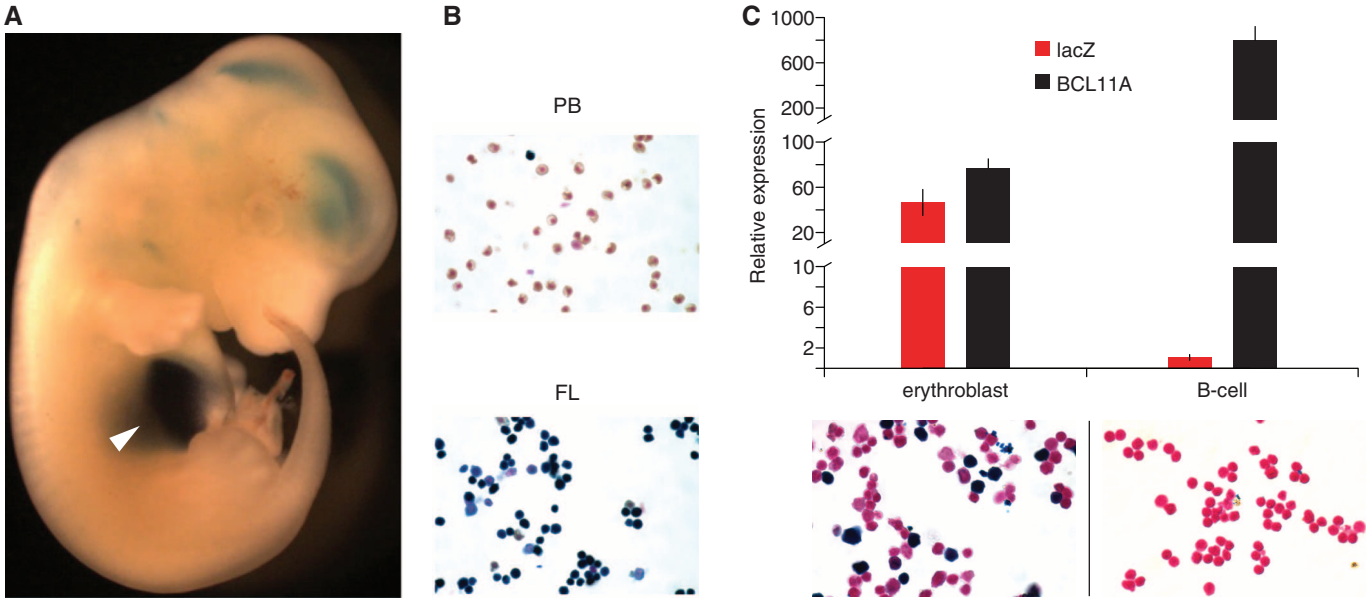
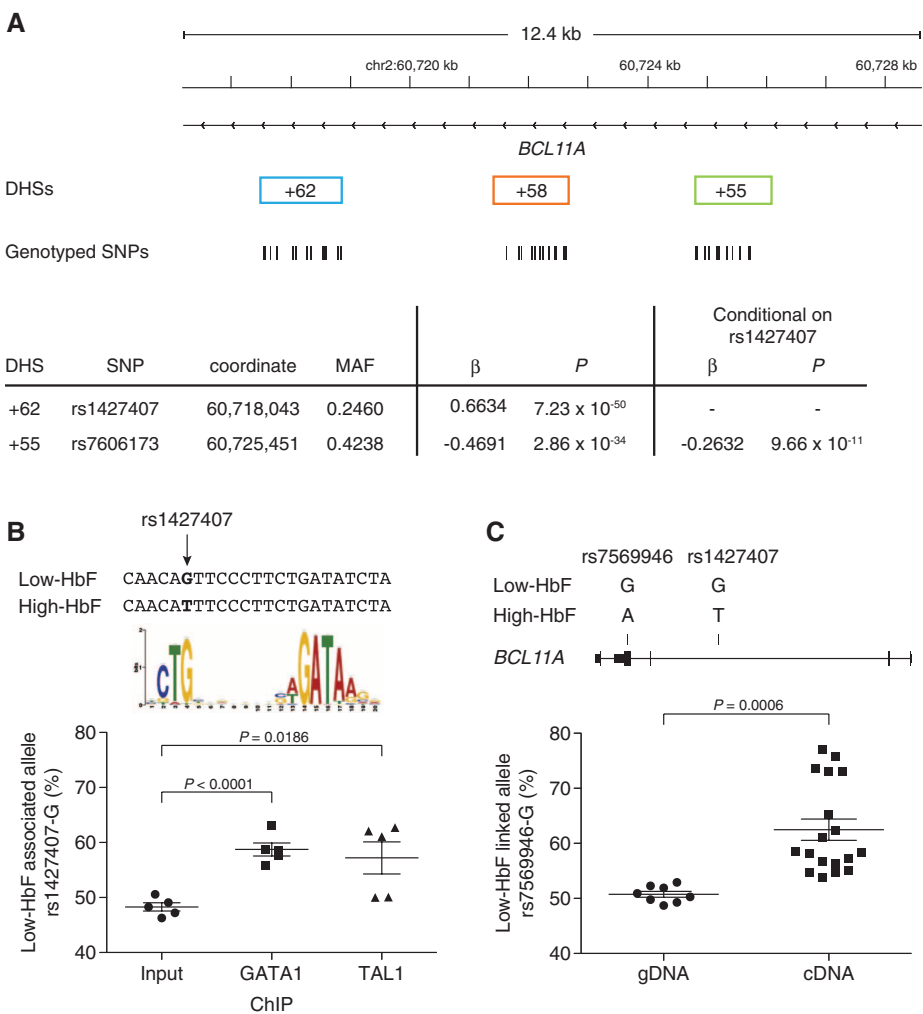


Fig. 3. The GWAS-marked *BCL11A* enhancer is sufficient for adult-stage erythroid expression. (A) A 12.4-kb fragment of *BCL11A* intron-2 (+52.0 to 64.4 kb from TSS) was cloned to a *lacZ* reporter construct. Shown is a transient transgenic mouse embryo from 12.5 dpc X-gal stained. Arrowhead indicates liver. (B) Cell suspensions isolated from peripheral blood (PB) and fetal liver

(FL) of stable transgenic embryos at 12.5 dpc X-gal stained. (C) Sorted erythroblasts and B-lymphocytes from young adult stable transgenic mice subject to X-gal staining or RNA isolation followed by quantitative reverse transcription (RT)–PCR. Gene expression was normalized to glyceraldehyde-3-phosphate dehydrogenase and expressed relative to T-lymphocytes. Error bars indicate SD.

level ($P = 7.23 \times 10^{-50}$) (Fig. 2A, fig. S2B, and supplementary text). We identified associations to HbF level within the three DHSs that remained after conditioning on rs1427407 (Fig. 2A and fig. S2B), which is consistent with the hypothesis that multiple functional SNPs within the composite enhancer act combinatorially to influence BCL11A regulation. The most significant residual association was for rs7606173 in DHS +55 ($P = 9.66 \times 10^{-11}$).

The SNP rs1427407 falls within a peak of GATA1 and TAL1 binding (Fig. 1, A and B). The minor T-allele disrupts the G-nucleotide of a sequence element resembling a half E-box/GATA composite motif [CTG(n_9)GATA], a consensus sequence enriched for chromatin bound by GATA1 and TAL1 complexes in erythroid cells (22, 23). We identified five primary erythroblast samples from individuals heterozygous for the major G-allele and minor T-allele at rs1427407 and subjected these samples to ChIP followed by pyrosequencing. As anticipated, we observed an even balance of alleles in the input DNA. However, we detected more frequent binding to the G-allele as compared with the T-allele in both the GATA1 and TAL1 immunoprecipitated chromatin samples (Fig. 2B).

Because the common synonymous SNP rs7569946 lies within exon-4 of *BCL11A*, it can be used to discriminate expression of alleles. We identified three primary erythroblast samples doubly heterozygous for the rs1427407–rs7606173 haplotype and rs7569946. For each sample, we determined by means of molecular haplotyping that the major rs7569946 G-allele was in phase with the low-HbF-associated rs1427407–rs7606173 G–C haplotype (table S4) (24, 25). Pyrosequencing revealed that whereas the alleles were balanced in genomic DNA (gDNA), significant imbalance was observed in complementary DNA (cDNA) with 1.7-fold increased expression of the low-HbF-linked G-allele of rs7569946 (Fig. 2C and supplementary text).

To understand the context within which these apparent regulatory trait-associated SNPs play their role, we explored the function of the harboring composite element. We cloned a 12.4-kb (+52.0 to 64.4 kb from TSS) human gDNA fragment containing the three erythroid DHSs in order to assay enhancer potential in a mouse transgenic *lacZ* reporter assay (fig. S4). Endogenous BCL11A shows abundant expression throughout the developing central nervous system, with much lower expression observed in the fetal liver (26). In contrast, we observed in the transgenic embryos reporter gene expression largely confined to the fetal liver, the site of definitive erythropoiesis, with weaker expression noted in the central nervous system (Fig. 3A).

A characteristic feature of globin gene and *BCL11A* expression is developmental regulation (supplementary text). In stable transgenic *BCL11A* +52.0- to 64.4-kb reporter lines at 12.5 days post coitum (dpc), circulating primitive erythrocytes failed to stain for X-gal, whereas definitive erythroblasts in fetal liver robustly stained positive (Fig. 3B). Endogenous BCL11A was expressed at 10.4-fold-higher levels in B-lymphocytes as compared with erythroblasts. LacZ expression was restricted to erythroblasts and not observed in B-lymphocytes (Fig. 3C). These results indicate that the GWAS-marked *BCL11A* intron-2 regulatory sequences are sufficient to specify developmentally restricted, erythroid-specific gene expression.

We aimed to disrupt the enhancer to investigate its requirement for BCL11A expression. Because there are no suitable adult-stage human erythroid cell lines, we turned to the mouse erythroleukemia (MEL) cell line. We observed an orthologous enhancer signature at intron-2 of mouse *Bcl11a* indicated by sequence homology, erythroid-specific DNase I hypersensitivity, characteristic histone marks, and GATA1/TAL1 occupancy (fig. S6) (22, 27). Sequence-specific nucleases can produce small chromosomal deletions via nonhomologous

end joining (NHEJ)-mediated repair (28). We engineered transcription activator-like effector nucleases (TALENs) to introduce double-strand breaks to flank the orthologous 10-kb *Bcl11a* intron-2 sequences carrying the erythroid enhancer chromatin signature (fig. S7A). Three different clones were isolated that had undergone biallelic excision of the intronic segment (figs. S7 and S8 and supplementary text). BCL11A transcript was profoundly reduced in the absence of the orthologous erythroid composite enhancer (Fig. 4A). BCL11A protein expression was not detectable in the enhancer-deleted clones (Fig. 4B). In the absence of the BCL11A enhancer, embryonic globin gene derepression was pronounced, with the ratio of embryonic $\epsilon\gamma$ to adult $\beta 1/2$ globin increased by a mean of 364-fold (fig. S9).

To examine potential lineage-restriction of the requirement for the +50.4- to 60.4-kb intronic sequences for BCL11A expression, we evaluated their loss in a nonerythroid context. The same strategy of introduction of two pairs of TALENs to obtain clones with NHEJ-mediated deletion was used in a pre-B lymphocyte cell line. In contrast to the erythroid cells, BCL11A expression was retained in the $\Delta 50.4$ - to 60.4-kb enhancer deleted pre-B cell clones at both the RNA and protein levels (Fig. 4, A and B). These results indicate that the orthologous erythroid enhancer sequences are essential for erythroid gene expression but are not required in B-lymphoid cells for integrity of transcription from the *Bcl11a* locus.

The prior identification of BCL11A as a critical repressor of HbF levels has raised new hope for mechanism-based therapeutic approaches to the β -hemoglobinopathies (29). However, the paradox that genetic variation at *BCL11A* is common, well-tolerated, and disease-protective despite the critical roles of BCL11A in neurogenesis and lymphopoiesis (19, 20, 30) has remained unresolved. We have demonstrated that the HbF-associated variants localize to an erythroid enhancer of *BCL11A*. Through allele-specific analyses, we show that

Downloaded from https://www.science.org at EPF Lausanne on September 25, 2024

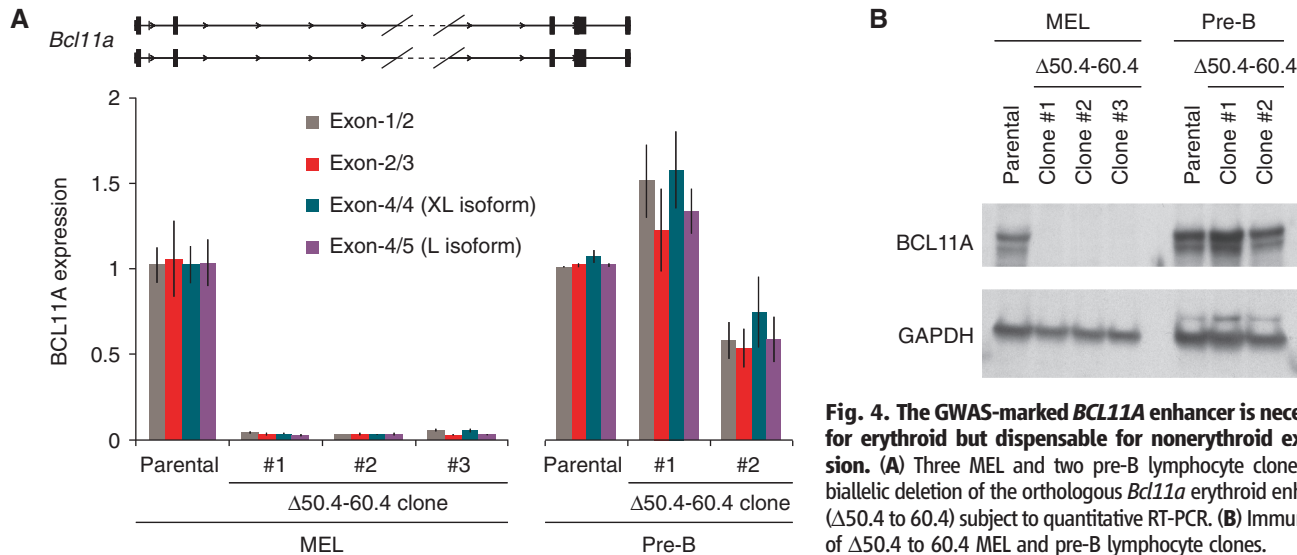


Fig. 4. The GWAS-marked *BCL11A* enhancer is necessary for erythroid but dispensable for nonerythroid expression. (A) Three MEL and two pre-B lymphocyte clones with biallelic deletion of the orthologous *Bcl11a* erythroid enhancer ($\Delta 50.4$ to 60.4) subject to quantitative RT-PCR. (B) Immunoblot of $\Delta 50.4$ to 60.4 MEL and pre-B lymphocyte clones.

genetic variation within this enhancer is associated with modest impact on TF binding, BCL11A expression, and HbF level. Relatively small effect sizes associated with individual variants may not be surprising given that most single-nucleotide substitutions, even within critical motifs, result in only modest loss of enhancer activity (31, 32). In contrast, loss of the *BCL11A* enhancer results in the absence of BCL11A expression in the erythroid lineage. Most trait-associated SNPs identified by GWAS are noncoding and have small effect sizes (1, 33). The impact of GWAS-identified SNPs on biological processes is often uncertain. Our findings underscore how a modest influence engendered by an individual noncoding variant neither predicts nor precludes a profound contribution of an underlying regulatory element.

Challenges to inhibiting BCL11A for mechanism-based reactivation of HbF include the supposedly “undruggable” nature of transcription factors (34) and its important nonerythroid functions (20, 30). With recent developments in their efficiency and precision, sequence-specific nucleases can be designed to exquisitely target genomic sequences of interest (35–37). We propose the GWAS-identified enhancer of *BCL11A* as a particularly promising therapeutic target for genome engineering in the β -hemoglobinopathies. Disruption of this enhancer would impair BCL11A expression in erythroid precursors with resultant HbF derepression while sparing BCL11A expression in nonerythroid lineages. Rational intervention might mimic common protective genetic variation.

References and Notes

1. L. Fugger, G. McVean, J. I. Bell, *N. Engl. J. Med.* **367**, 2370–2371 (2012).
2. J. Ernst et al., *Nature* **473**, 43–49 (2011).
3. D. S. Paul et al., *PLoS Genet.* **7**, e1002139 (2011).
4. M. T. Maurano et al., *Science* **337**, 1190–1195 (2012).
5. P. van der Harst et al., *Nature* **492**, 369–375 (2012).
6. ENCODE Project Consortium et al., *Nature* **489**, 57–74 (2012).
7. S. Menzel et al., *Nat. Genet.* **39**, 1197–1199 (2007).
8. M. Uda et al., *Proc. Natl. Acad. Sci. U.S.A.* **105**, 1620–1625 (2008).
9. G. Lettre et al., *Proc. Natl. Acad. Sci. U.S.A.* **105**, 11869–11874 (2008).
10. M. Nuinon et al., *Hum. Genet.* **127**, 303–314 (2010).
11. N. Solovieff et al., *Blood* **115**, 1815–1822 (2010).
12. P. Bhatnagar et al., *J. Hum. Genet.* **56**, 316–323 (2011).
13. V. G. Sankaran et al., *Science* **322**, 1839–1842 (2008).
14. J. Xu et al., *Genes Dev.* **24**, 783–798 (2010).
15. J. Xu et al., *Proc. Natl. Acad. Sci. U.S.A.* **110**, 6518–6523 (2013).
16. V. G. Sankaran et al., *Nature* **460**, 1093–1097 (2009).
17. J. Xu et al., *Science* **334**, 993–996 (2011).
18. F. Esteghamat et al., *Blood* **121**, 2553–2562 (2013).
19. P. Liu et al., *Nat. Immunol.* **4**, 525–532 (2003).
20. Y. Yu et al., *J. Exp. Med.* **209**, 2467–2483 (2012).
21. M. D. Farber, M. Koshy, T. R. Kinney, The Cooperative Study of Sickle Cell Disease, *J. Chronic Dis.* **38**, 495–505 (1985).
22. E. Soler et al., *Genes Dev.* **24**, 277–289 (2010).
23. M. T. Kassouf et al., *Genome Res.* **20**, 1064–1083 (2010).

24. D. J. Turner, M. E. Hurles, *Nat. Protoc.* **4**, 1771–1783 (2009).
25. J. Tyson, J. A. Armour, *BMC Genomics* **13**, 693 (2012).
26. M. Leid et al., *Gene Expr. Patterns* **4**, 733–739 (2004).
27. M. S. Kowalczyk et al., *Mol. Cell* **45**, 447–458 (2012).
28. H. J. Lee, E. Kim, J. S. Kim, *Genome Res.* **20**, 81–89 (2010).
29. D. E. Bauer, S. C. Kamran, S. H. Orkin, *Blood* **120**, 2945–2953 (2012).
30. A. John et al., *Development* **139**, 1831–1841 (2012).
31. R. P. Patwardhan et al., *Nat. Biotechnol.* **30**, 265–270 (2012).
32. A. Melnikov et al., *Nat. Biotechnol.* **30**, 271–277 (2012).
33. K. A. Frazer, S. S. Murray, N. J. Schork, E. J. Topol, *Nat. Rev. Genet.* **10**, 241–251 (2009).
34. A. N. Koehler, *Curr. Opin. Chem. Biol.* **14**, 331–340 (2010).
35. F. D. Urnov, E. J. Rebar, M. C. Holmes, H. S. Zhang, P. D. Gregory, *Nat. Rev. Genet.* **11**, 636–646 (2010).
36. J. K. Joung, J. D. Sander, *Nat. Rev. Mol. Cell Biol.* **14**, 49–55 (2013).
37. J. van der Oost, *Science* **339**, 768–770 (2013).

Acknowledgments: We thank A. Woo, A. Cantor, M. Kowalczyk, S. Burns, J. Wright, J. Snow, J. Trowbridge, and members of the Orkin laboratory—particularly C. Peng, P. Das, G. Guo, M. Kerenyi, and E. Baena—for discussions. C. Guo and F. Alt provided the pre-B cell line; A. He and W. Pu provided the pWHERE lacZ reporter construct; C. Currie and M. Nguyen

provided technical assistance; D. Bates and T. Kutayin provided expertise with sequence analysis; R. Sandstrom provided help with data management; G. Losyev and J. Daley provided aid with flow cytometry; and J. Desimini provided graphical assistance. L. Yan at EpigenDx (Hopkinton, Massachusetts) conducted the custom pyrosequencing reactions. This work was funded by grants from the Doris Duke Charitable Foundation (2009089) and Canadian Institute of Health Research (123382) to G.L.; Amon Carter Foundation, Hyundai Hope on Wheels, NIH, Lucille Packard Foundation to M.H.P.; NIH grants U54HG004594 and U54HG007010 to J.A.S.; and NIH R01HL032259, P01HL032262, and P30DK049216 (Center of Excellence in Molecular Hematology) to S.H.O. D.E.B. is supported by National Institute of Diabetes and Digestive and Kidney Diseases Career Development Award K08DK093705. D.E.B., J.X., and S.H.O. are inventors on a patent application related to this work, filed by Boston Children’s Hospital. The CSSCD samples with DNA and associated phenotype information are available from the National Heart, Lung, and Blood Institute to researchers that have appropriate institutional review board approval to use the materials.

Supplementary Materials

www.sciencemag.org/content/342/6155/253/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S9
Tables S1 to S6
References (38–54)

18 June 2013; accepted 13 September 2013
10.1126/science.1242088

Ancient DNA Reveals Key Stages in the Formation of Central European Mitochondrial Genetic Diversity

Guido Brandt,^{1*†} Wolfgang Haak,^{2*†} Christina J. Adler,³ Christina Roth,¹ Anna Szécsényi-Nagy,¹ Sarah Karimnia,¹ Sabine Möller-Rieker,¹ Harald Meller,⁴ Robert Ganslmeier,⁴ Susanne Friederich,⁴ Veit Dresely,⁴ Nicole Nicklisch,¹ Joseph K. Pickrell,⁵ Frank Sirocko,⁶ David Reich,⁵ Alan Cooper,^{2†} Kurt W. Alt,^{1†} The Genographic Consortium[§]

The processes that shaped modern European mitochondrial DNA (mtDNA) variation remain unclear. The initial peopling by Palaeolithic hunter-gatherers ~42,000 years ago and the immigration of Neolithic farmers into Europe ~8000 years ago appear to have played important roles but do not explain present-day mtDNA diversity. We generated mtDNA profiles of 364 individuals from prehistoric cultures in Central Europe to perform a chronological study, spanning the Early Neolithic to the Early Bronze Age (5500 to 1550 calibrated years before the common era). We used this transect through time to identify four marked shifts in genetic composition during the Neolithic period, revealing a key role for Late Neolithic cultures in shaping modern Central European genetic diversity.

The Central European Neolithic and the subsequent Early Bronze Age (EBA) reflect periods of momentous cultural changes (1–4). However, the extent to which such prehistoric cultural changes were accompanied by differences in the underlying genetics of local populations (1–5) and how such population shifts contributed to the present-day genetic diversity of Central Europe (6–9) are yet to be understood. Ancient DNA studies have revealed genetic discontinuities between indigenous hunter-gatherers and early farmers and between the latter and present-day Europeans (10, 11). Although this

confirms the importance of genetic shifts after the arrival of farming, the number and sequence of events and their potential origins and contributions to the genetic composition of modern-day Central Europe remain unclear (5, 6, 12).

We collected samples from 25 sites of the Mittelbe-Saale region in Saxony-Anhalt, Germany, attributed to nine archaeological cultures of the Early, Middle, and Late Neolithic period and the EBA, spanning ~4000 years (Fig. 1A, figs. S1 and S2, and table S1) (13). Mittelbe-Saale played a key role in human prehistory in Central Europe (4, 13), and the continuous settlement activity