Supplementary materials for Resolving the chromatin impact of somatic variants with targeted Fiber-seq

Stephanie C. Bohaczuk[1], Zachary J. Amador[2], Chang Li[1], Benjamin J. Mallory[2], Elliott Swanson[2], Jane Ranchalis[1], Katherine M. Munson[2], Tomas Walsh[1], Morgan O. Hamm[2], Yizi Mao[1], Andre Lieber[1], Mitchell R. Vollger[1], Andrew B. Stergachis[1,2,3,†]

1. Division of Medical Genetics, University of Washington School of Medicine, Seattle, WA, USA
2. Department of Genome Sciences, University of Washington, Seattle, WA, USA
3. Brotman Baty Institute for Precision Medicine, Seattle, WA USA

† Corresponding author. absterga@uw.edu

Supplemental Figures

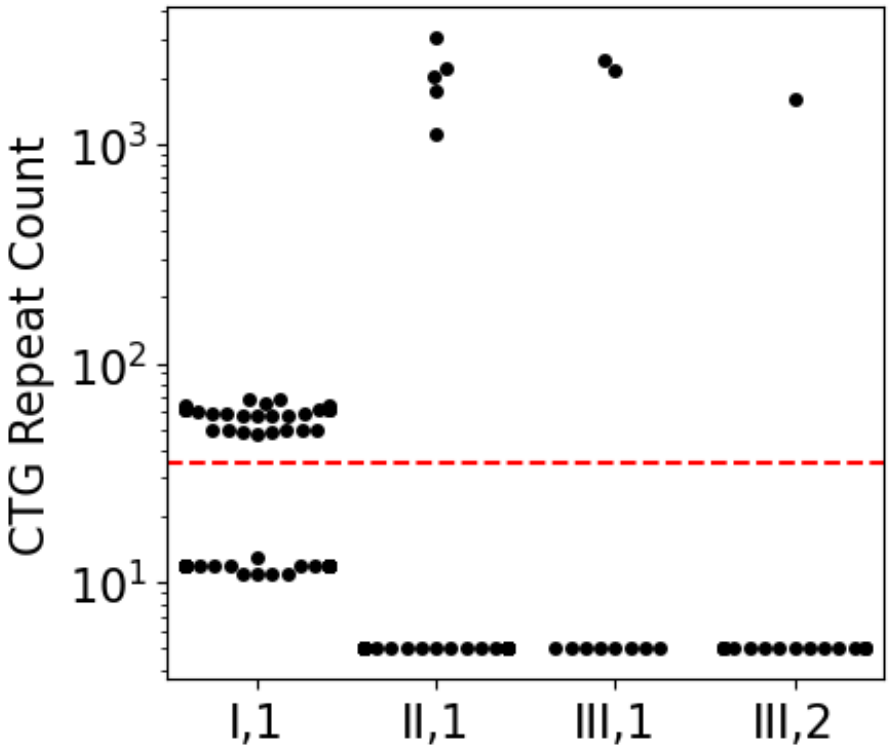

Figure S1. CTG repeat count per donor. Swarm plot of CTG repeat count in reads fully spanning the CTG repeat, separated by individual donor.

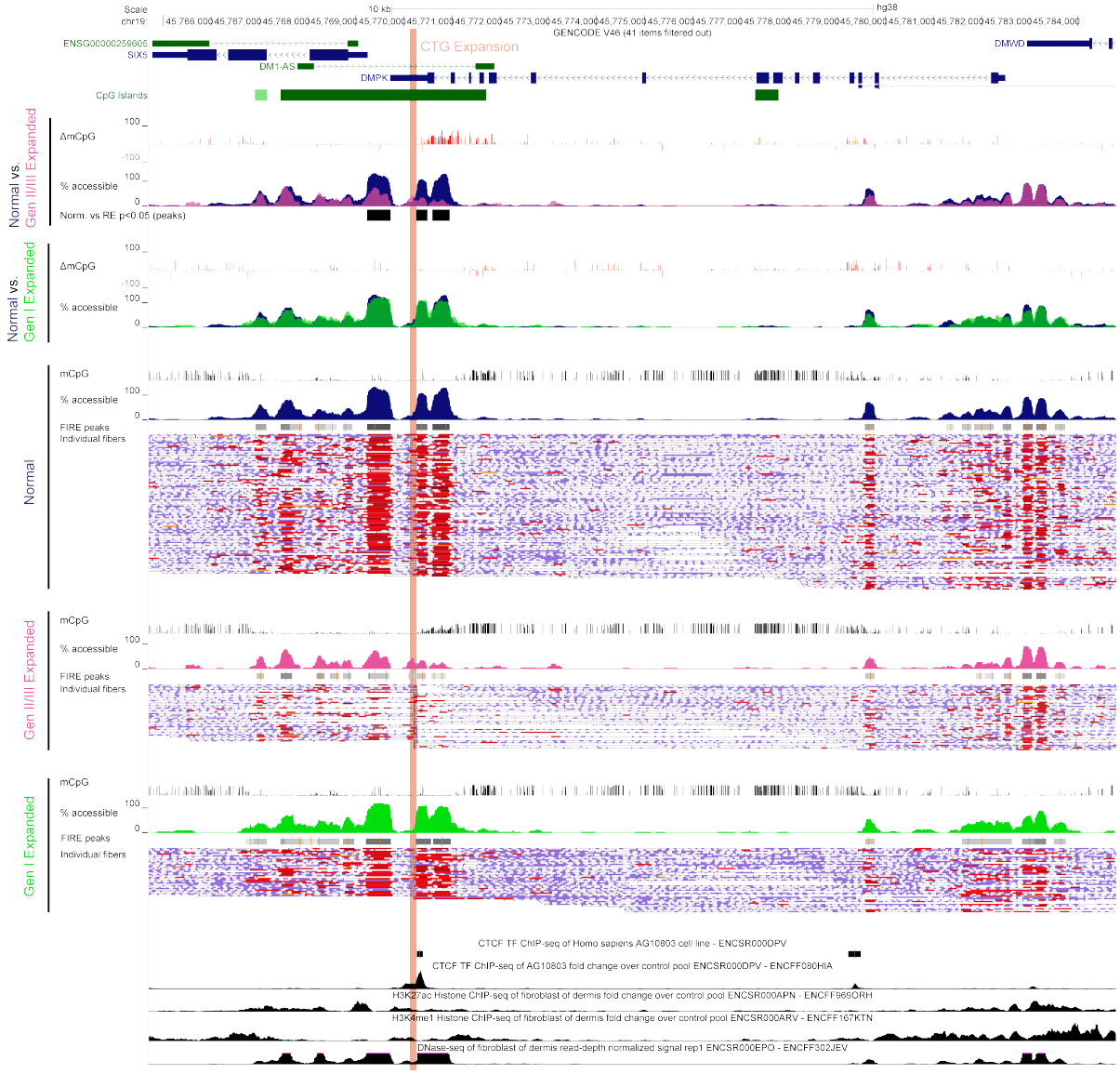

Figure S2. Targeted Fiber-seq of DM1 fibroblasts at *DMPK*. The top two panels show a comparison of mCpG and actuation between normal and expanded haplotypes. Statistical significance of percent actuation was tested by Fisher's exact test with Benjamini-Hochberg FDR correction at all FIRE peaks identified for the normal haplotype (Gen I expanded had no significant differences within this window). The bottom three panels represent Fiber-seq FIRE results for each group of haplotypes. ENCODE fibroblast DNase-seq and ChIP-seq tracks are shown below for reference.

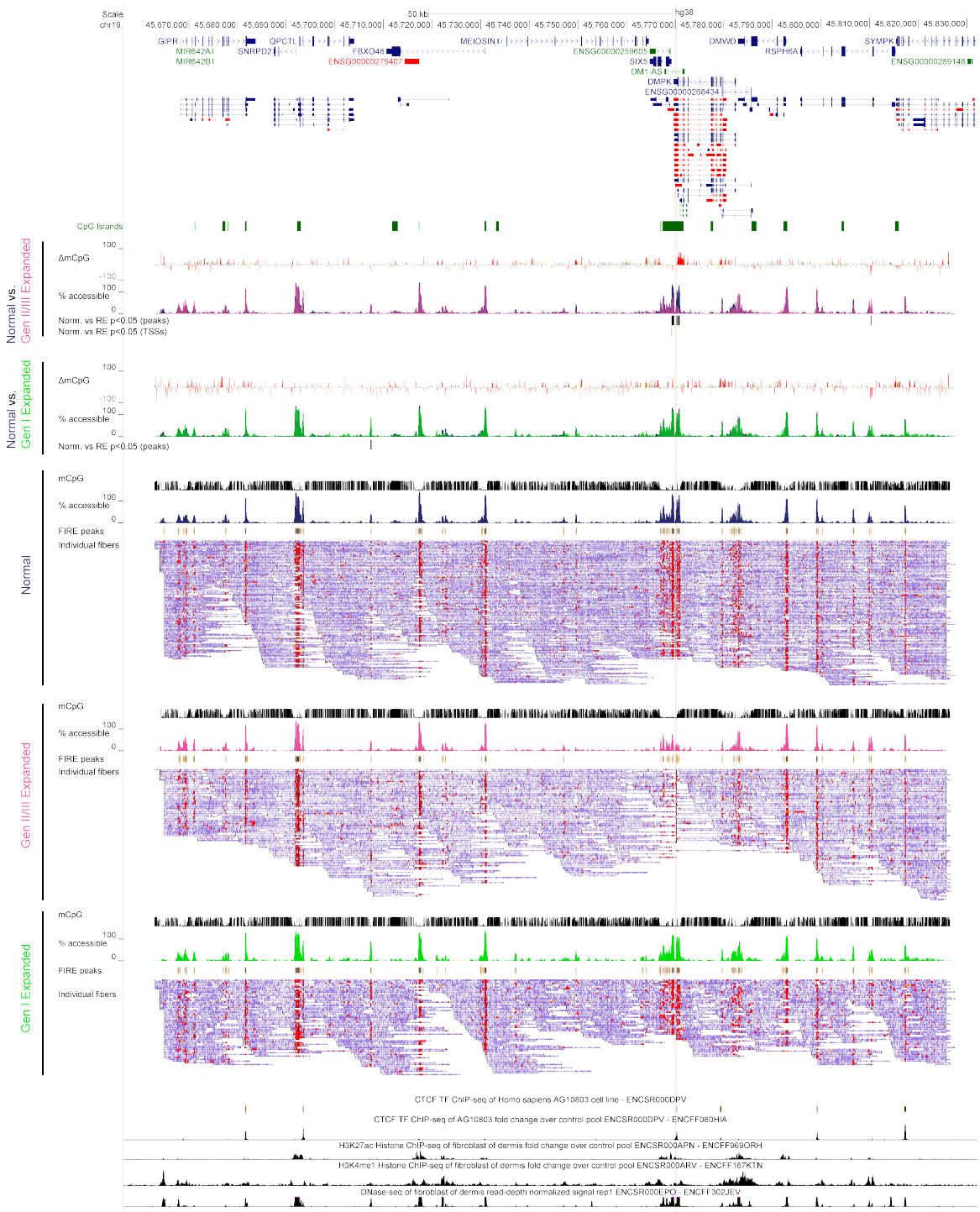

Figure S3. Targeted Fiber-seq across the targeted locus containing *DMPK*. A zoom-out of Supplemental Fig. S2 across the entire targeted region. The “Normal vs. Gen II/III Expanded” panel also contains a track showing the *SIX5* promoter as the only GENCODE V46 annotated transcriptional start site with a statistically significant difference in percent actuation compared to normal haplotypes (Fisher’s exact test with Benjamini-Hochberg FDR correction). GENCODE V46 transcripts are shown above There were no significantly different TSSs for the Gen I expanded haplotype.

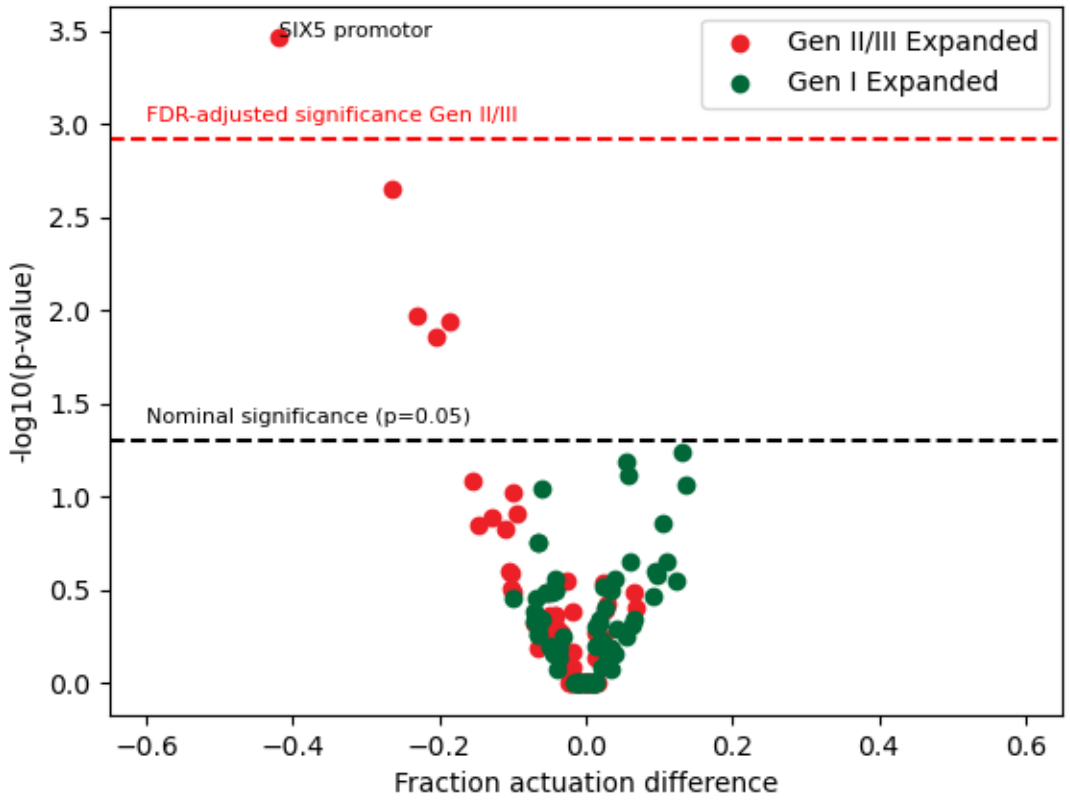

Figure S4. Volcano plot of actuation difference at all GENCODE V46 TSSs within the targeted locus. Y-axis p-values were calculated by Fisher's exact test. The fraction actuation difference (x-axis) compares actuation among fibers from normal haplotypes to the Gen II/III expanded haplotypes (red) and Gen I expanded haplotype (green). The black dashed line denotes the threshold for statistical significance ($p<0.05$). Points above the red dashed line are significant with Benjamini-Hochberg FDR correction.

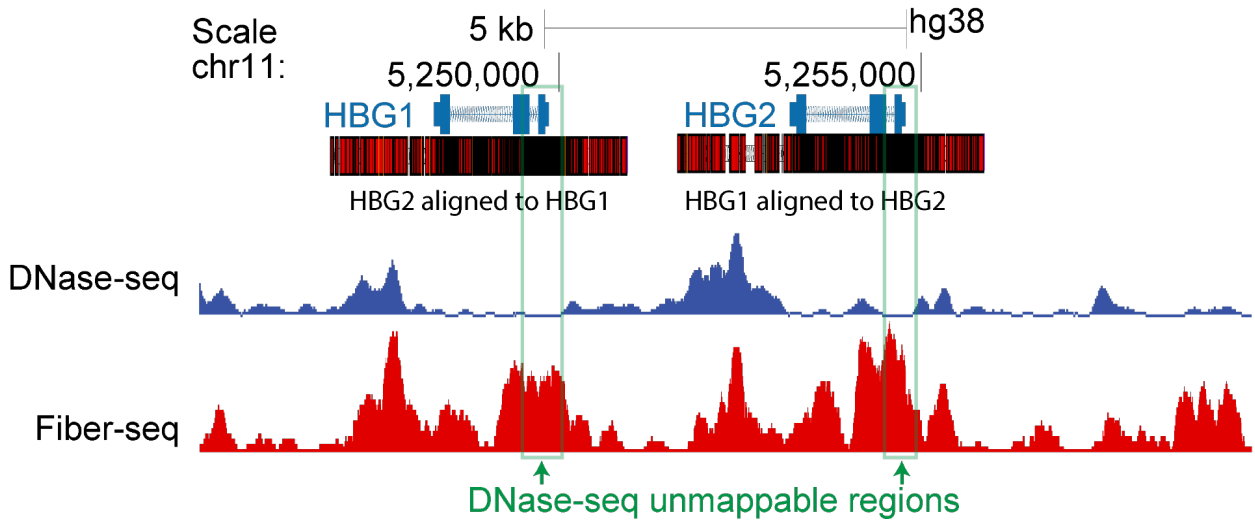

Figure S5. Sequence identity of *HBG1* and *HBG2*. Identical sequence is shown in black, with mismatches in red. Erythroid DNaseI-seq (blue, ENCODE track ENCFF800PMS) is shown above Fiber-seq percent actuated (red). Signal at the *HBG1*/*HBG2* promoters is missing from DNase-seq as short reads are not uniquely mappable due to 100% sequence identity.

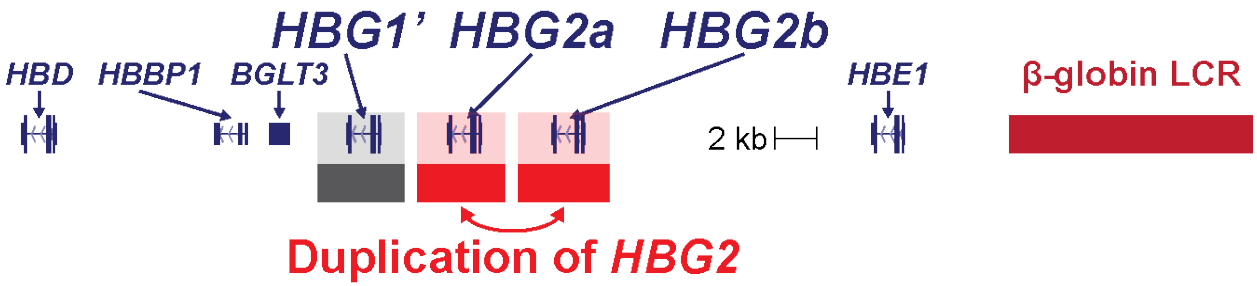

Figure S6. Schematic of the *HBG2* duplication in donor-derived CD34^{+} cells.

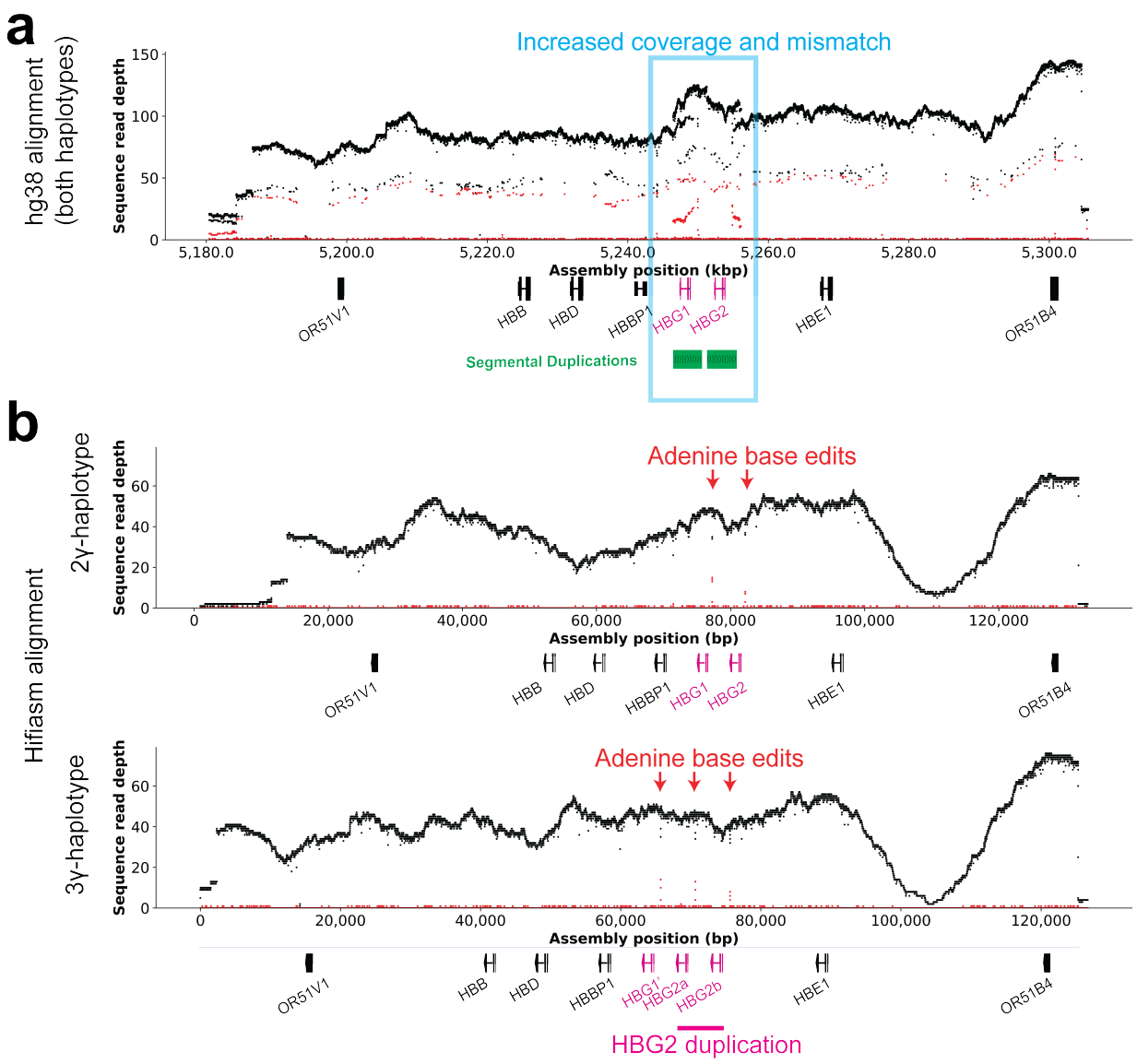

Figure S7. NucFreq plots (Vollger et al. 2019) **of donor 1 at the β-globin locus.** Read depth is shown in black, and mismatches are shown in red. **a)** Alignment to hg38, with the blue boxed region representing a spike in coverage and mismatch across the segmental duplications containing the *HBG1/HBG2* genes. **b)** Reads mapped to the diploid donor-specific assembly created with Hifiasm using HiFi reads from the same donor. Note that the spike in coverage and mismatch noted in a are resolved with inclusion of an apparent *HBG2* duplication event along the 3γ-haplotype.

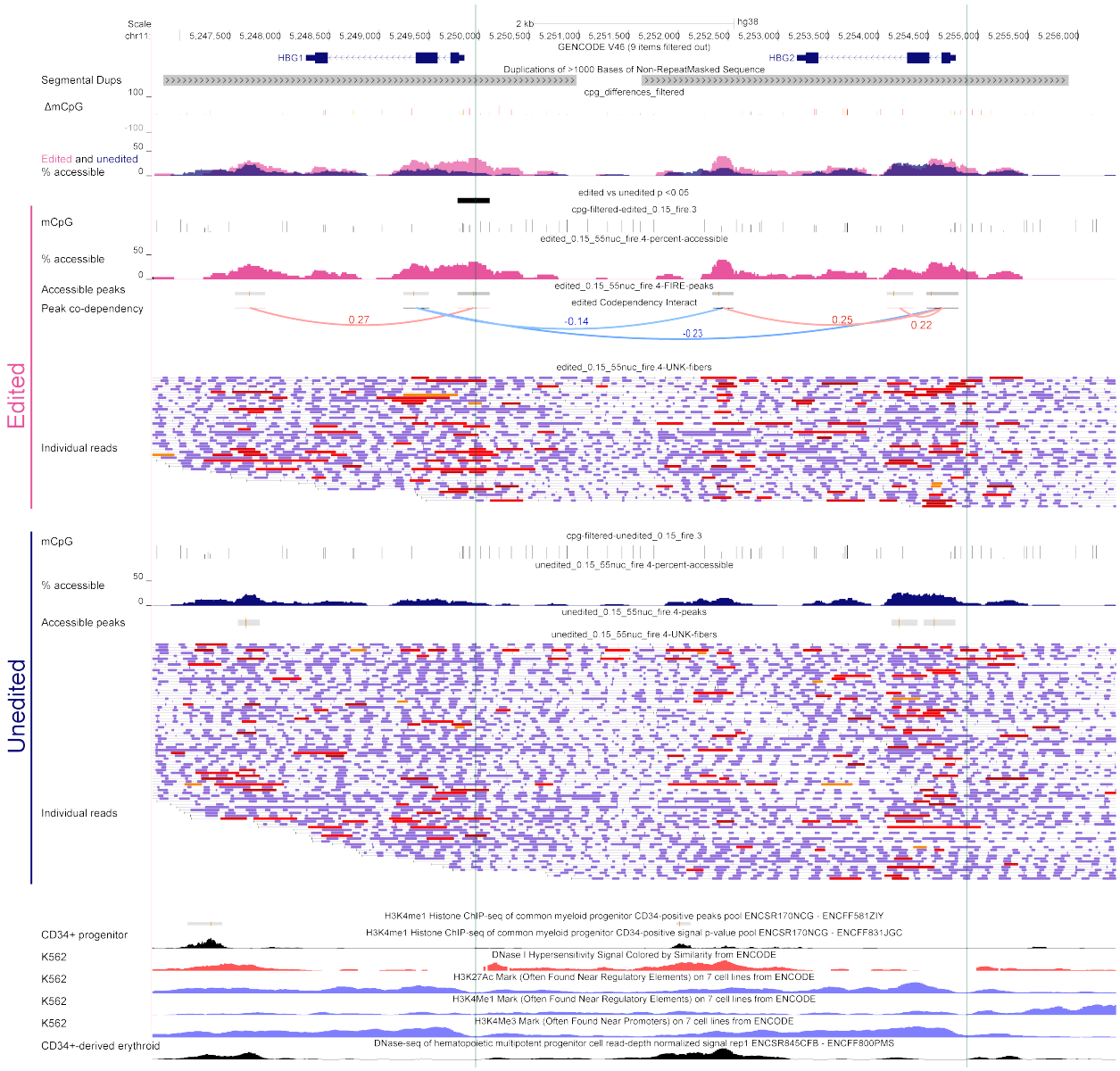

Figure S8. Targeted Fiber-seq across the γ-globin genes in ABE-edited CD34+ derived erythroid cells. Positive and negative codependencies of magnitude 0.1 or larger are displayed in red and blue, respectively. ABE-edited sites are highlighted in green. ENCODE tracks of H3K4me1 ChIP-seq from undifferentiated CD34$^+$ progenitor cells, DNaseI and histone ChIP-seq from K562 cells (which express *HBG1*/*HBG2*), and DNaseI of CD34$^+$ erythroid cells are included below.

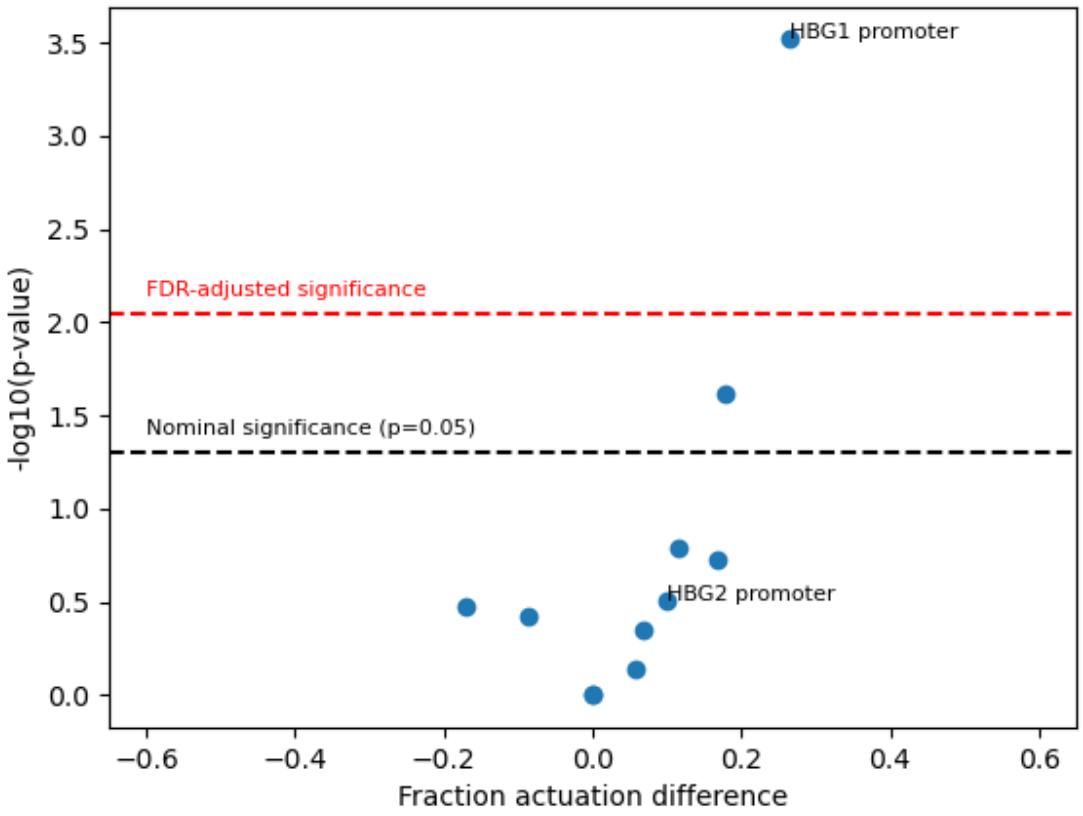

Figure S9. Volcano plot of actuation difference at all FIRE peaks from edited fibers. Y-axis p-values were calculated by Fisher's exact test. The fraction actuation difference (x-axis) compares actuation among edited and unedited fibers from ABE-edited erythroid cells. The black dashed line denotes the threshold for statistical significance ($p<0.05$). Points above the red dashed line are significant with Benjamini-Hochberg FDR correction.

Supplemental Tables

Supplemental Tables 1-2 can be found in the supplemental materials file called Supplemental_Tables.xlsx.

Supplemental Data

Hifiasm assembly of the β-globin locus from the donor with the *HBG2* duplication is available on GitHub (https://github.com/StephanieBohaczuk/Targeted-Fiber-seq/tree/main/ABE_hifiasm_assembly) and in the supplemental materials file Supplemental_Data.zip.

Supplemental Code

The code to reproduce figures and analyses in this study is available on GitHub (https://github.com/StephanieBohaczuk/Targeted-Fiber-seq). A permanent copy of the code used for figures and analyses can be found in the supplemental materials file Supplemental_Code.zip.

Supplemental Note

(1) The two peaks with increased actuation in expanded vs. normal haplotypes are both associated with variants that are consistent with the observed results:

Peak chr19:45,810,189-45,810,359 (significantly increased in Generations II/III)- There is a variant at chr19:45810295 (G/C). At this position, all normal haplotypes are C, and all expanded haplotypes from all generations are G (reference). This peak is also increased in Generation I but doesn't meet significance.

Peak chr19:45707335-45707564 (significantly increased in Gen I): There is a variant at chr19:45707265 (C/T). At this position, all normal haplotypes are C (reference), the Generation I expanded haplotype is T, and the Gen II/III expanded haplotypes are C. This peak is not increased in the GenII/IIII expanded haplotype. Although the variant does not strictly overlap, the corresponding peak from the grandfather sample is wider (chr19:45,707,296-45,707,576) and there are several FIRE elements within the peak that extend to it.

References

Vollger, Mitchell R., Philip C. Dishuck, Melanie Sorensen, Annemarie E. Welch, Vy Dang, Max L. Dougherty, Tina A. Graves-Lindsay, Richard K. Wilson, Mark J. P. Chaisson, and Evan E. Eichler. 2019. “Long-Read Sequence and Assembly of Segmental Duplications.” *Nature Methods* 16 (1): 88–94.