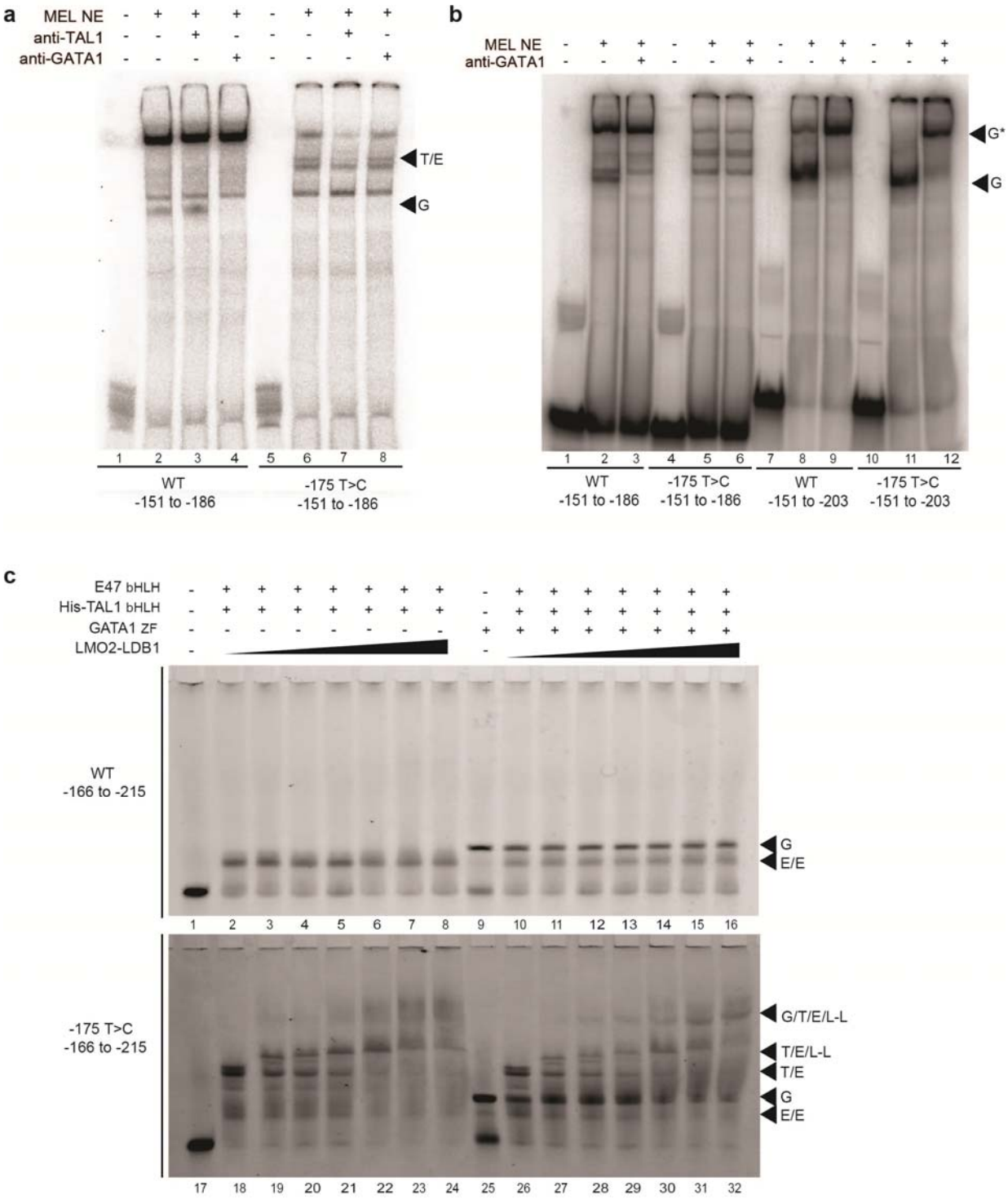
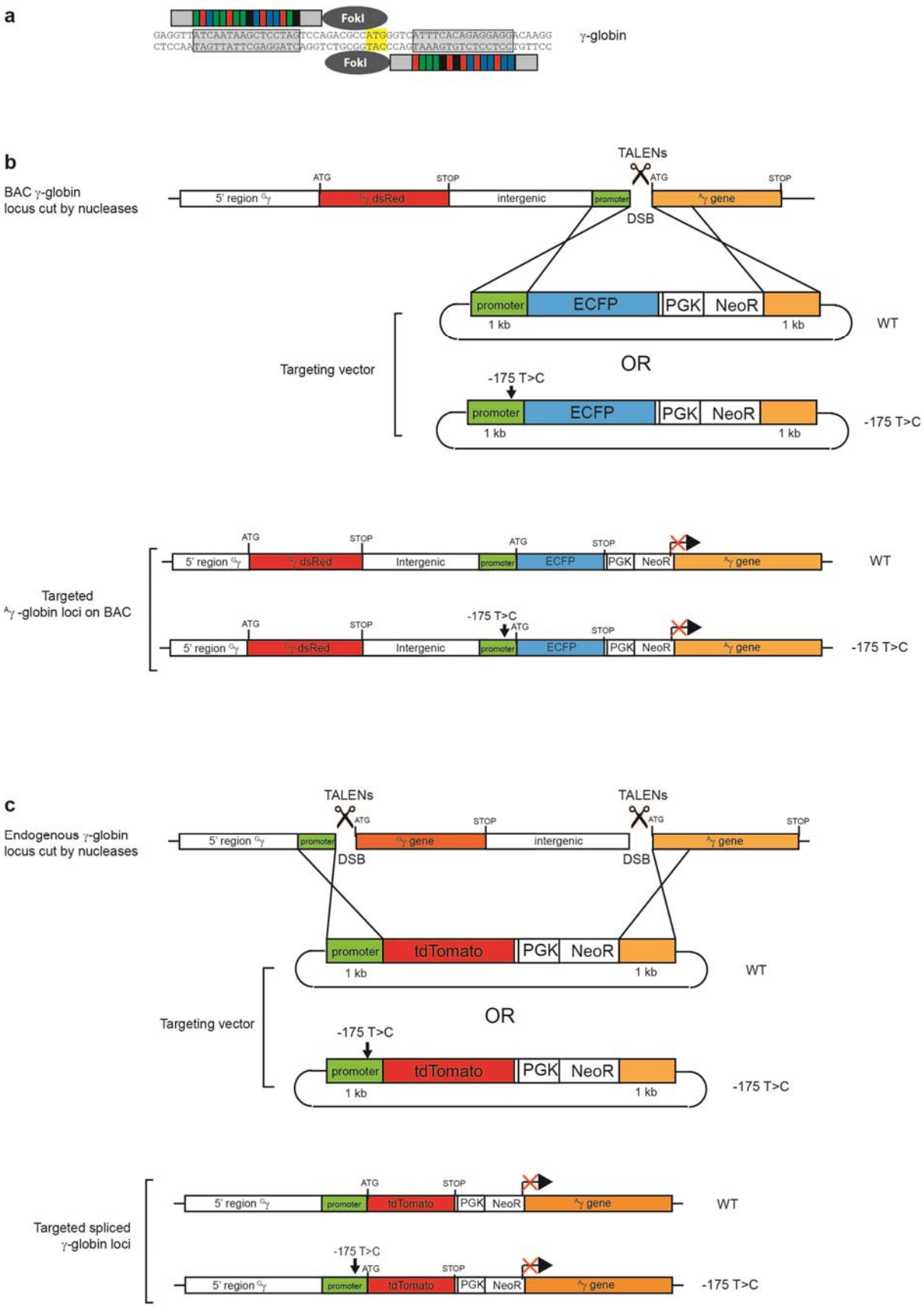
SUPPLEMENTARY INFORMATION

Supplementary Figures

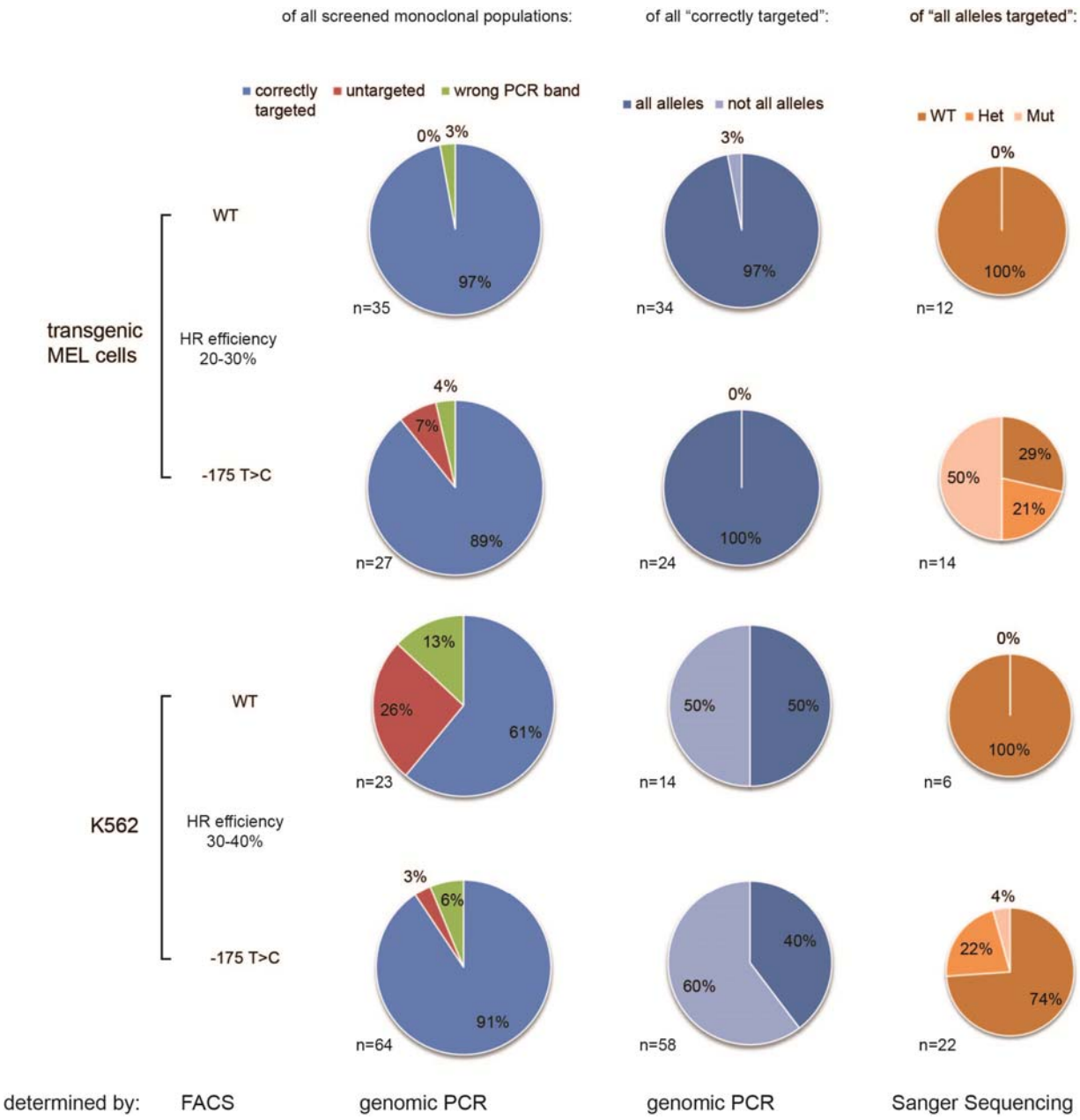


Supplementary Fig. 1: EMSAs of TAL1 and GATA1 binding to the γ-globin promoter. (**a**) EMSA showing endogenous proteins from MEL cell nuclear extract (NE) (induced with 2% DMSO for 72h) binding to the γ–globin promoter (-151 to -186). GATA1 (G) binds to the

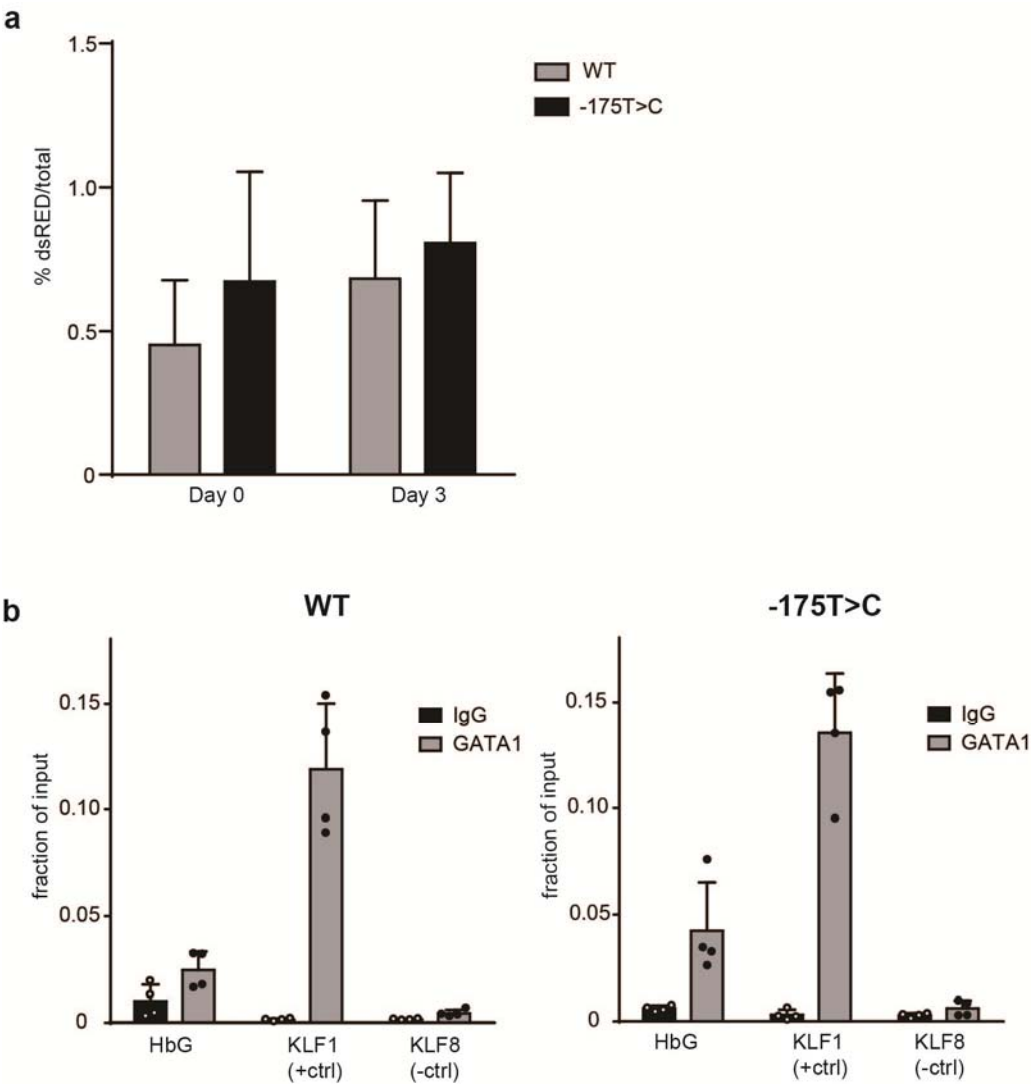
WT probe (lane 2) as confirmed by supershift with anti-GATA1 antibody (lane 4), but binding is abolished in the -175T>C probe (lane 6). Endogenous TAL1/E47 heterodimer (T/E) binding is only detected in the -175T>C probe (lane 6) as confirmed by supershift with TAL1 antibody (lane 7). (**b**) EMSA showing endogenous GATA1 from MEL cell NE (induced with 2% DMSO for 72h) binding to the γ–globin promoter. Probes in lanes 1-6 contain only the -175 GATA motif whereas probes in lanes 7-12 contain both GATA motifs (-175 and -185). Binding of GATA1 to the -175T>C mutant probe is abolished when only the -175 GATA site is present (lane 5) but is retained when both GATA sites are available (lane 11). Supershift with anti-GATA1 confirmed the identity of the protein (lanes 3, 6, 9 and 12). (**c**) EMSA showing a complex of GATA1-TAL1-E47 and LMO2-LDB1 binding to the mutant -175T>C promoter. Proteins were bacterially expressed and purified by ion-exchange chromatography. E47 and TAL1 were incubated with increasing amounts of an LMO2-LDB1 fusion protein and either the WT γ–globin promoter probe (lanes 1-8) or the mutant -175T>C γ–globin promoter probe (lanes 17-24). Only the -175T>C probe shows binding of TAL1/E47 heterodimer (T/E) and TAL1/E47/LMO2-LDB1 complex (T/E/L-L). Both WT and -175T>C probes show background binding of E47 heterodimer (E/E). Addition of GATA1 to the complex (lanes 9-16 for WT and 25 to 32 for -175T>C) results in a slower migrating retarded band only in the -175T>C mutant probe (visible in lanes 28 to 32). This complex is presumably GATA1/TAL1/E47/LMO2-LDB1 (G/T/E/L-L).


a
FokI
GAGGTTATCAATAAGCTCCTAGTCCAGACGCCATGGGTCATTTCACAGAGGAGGACAAGG
CTCCAATAGTTATTCGAGGATCAGGTCTGCGGTACCCAGTAAAGTGTCTCCTCCTGTTCC
FokI
γ-globin
b
BAC γ-globin locus cut by nucleases
TALENs
ATG
STOP
5' region Gγ
Gγ dsRed
intergenic
promoter
Aγ gene
DSB
promoter
ECFP
PGK
NeoR
1 kb
1 kb
WT
Targeting vector
OR
-175 T>C
-175 T>C
Targeted Aγ -globin loci on BAC
Intergenic
ECFP
c
Endogenous γ-globin locus cut by nucleases
Gγ gene
tdTomato
Targeted spliced γ-globin loci

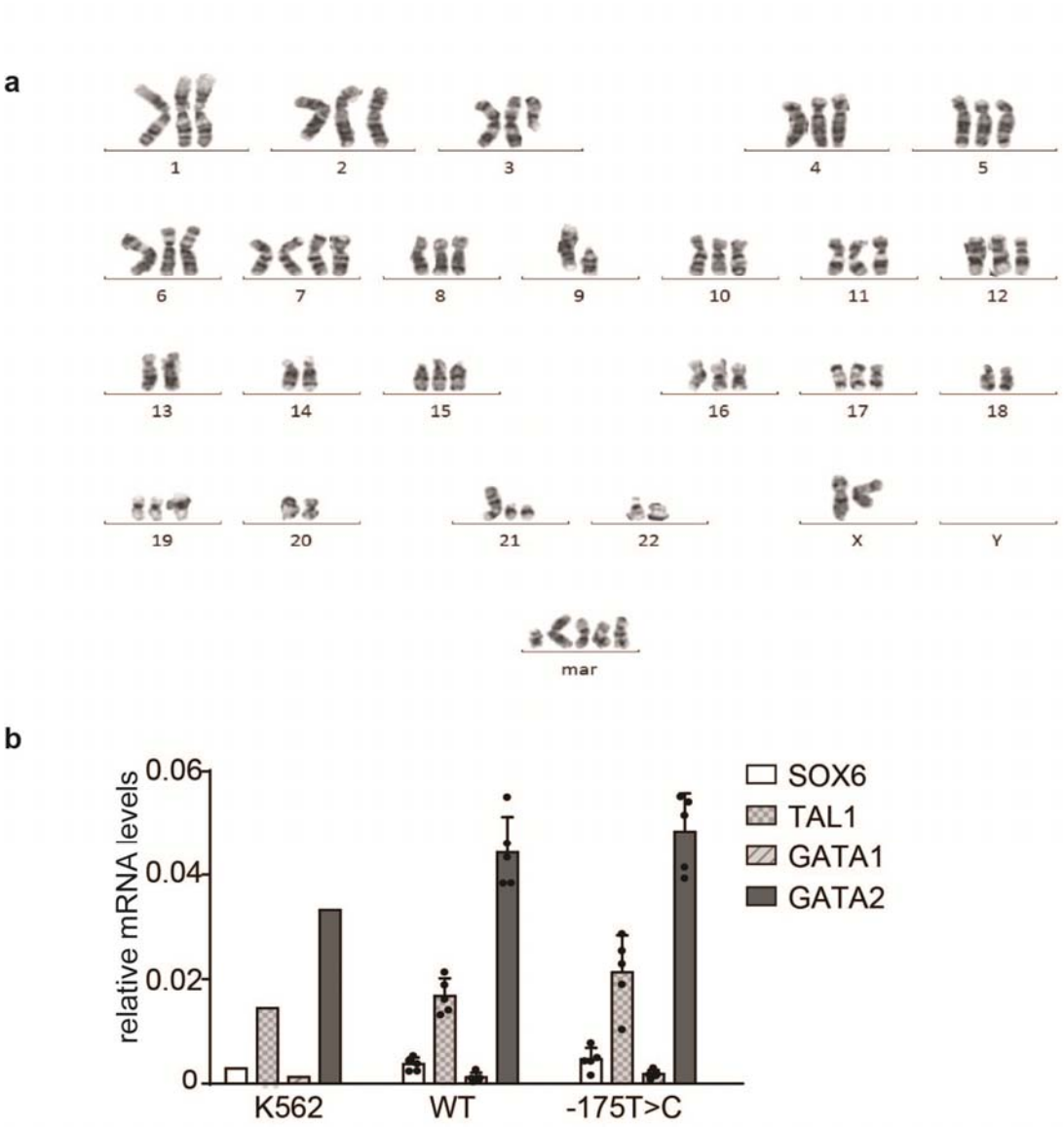
Supplementary Fig. 2. Genome-editing strategy introducing the -175 T>C mutation into the *γ-globin locus*. (**a**) Schematic of Tal-Effector Nucleases (TALENs) targeting the translation start site of the *γ-globin* gene. DNA sequences recognized by TALENs are highlighted and boxed. (**b**) Gene editing strategy in MEL cells ($^{G}\gamma$ dsRED βEGFP): transfected TALENs create a DSB at the ATG start codon of $^{A}\gamma$-globin gene on the transgenic BAC. Cotransfection of targeting vector with 1kb arms of homology 5' and 3' from DSB results in integration of ECFP via homologous recombination. ECFP is driven by the $^{A}\gamma$ promoter whereas the $^{A}\gamma$ gene is lacking a promoter region and is not expressed. (**c**) Gene editing strategy in K562 cells: TALENs create a double strand break (DSB) at the ATG of both endogenous *γ-globin* genes. TdTomato is integrated by homologous recombination after cotransfection with a targeting vector with 1kb arms of homology on the 5' and 3' end of the DSB. TdTomato reporter is now driven by the endogenous $^{G}\gamma$ promoter. The $^{G}\gamma$ gene was spliced in this process and the $^{A}\gamma$ gene is lacking a promoter region and cannot be expressed.



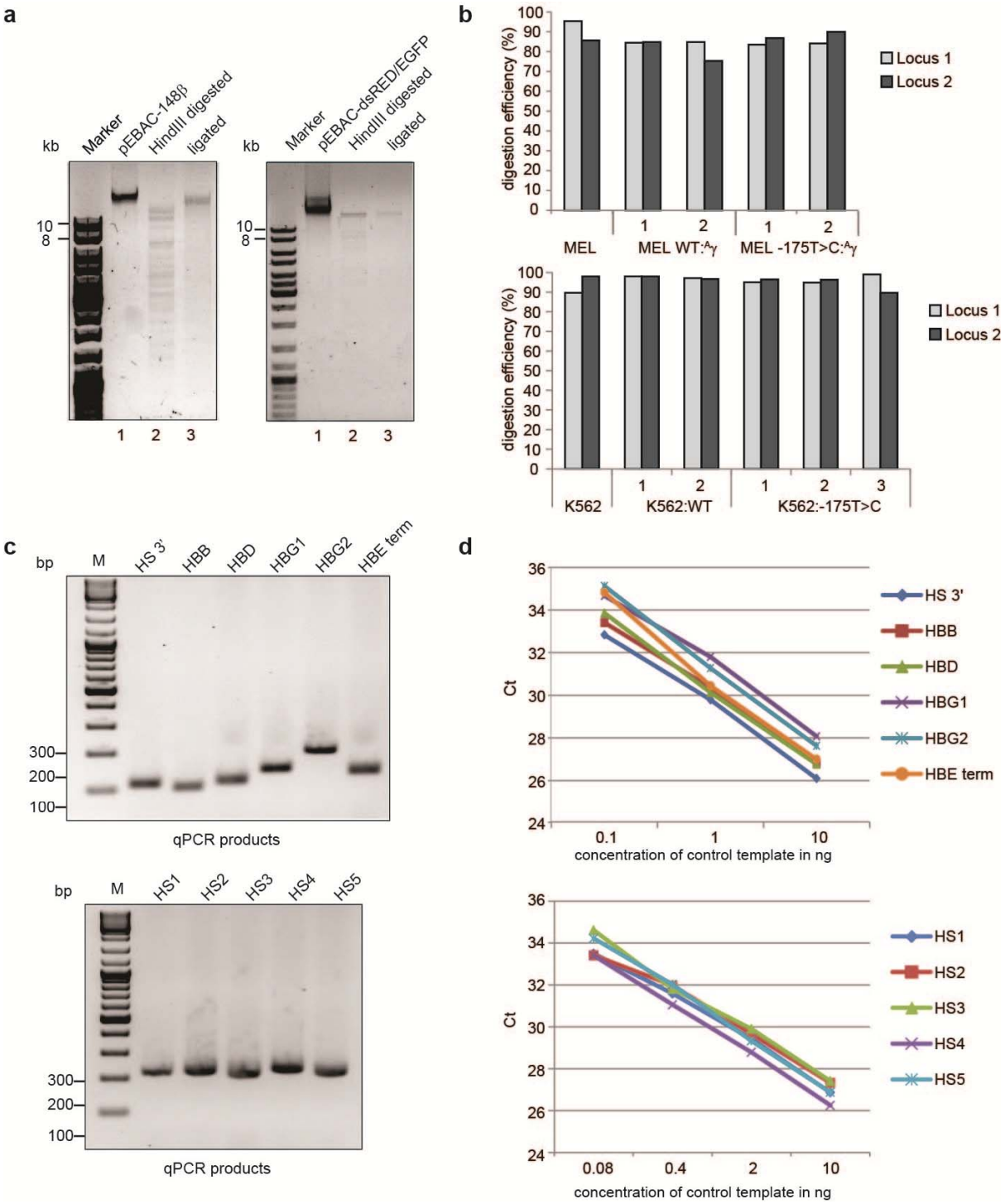
Supplementary Fig. 3. Statistics on genome-editing in transgenic MEL and K562 cells. After selection with G418 for 3-6 days cells were monitored for fluorescence (% fluorescent cells = % homologous recombination (HR) efficiency) and then sorted by FACS into 96-well plates to obtain monoclonal populations. Monoclonal populations were then screened by genomic PCR and correctly targeted clones were further screened for homozygous or heterozygous targeting. Clones that were targeted in all alleles were further analyzed and the genomic PCR products were Sanger sequenced to determine the alleleic distribution of the -175T>C mutation (Wildtype, Heterozygote or Mutant).



Supplementary Fig. 4. Additional data for MEL $^{G}\gamma$dsRED $^{A}\gamma$ECFP βEGFP cells. (**a**) Expression of dsRED reporter gene as determined by flow cytometry. Shown is the percentage of dsRED ($^{G}\gamma$-globin) over total measurable globin reporters (ECFP, EGFP and dsRED) comparing clonal MEL WT:$^{A}\gamma$ and -175T>C:$^{A}\gamma$ cell populations (n=3) in uninduced state (Day 0) and after 72h of induction with 2% DMSO (Day 3). Difference between WT and -175T>C cells is not statistically significant (p=0.43 (Day 0) and p=0.59 (Day 3) as determined by unpaired two-tailed t-test). (**b**) Anti-GATA1 chromatin immunoprecipiation (ChIP) in MEL $^{G}\gamma$dsRED $^{A}\gamma$ECFP βEGFP cells carrying either the WT or the -175T>C mutant promoter. ChIP was performed on four independently engineered monoclonal populations of MEL $^{G}\gamma$dsRED $^{A}\gamma$ECFP βEGFP WT or -175T>C. Enrichment of the *γ-globin* promoter region is slightly but insignificantly higher in cells carrying the -175T>C mutation (p=0.19, as determined by unpaired two-tailed t-test).



Supplementary Fig. 5. Additional data for K562 cells. (**a**) Karyotype of K562 cells used in this manuscript. (**b**) Transcription factor expression levels in unmodified K562 cells (n=1), K562 WT:$^{G}\gamma$-$^{A}\gamma$ and K562 -175T>C:$^{G}\gamma$-$^{A}\gamma$ cells (n=5). No significant difference in expression of transcription factors could be detected between the three cell lines. BCL11A expression levels were also assessed but have not been demonstrated because they were low to undetectable.



Supplementary Fig. 6. 3C controls. (**a**) Representative gel electrophoresis of pEBAC^{G}γdsREDβEGFP and pBAC148β DNA. BAC DNA was purified (lane 1), digested with *Hind*III (lane 2), and ligated with T4 ligase (lane 3) to generate random ligation products of *Hind*III fragments. Ligated DNA served as control DNA in 3C assay. (**b**) Digestion efficiency of 3C samples at two different loci. Only samples with efficiency greater than 75% were considered for analysis. (**c**) Gel electrophoresis of amplification products of 3C primers to

verify primer specificity. (**d**) 3C primers were tested for linearity in qPCR by using serial diluted ligated BAC DNA as template.

Supplementary Tables

Supplementary Table 1: Clinical data on the -175T>C HPFH mutation in humans.

Ethnic background	Relation	Gender /Age	Health status	Genetic background			Hematological data					Ref
				5' $^{G}\gamma$	5' $^{A}\gamma$	other abnormalities	HbA2 [%]	HbA [%]	HbS/C [%]	HbF [%]	Composition of HbF	
-	-	Adult	healthy	normal	normal	-	2-3%	95-98%	0	<2%	66% $^{A}\gamma$ 33% $^{G}\gamma$	[1,2]
Italian (Sardinian)	Propositа	F49	healthy	**-175 T>C (Het)**	Normal	-	1.1	?	-	17	90-92% $^{G}\gamma$	[3]
	Daughter	F20	healthy	**-175 T>C (Het)**	Normal	-	1.3	?	-	21		
Black American		F27	healthy	**-175 T>C (Het)**	Normal	β^{S} (Het, in trans to -175 T>C) ǂ	1.9	27.5	40.9	29.5	100% $^{G}\gamma$	[4,5]
Black		M25	mild microcytosis	-158 C>T (Het, in cis to -175T>C) †	**-175T>C (Het)** -369C>G‡ -16C>Gɸ +24A>C	β^{C} (Het, in trans to -175 T>C) Ł	1.6	15.7	45.4	38	~80% $^{A}\gamma$ ~20% $^{G}\gamma$	[6]
Black		F30	healthy	-158 C>T (Het, in cis to -175T>C)†	**175T>C (Het)**	β^{S} (Het, in trans to -175 T>C) ǂ	1.9	16.8	40.4	40.9	~66% $^{A}\gamma$ ~34% $^{G}\gamma$	[7]
British	Propositus	M3	Thrombocy-topenic, macrocytic	**-175 T>C (Het)** -158 C>T	4 bp deletion (Het, -222 to -225)*	Trisomy 8 (mosaic)	1.1	?	-	21.6	>90% $^{G}\gamma$	[8]
	Mother	F	healthy	**-175 T>C (Het)**	4 bp deletion (Het, -222 to -225)*	-	1.5	?	-	20.1	>90% $^{G}\gamma$	
	Aunt	F	healthy	**-175 T>C (Het)**	4 bp deletion (Het, -222 to -225)*	-	?	?	-	16.2	?	

† This variant (T at position -158) has been shown to result in high $^{G}\gamma/^{A}\gamma$ ratios in patients with sickle cell disease or beta thalassemia [9]. It was first described in the first gene sequenced (from a fetus) but it is likely that the presence of C at -158 is the more common polymorphism [4,10].

‡ appears to be common as it has been found in five additional cases from adults without ndHPFH [11,12]

ɸ this variant was also found in the mother who is a HbC heterozygote

* this deletion has been shown to result in low expression levels of Agamma in adults [13]

ǂ β^{S} is a mutation in the β-globin gene that results in an amino acid change.glutamic acid is replaced with valine at position 6 in β-globin resulting in the prodcution of hemoglobin S (HbS) instead of normal HbA. HbS polymerizes and red blood cells appear sickle-shaped.

Ł β^{C} is a mutation in the β-globin gene that results in an amino acid change.Glutamic acid is replaced with lysine at position 6 in β-globin resulting in the prodcution of hemoglobin C (HbC) instead of normal HbA.

Supplementary Table 2: Primer lists

ChIP primers		
Locus	**F/R**	**Oligo sequence 5' to 3'**
human HBB promoter	F	GGAGGGCTGAGGGTTTGAAGTCC
	R	TGTCCTTGGCTCTTCTGGCACTG
human HBG promoter	F	CAAGGCTATTGGTCAAGGCAA
	R	TTCCCCACACTATCTCAATGCAAA
human ZFPM1 intron1	F	AGTCGATGTGAGCTCCGATAA
	R	GGCCAAAGATAAGGCCTCTT
mouse KLF1 promoter	F	AGCACACCACACACATATCG
	R	ATGGGCTATGAGGCTAGGAA
mouse KLF8 -4.5 kb of exon 1	F	GGTTTCTGAGACCTAACACTTCACACT
	R	CCATTTAGTCATCCAGCGAACAA

3C primers		
		Primers for digestion efficiency
Locus	**F/R**	**Oligo sequence 5' to 3'**
HS2 uncut	F	CATCACTCTAGGCTGAGAACATCTG
	R	GGCTCAAGCACAGCAATGC
HBB cut	F	GTCAGTGGGGCTGGA ATA AA
	R	TGGTCAGAGCCTCAGTTTCA
HBE term cut	F	CGTAGAGGACTAGGAAAGACCAGA
	R	TGTGCACATAAGCAGATTACTTTTT
		Primers for 3C assay
Locus	**F/R**	**Oligo sequence 5' to 3'**
HS2 (internal ctrl)	F	CATCACTCTAGGCTGAGAACATCTG
	R	GGCTCAAGCACAGCAATGC
3C HS2 (constant)	F	CATAGTTGTCAGCACAATGCCTA
3C HS 3'	R	GCCTGGTGGTGACAAAATCT
3C HBB	R	TGGTCAGAGCCTCAGTTTCA
3C HBD	R	ATACTGAAACATAGGGGACGAG
3C HBG1	R	GGAGGCAAGCTGTATCTTCAAATT
3C HBG2	R	GAGCTGTGAGGTGAAACTACCA
3C HBE term	R	TGTGCACATAAGCAGATTACTTTTT
3C HBG2 (constant)	R	GAGCTGTGAGGTGAAACTACCA
3C HS1	R	CCTGATGAGTTTTTCCTCCA
3C HS2	R	GCTTGGACTATGGGAGGTCA
3C HS3	R	CAGCCTTTTGCTCAGGGTAG
3C HS4	R	CCAAATGGGTGACTGTAGGG
3C HS5	R	AGGACATGGCCATCAGTACA

RT PCR primers		
Locus	**F/R**	**Oligo sequence 5' to 3'**
tdtomato	F	GCCGACATCCCCGATTACAAGA
	R	CGATGGTGTAGTCCTCGTTGTGG
HBG	F	CCTGTCCTCTGCCTCTGCC
	R	GGATTGCCAAAACGGTCAC
HBB	F	CACGTGGATCCTGAGAACTTCAG
	R	GGTGAATTCTTTGCCAAAGTGAT
HBE	F	TGCTGAGGAGAAGGCTGCCG
	R	TGGGTCCAGGGGTAAACAACGAGG
HBA	F	GGGTGGACCCGGTCAACTT
	R	GAGGTGGGCGGCCAGGGT
HBZ	F	GAGGACCATCATTGTGTCCA
	R	AGTGCGGGAAGTAGGTCTTG
18S	F	CACGGCCGGTACAGTGAAAC
	R	AGAGGAGCGAGCGACCAA
BCL11A	F	CGAGCACAAACGGAAACAATG
	R	GATTAGAGCTCCATGTGCAGAACG
GATA2	F	AAGGCTCGTTCCTGTTCAGA
	R	TGCCCATTCATCTTGTGGTA
GATA1	F	TGCTCTGGTGTCCTCCACAC
	R	TGGGAGAGGAATAGGCTGCT
TAL1	F	AGCCGGATGCCTTCCCTAT
	R	CCGCACAACTTTGGTGTGG
SOX6	F	GGCGTCCCCCTACCCTGTCATCC
	R	TGCTGCACACGGCTCCTCACTG

Supplementary References

1. Comi, P. *et al.* Globin chain synthesis in single erythroid bursts from cord blood: studies on gamma leads to beta and G gamma leads to A gamma switches. *Proc Natl Acad Sci U S A* **77**, 362-5 (1980).
2. http://www.nlm.nih.gov/medlineplus/ency/article/003639.htm.
3. Ottolenghi, S. *et al.* Sardinian G gamma-HPFH: a T----C substitution in a conserved "octamer" sequence in the G gamma-globin promoter. *Blood* **71**, 815-7 (1988).
4. Surrey, S., Delgrosso, K., Malladi, P. & Schwartz, E. A single-base change at position -175 in the 5'-flanking region of the G gamma-globin gene from a black with G gamma-beta+ HPFH. *Blood* **71**, 807-10 (1988).

5. Friedman, S. & Schwartz, E. Hereditary persistence of foetal haemoglobin with beta-chain synthesis in cis position (Ggamma-beta+-HPFH) in a negro family. *Nature* **259**, 138-40 (1976).
6. Stoming, T.A. *et al.* An A gamma type of nondeletional hereditary persistence of fetal hemoglobin with a T----C mutation at position -175 to the cap site of the A gamma globin gene. *Blood* **73**, 329-33 (1989).
7. Coleman, M.B. *et al.* G gamma A gamma (beta+) hereditary persistence of fetal hemoglobin: the G gamma -158 C-->T mutation in cis to the -175 T-->C mutation of the A gamma-globin gene results in increased G gamma-globin synthesis. *Am J Hematol* **42**, 186-90 (1993).
8. Craig, J.E., Sheerin, S.M., Barnetson, R. & Thein, S.L. The molecular basis of HPFH in a British family identified by heteroduplex formation. *Br J Haematol* **84**, 106-10 (1993).
9. Labie, D. *et al.* The -158 site 5' to the G gamma gene and G gamma expression. *Blood* **66**, 1463-5 (1985).
10. Gilman, J.G. & Huisman, T.H. DNA sequence variation associated with elevated fetal G gamma globin production. *Blood* **66**, 783-7 (1985).
11. Slightom, J.L., Blechl, A.E. & Smithies, O. Human fetal G gamma- and A gamma-globin genes: complete nucleotide sequences suggest that DNA can be exchanged between these duplicated genes. *Cell* **21**, 627-38 (1980).
12. Shen, S.H., Slightom, J.L. & Smithies, O. A history of the human fetal globin gene duplication. *Cell* **26**, 191-203 (1981).
13. Gilman, J.G., Johnson, M.E. & Mishima, N. Four base-pair DNA deletion in human A gamma globin-gene promoter associated with low A gamma expression in adults. *Br J Haematol* **68**, 455-8 (1988).