***In vivo* HSPC gene therapy with base editors allows for efficient reactivation of fetal γ-globin in β-YAC mice**

Chang Li[1], Aphrodite Georgakopoulou[1,3], Arpit Mishra[1], Sucheol Gil[1], R. David Hawkins[1], Evangelia Yannaki[3], André Lieber[1,2]

SUPPLEMENTARY INFORMATION

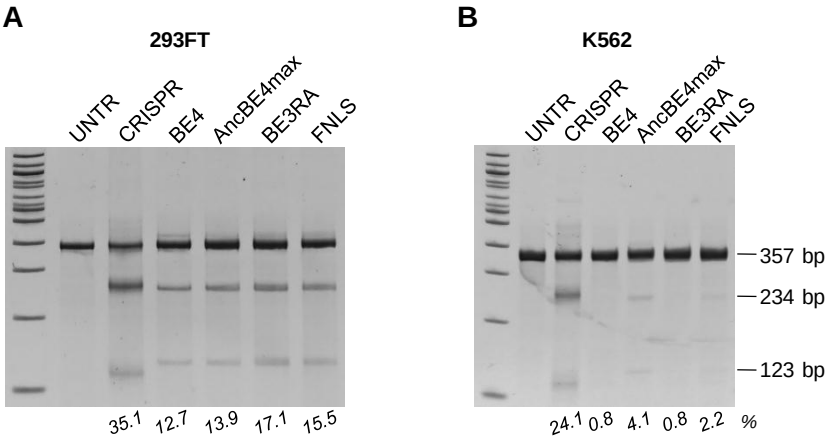

Figure S1. Comparison of different version of cytidine base editors. Cytidine base editors (BE4, AncBE4max, BE3RA, and FNLS) were subcloned under the control of a ubiquitously active EF1α promoter. A second plasmid pSP-sgBCL11AE1 expressing a guide sequence (5'-CACAGGCTCCAGGAAGGGTT-3') specific to the +58 BCL11A enhancer was used for co-transfection. **A**) Comparison of base editor activity in 293FT cells. The cells were transfected with base editor:pSP-sgBCL11AE1 (3µg:1µg) plasmids. The editing frequency at the *BCL11A* enhancer target site was analyzed 4 days after transfection by T7EI assay. The first lane is the 100bp DNA ladder from New England Biolabs. **B**) The same study was performed in an erythroleukemia cell line (K562). Expected sizes of the PCR amplicon and cleaved bands are indicated on the right. The percentages of cleavage are shown below the gel pictures.

A

Full name	Editor	Guide RNA	Yield (× 10^{12} vp)
HDAd-CBE-sgBCL#1	CBE	sgBCL#1	1.2
HDAd-CBE-sgHBG#1	CBE	sgHBG#1	0.8
HDAd-ABE-sgHBG#2	ABE	sgHBG#2	3.8
HDAd-ABE-sgHBG#4	ABE	sgHBG#4	2.8
HDAd-CBE-Dual#1	CBE	sgBCL#1 + sgHBG#1	0.5
HDAd-CBE-sgNeg	CBE	sgNeg	1.5

B

```
TadA^N+32aa   1 TCTGAAGTCGAGTTCAGCCACGAGTATTGGATGAGGCACGCACTGACCCTGGCAAAACGAGCCTGGGACG
TadA^N*+32aa  1 .....G..G.....TTC.........C.........A..T..C...........C..GA.G..AC.C..T.

TadA^N+32aa  71 AAAGAGAAGTCCCCGTCGGCGCCGTGCTGGTGCACAACAATCGCGTGATCGGAGAAGGATGGAACAGGCC
TadA^N*+32aa 71 .G..G..G..G..T..G..A.............TG......A.A........C..G..C........AG.

TadA^N+32aa 141 AATCGGACGCCACGACCCTACCGCACACGCAGAGATCATGGCACTGAGGCAGGGAGGACTGGTCATGCAG
TadA^N*+32aa 141 C.....C.TG........A..A..C..T..C..A..T.....C.....A.....C..C...........

TadA^N+32aa 211 AATTATCGCCTGATCGATGCCACCCTGTATGTGACACTGGAGCCATGCGTGATGTGCGCAGGAGCAATGA
TadA^N*+32aa 211 ..C..CA.A.....T..C...........C......T.C.....T.............C..C..C....

TadA^N+32aa 281 TCCACAGCCGGATCGGAAGAGTGGTGTTCGGAGCACGAGACGCCAAGACCGGAGCAGCAGGCAGCCTGAT
TadA^N*+32aa 281 .....TCTA.......CC.C........T..C.TGA.GA.T..A..A..T..C..C......TCT.....

TadA^N+32aa 351 GGATGTGCTGCACCACCCAGGCATGAACCACCGGGTGGAGATCACAGAGGGCATCCTGGCAGACGAGTGC
TadA^N*+32aa 351 ...C.........T....C........T...A.A..C..A..T..C.....A...........T..A..T

TadA^N+32aa 421 GCCGCCCTGCTGAGCGATTTCTTTAGAATGAGGAGACAGGAGATCAAGGCCCAGAAGAAGGCACAGAGCT
TadA^N*+32aa 421 ............T..T........C.G...CCT.......T.T.T..T..T...........C.......

TadA^N+32aa 491 CCACCGACTCTGGAGGCTCTAGCGGAGGATCCTCTGGAAGCGAGACACCAGGCACATCCGAGTCCGCCAC
TadA^N*+32aa 491 .......TAG...C..A........G..C........CTCT.....C..T......AG....AG...A..

TadA^N+32aa 561 ACCAGAGTCTTCCGGCGGCTCCTCCGGAGGATCC
TadA^N*+32aa 561 ...T..AAGCAG...G...AG.AG...G..GAG.
```

Figure S2. Virus preparations and alignment of TadA repeats after codon optimization. A) Information of generated HDAd vectors. Yields per 2L spinner were show. vp, viral particles.**B**) Sequence alignment of two 594bp TadA+32aa repeats after alternative codon usage to reduce repetitiveness. TadA^{N}+32aa is the first repeat; TadA^{N*}+32aa indicates the second repeat.

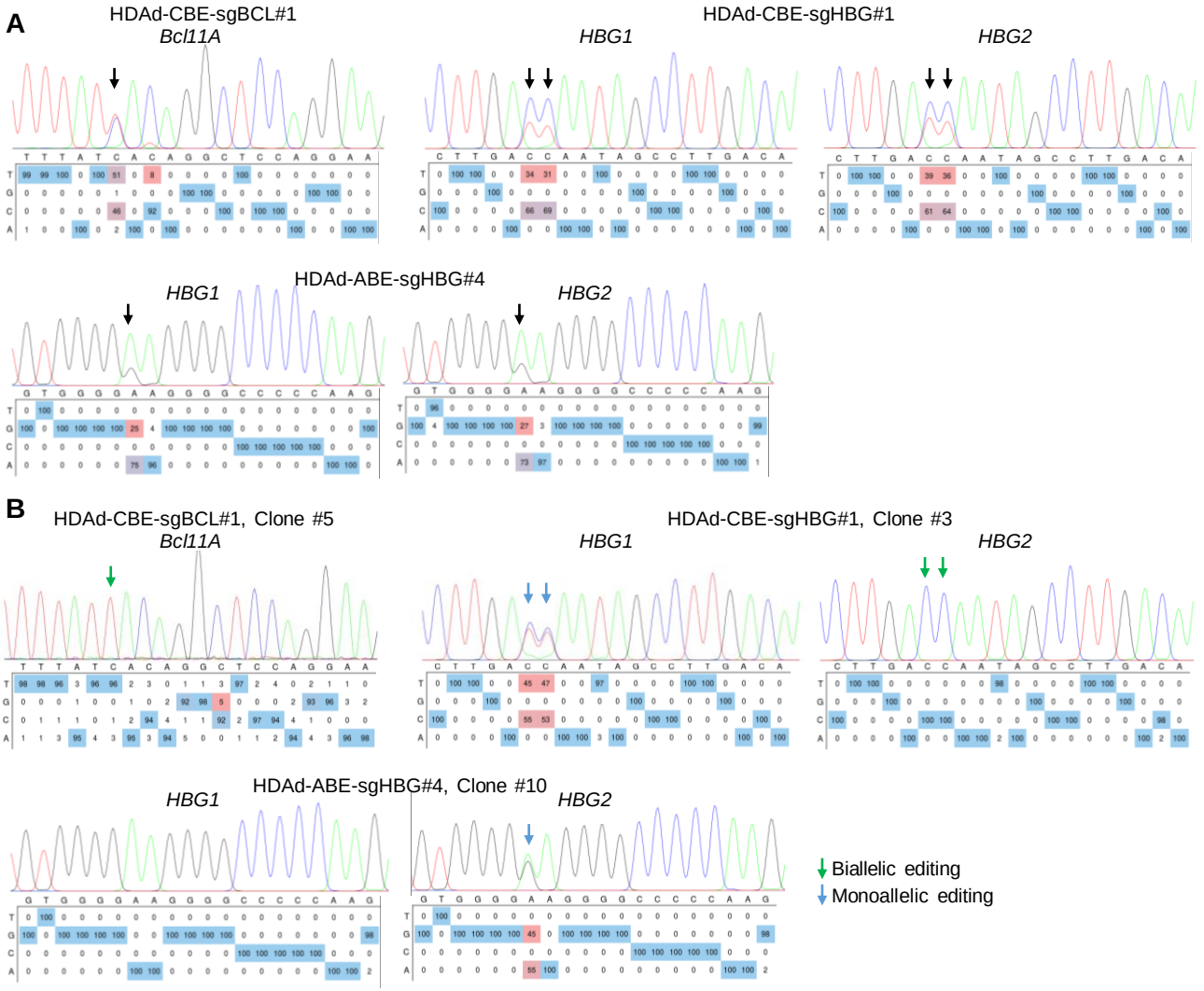

Figure S3. Base editing and γ-globin induction in HUDEP-2 cells (Data supporting Fig.2). A) Targeting base conversion at day 4 after transduction with HDAd-CBE-sgBCL#1, HDAd-CBE-sgHBG#1, and HDAd-ABE-sgHBG#4 (MOI = 1000vp/cell). The conversion was examined by Sanger sequencing. Data are supplementary to Fig.2D. **B)** Base substitutions in representative single cell-derived HUDEP-2 clones. The used vectors were indicated. Data are supplementary to Fig.2E. **C)** γ-globin expression in corresponding single cell HUDEP-2 clones in B). Data are supplementary to Fig.2F.

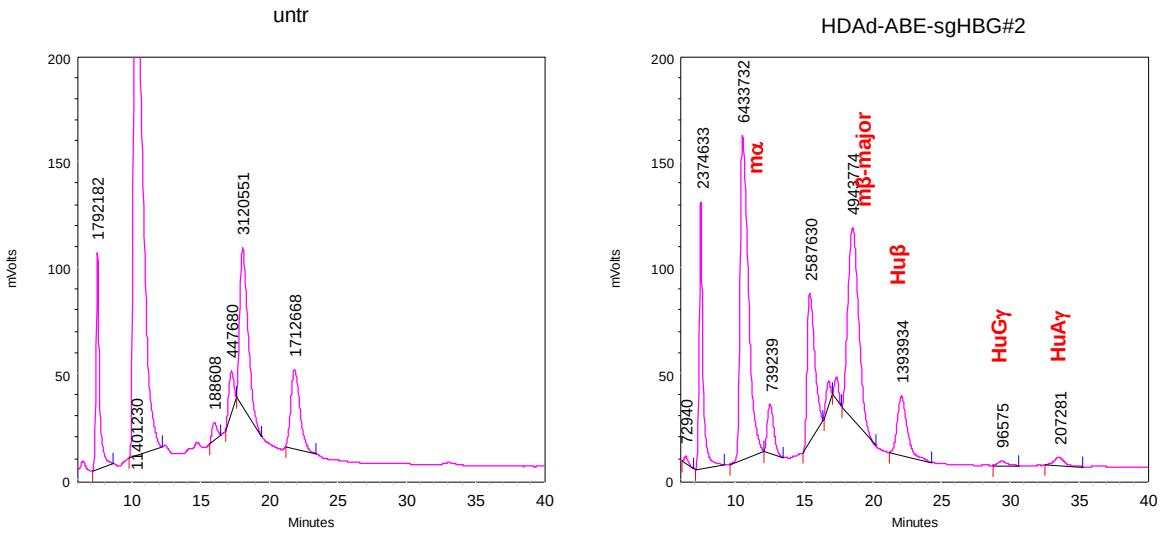

Figure S4. Representative HPLC chromatogram. Lysates from peripheral red blood cells at week 16 after transduction were used. untr, untransduced control.

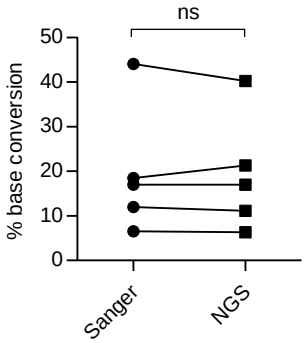

Figure S5. Comparison of editing levels measured by Sanger sequencing versus NGS. Editing levels at A8 site are shown. Each symbol represents one animal. ns, not significant.

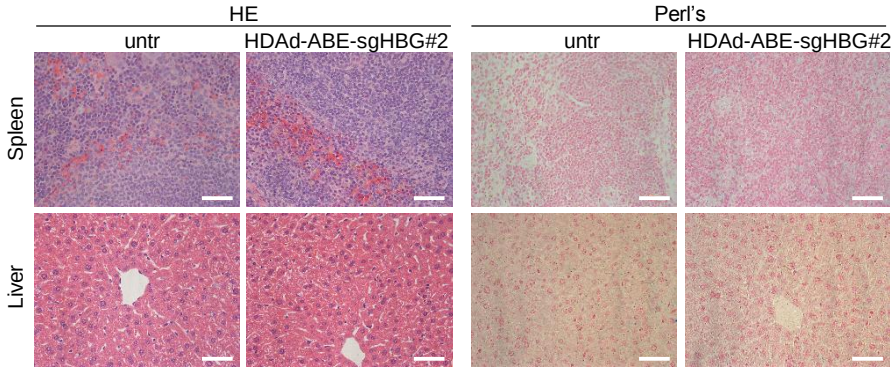

Figure S6. Histology analyses showing no aberrant extramedullary erythropoiesis after HDAd-ABE-sgHBG#2 treatment. Spleen and liver sections were prepared at week 16 to examine potential abnormal extramedullary hemopoiesis by hematoxylin/eosin (HE) staining, and iron deposition by Perl's staining. The scale bars are 20μm. Representative sections are shown. untr, untransduced control.

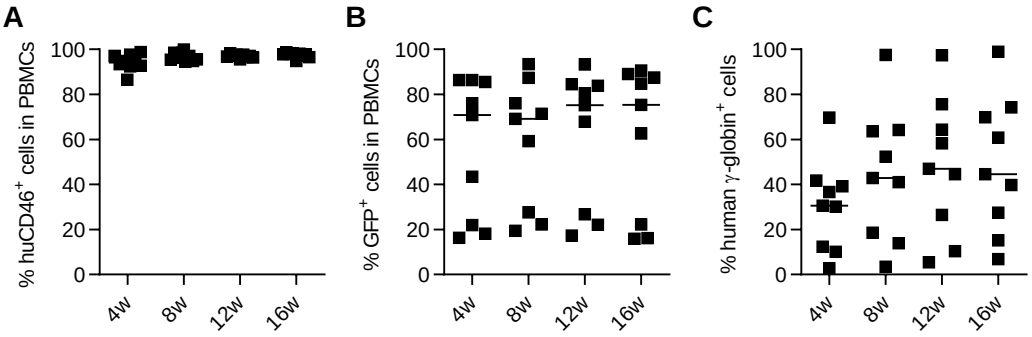

Figure S7. Secondary transplantation. Lineage minus cells collected from primary mice at week 16 after transduction were IV infused into lethally irradiated C57BL/6J secondary recipients. **A**) Engraftment measured by flow cytometry based on the percentage of human CD46-positive PBMCs at various weeks after transplantation. **B**) GFP expression in PBMCs. **C**) γ-globin expression in peripheral RBCs measured by flow cytometry. Each dot represents one animal.

Table S1. Top-scored off-target sites for sgHBG#2 in mouse genome.

ID	DNA sequence (5' to 3')	Chromosome	Position	Direction	Mismatches (bp)
mOTS1	gTTGACCAAgAGCCcTGACAAGG	chr6	145584801	-	3
mOTS2	CTTGACCAATAGCCagGAtATGG	chr5	128173360	+	3
mOTS3	CTTcACCAATAGCCTTtcaAAGG	chr8	4585505	+	4
mOTS4	CTTGACggATAaCCTTGtCATGG	chr8	48612946	-	4
mOTS5	CTTGACCAgcAcCCTgGACAGGG	chr8	120149968	-	4
mOTS6	CTTGtCCAAgtGCCTTGAaAAGG	chr12	71247526	+	4
mOTS7	CaTGACCAATAGtCTgaACATGG	chr12	100624810	-	4
mOTS8	CTTtACCAATAGCCTTttaAAGG	chr3	11687998	-	4
mOTS9	CTTGACtgATAaCCTTGACcAGG	chr3	63837608	-	4
mOTS10	CcTGAgCcATAGCtTTGACACGG	chr3	145100290	-	4
mOTS11	CTTGAgCcATAGCCcTGcCAGGG	chr7	50698141	+	4
mOTS12	CTTGAttAATAGCCgTGAtAAGG	chr7	114757537	-	4
mOTS13	CTacAtCAcTAGCCTTGACAGGG	chr7	128135370	-	4
mOTS14	tTaGACCAgTAtCCTTGACAGGG	chr7	137397027	-	4
mOTS15	CTTGACCAATAGCCTctcaAAGG	chr4	73893992	-	4
mOTS16	CTTGACCAATAGCCTctcaAAGG	chr4	73908833	-	4
mOTS17	CTTGACCAATAGCCTctcaAAGG	chr4	73879146	-	4
mOTS18	CTTGACCcATtGCCccGACACGG	chr5	36483628	-	4
mOTS19	aTTGcCCAATAaCCTTGgCAAGG	chr5	40909043	-	4
mOTS20	CTTGACCtATAGCgaTGcCAAGG	chr5	106443231	-	4
mOTS21	CTTGAtCtATAGCgTTGcCATGG	chr5	150058734	-	4
mOTS22	CTatACCcATtGCCTTGACATGG	chr16	3180108	+	4
mOTS23	CTatACCcATtGCCTTGACATGG	chr16	3286314	-	4
mOTS24	aTTGACCAcTAGCCTTGAatCGG	chr16	25083021	+	4
mOTS25	CTTGACCctTAGCCaTGgCAAGG	chr16	93325575	-	4
mOTS26	CTTcACCAATgGCCTTccCATGG	chr1	30384032	-	4
mOTS27	CTTGACCAATAatCcTGcCAAGG	chr1	38321903	+	4
mOTS28	CTgGACCAAcAGtCTTGgCATGG	chr1	64285756	-	4
mOTS29	CTTGgCaAATAGCaTgGACAGGG	chr1	87645017	-	4
mOTS30	CTgGAaCAcTAGCCTTGAgAGGG	chr1	94033869	-	4
mOTS31	aTTGACaAAaAGCCTTGAaATGG	chr1	169687353	+	4
mOTS32	CTTGACCAATgaCtTTGAtATGG	chr13	71283102	-	4
mOTS33	CTcGACCAATAcCCTTtgCAGGG	chr13	72640313	-	4
mOTS34	aTTGAaCAATttCCTTGACATGG	chr13	88281393	+	4
mOTS35	CaTGACtAATAGaCTgGACAGGG	chr2	65329762	+	4
mOTS36	tTaGACCAATAGCCTTtAaAGGG	chr2	67570093	+	4
mOTS37	CTTGgCaAATAGCtTTGgCAGGG	chr2	93194588	+	4
mOTS38	aTTGcCCAATAaCCTTGgCAGGG	chr2	135785682	-	4
mOTS39	CTTGACCAATgGggTTGgCAAGG	chr2	144936425	-	4
mOTS40	CTTGAgCAAaAGCCTgGAgAGGG	chr2	158949893	+	4
mOTS41	CTTGACattTAGCCTgGACATGG	chr2	159053053	+	4
mOTS42	CTgGACCAAgAGaCTaGACAAGG	chr2	159371546	-	4
mOTS43	gTaGAaCAATAGCCTTaACAAGG	chr15	34220561	-	4
mOTS44	CTTcACCAATAGCCTaGAatAGG	chr15	46262820	+	4
mOTS45	tTTcACCAAaAGCCTTaACAGGG	chr17	84741107	+	4
mOTS46	CTgGACCccTAGtCTTGACAAGG	chr10	78607651	-	4
mOTS47	CTTcACCAATAGCCTcaAgATGG	chr10	92363995	+	4
mOTS48	CTTGgCCAgTgaCCTTGACAAGG	chr10	105506916	-	4
mOTS49	CTTGgCtcATgGCCTTGACAGGG	chr6	52282642	+	4
mOTS50	CTTGcCtgATAaCCTTGACAGGG	chr6	87546505	+	4
mOTS51	gTTGAgCAAaAGCCTaGACATGG	chr6	99114295	+	4
mOTS52	gTTcACCAAcAGCCaTGACATGG	chr6	127761697	+	4
mOTS53	CTTGACCAAgAGCaTTcACtTGG	chr14	9189966	+	4
mOTS54	CTaGACCAATgGCCTTccCAGGG	chr14	23807134	-	4
mOTS55	gcTGAgCAATgGCCTTGACAAGG	chr14	71558408	-	4
mOTS56	CTTcACCAATgGCCTTtcCATGG	chr14	107589002	+	4
mOTS57	CTTGAaCAATgGCCTTcAgAAGG	chr9	51374420	-	4
mOTS58	CTTatCCAAaAGCCTTGAtATGG	chr9	59923230	+	4
mOTS59	CTTctCCAtTcGCCTTGACAGGG	chr9	60002413	+	4
mOTS60	tTTcACCAATAGCCTTtcCAGGG	chr9	72527429	+	4

mOTS61	CTTGACCAATgGCCccGACtAGG	chr9	108103769	-	4
mOTS62	aTgGACCAcTAGCCTTcACATGG	chr9	116730509	+	4
mOTS63	CTgGACaAAcAGCCTTGcCAGGG	chr9	120739525	+	4
mOTS64	CTaGACtcATAGaCTTGACAGGG	chrX	41531233	-	4
mOTS65	CTTGcCtcAcAGCCTTGACAAGG	chr18	27679671	+	4
mOTS66	CTTGACCAcTAttCTgGACATGG	chr18	35302193	+	4
mOTS67	CaaGACCAcTAGCCTTGAgATGG	chr11	10226829	-	4
mOTS68	tTTGgCCAtTAcCCTTGACATGG	chr11	32304533	-	4
mOTS69	CTTGACCttTtGgCTTGACAGGG	chr11	51933879	+	4
mOTS70	CTTGACaAcTgcCCTTGACAGGG	chr11	60527881	+	4
mOTS71	CTTGAtCAATgGCCTTaAgAAGG	chr11	71798128	-	4
mOTS72	aTTcACCAATAaCCTTGgCAGGG	chr11	78553092	-	4
mOTS73	CaTGAtCAATAGCaTTGcCAAGG	chr11	87234963	+	4
mOTS74	CTaGAgCAATAcCCTTcACAAGG	chr11	92762429	+	4
mOTS75	CTTGACtAATctCCcTGACAGGG	chr11	98595796	+	4
hOTS76	CTTcACCAATAGaCTcGtCATGG	chr11	119291713	-	4

Table S2. Top-scored off-target sites for sgHBG#2 in human genome.

ID	DNA sequence (5' to 3')	Chromosome	Position	Direction	Mismatches (bp)
hOTS1	CgTGACCAAaAGCCTTGtCATGG	chr9	132846597	-	3
hOTS2	CTTGgCCAgTAGCCTTGAtATGG	chr15	68936306	-	3
hOTS3	CTcGACCAAcAGCCcTGACAAGG	chr5	137486079	+	3
hOTS4	CTaGAaCAATAGCCTaGACATGG	chr5	13621061	+	3
hOTS5	CTTGACCgATAGCaTTGAaAAGG	chr7	89493088	+	3
hOTS6	tTTGACCAATAGCCTgGAaAAGG	chr7	18458968	-	3
hOTS7	CTTGACCAATAtCtTTGAgAAGG	chr3	28730447	-	3
hOTS8	CTTcACCAAgAGCCTTcACAGGG	chr12	116513739	-	3
hOTS9	CTTGAaCAAaAGCCTTGACgGGG	chr12	68668156	-	3
hOTS10	CTTGACCAtTAGCCcTtACAAGG	chr8	145029259	-	3
hOTS11	aTTGACaAAgAGtCTTGACAAGG	chr8	15857986	-	4
hOTS12	CTTGACCAATgGaCTgtACAAGG	chr8	117084520	-	4
hOTS13	CTTtACCAATAtCCTgGcCAGGG	chr12	38147787	-	4
hOTS14	CTTGcCCAAgAGCCTgGgCATGG	chr12	66373900	+	4
hOTS15	CTTGACCcAaAGCCcTGcCATGG	chr12	94180100	+	4
hOTS16	CcTGcCCAtTAGCCcTGACATGG	chr12	123023036	+	4
hOTS17	CTTGcCtAATtGCCTTGgCAAGG	chr3	5522774	+	4
hOTS18	gTTGAtgcATAGCCTTGACAAGG	chr3	28057261	+	4
hOTS19	CTTGcCCAATAGCtcTGACtAGG	chr3	49582639	+	4
hOTS20	CcTGAtCAgTAGCtTTGACAAGG	chr3	102225064	+	4
hOTS21	CTTGAgCAATAcCaTTGgCAGGG	chr3	115365952	-	4
hOTS22	CTTGAgCAAagGCCTTGAgATGG	chr7	4417433	-	4
hOTS23	CTTGACCAATtcCCTTGAagCGG	chr7	25949766	-	4
hOTS24	CTTctCCAATAGCCTTcACcTGG	chr7	35939966	+	4
hOTS25	CaTGcCCAtTAcCCTTGACAGGG	chr7	42246251	+	4
hOTS26	CTTctCCAATAGCCTTcACcTGG	chr7	45833861	-	4
hOTS27	CTTGACtAccAcCCTTGACAGGG	chr7	123673373	+	4
hOTS28	CTaGACCAATAGttTTGACtGGG	chr4	18523567	-	4
hOTS29	CTTGAgCAtTAGCtTTGgCATGG	chr4	35612229	-	4
hOTS30	CTTaACCAATAGCCTcaAaATGG	chr4	37072405	+	4
hOTS31	CTgtcCCAATAGCCTTaACATGG	chr4	72985779	+	4
hOTS32	CTTGAtCAATAGCCTcagCAAGG	chr4	172902563	+	4
hOTS33	CTTcACaAAaAGCCTTGtCAGGG	chr4	185469798	-	4
hOTS34	gTTGcCCAATAGCCTaGAtACGG	chr5	66721886	-	4
hOTS35	aTTGcCCAATAGCCTgGACtTGG	chr5	76136808	+	4
hOTS36	CTTGACCAAgAGCtTTGAtgAGG	chr16	5742590	-	4
hOTS37	CTTGAtCAtTAGCCcTtACAAGG	chr16	19397024	+	4
hOTS38	CTTGACCcATAGaaTTGAtATGG	chr16	22693308	+	4
hOTS39	CTTGAatAAgAGCCTTGACtAGG	chr16	27890144	-	4
hOTS40	CTgcACCAgcAGCCTTGACAGGG	chr16	85393040	-	4
hOTS41	aTTGACaAATtGtCTTGACAAGG	chr1	30388284	-	4
hOTS42	aTTctCCAATAaCCTTGACATGG	chr1	34226912	+	4
hOTS43	CTTGACCtATAtCCgTGAaATGG	chr1	89314094	+	4
hOTS44	CTTGcCCAATAGtCTatACAGGG	chr1	191365274	-	4
hOTS45	CTTGgCCAAgAGCCTgGcCACGG	chr13	98483238	+	4
hOTS46	CTTGACCAATActCTaGACcTGG	chr2	12091932	+	4
hOTS47	aTTGgCtAATAGCaTTGACATGG	chr2	18355208	-	4
hOTS48	CTTGACtAATAcaCTTtACATGG	chr2	34228360	+	4
hOTS49	CTTaACCAATAaCCcTaACAAGG	chr2	55183729	+	4
hOTS50	CTTGtCCAATtGCCTTGgCcAGG	chr2	117104607	+	4
hOTS51	CTTaACCAAaAGgaTTGACAAGG	chr2	151857147	+	4
hOTS52	tTTGACCAATAcCCTTGtgAGGG	chr2	220178975	+	4
hOTS53	CTTGcCCAgTgGCCTTtACAGGG	chr2	227407995	+	4
hOTS54	CTTGtCtAATtGCCTTGACtAGG	chr2	238074910	+	4
hOTS55	gTTGcCCAATAGCtcTGACAAGG	chr19	54795919	+	4
hOTS56	aTaGACCAcTAGCCTTGACcAGG	chr21	42945806	+	4
hOTS57	CTgGAggAATAGCCaTGACATGG	chr15	23693005	-	4
hOTS58	gTTGACCAATAGgCTaGAaAAGG	chr15	34820103	+	4
hOTS59	aTTGACaAcTAGCCTTcACAGGG	chr15	47374562	-	4
hOTS60	CTaGACtAAagGCCTTGACAAGG	chr15	53629113	+	4

hOTS61	CTTGcCCAtctGCCTTGACATGG	chr15	66926446	-	4
hOTS62	tTTGAtCAAcAGCCaTGACAGGG	chr15	73865957	+	4
hOTS63	aTTGACCAtTAGaCTTaACATGG	chr15	79665725	+	4
hOTS64	CTTGcCtAATAGCtcTGACAGGG	chr15	87284223	-	4
hOTS65	CTTGACaAATAcCCTTtgCATGG	chr15	88053418	-	4
hOTS66	gTTGACCAAaAGCCTTacCAAGG	chr15	100394792	-	4
hOTS67	CTTGAaCAATAGCaTTcAaATGG	chr17	29800385	+	4
hOTS68	CTTGgCCAATtaCCTaGACAAGG	chr17	40703221	-	4
hOTS69	tTgGACCAATAGCtTTGAgAGGG	chr17	76979384	-	4
hOTS70	CTTGAtCAATAGtCcTGACtAGG	chr10	58498724	+	4
hOTS71	tTTGACCAAcAGCgTgGACATGG	chr10	124619710	+	4
hOTS72	aTTGAtCAATAGtgTTGACATGG	chr10	125120841	-	4
hOTS73	CTTcAtaAATAGCCcTGACAAGG	chr6	78461199	+	4
hOTS74	CaTGgCCAAcAGCCTTGAaAAGG	chr6	132912174	-	4
hOTS75	CTTcACCAATtGCtcTGACATGG	chr14	84987298	+	4
hOTS76	CTTcAtCAAaAGCCcTGACAGGG	chr14	93381467	-	4
hOTS77	aTTGACaAATAGgCTgGACATGG	chr14	105879172	-	4
hOTS78	tTTGACCtATAGCCaTGgCATGG	chr20	12049844	+	4
hOTS79	CTTGAtCAATAGCagTGACtTGG	chr9	119378536	-	4
hOTS80	CTTGAgCAATAGCCTTGgttTGG	chrX	21083883	-	4
hOTS81	CTTGACCcATAcCaTTtACAAGG	chrX	35027287	-	4
hOTS82	CTTcAttAATAGCCTgGACAGGG	chrX	41666874	+	4
hOTS83	CTTcACaAATAGCCTTGtgAAGG	chrX	50280608	-	4
hOTS84	CTTGcCCAATtGCtcTGACAAGG	chr18	5129873	-	4
hOTS85	tcTGACCAAgAGCCTaGACAAGG	chr18	34643629	-	4
hOTS86	CTTGtCaAtTAGCCTgGACATGG	chr11	11571645	-	4
hOTS87	CTTGACCAccActCTTGACAGGG	chr11	35673822	+	4
hOTS88	CTTGACCAAgAGCtcTtACAGGG	chr11	76186235	+	4
hOTS89	CgTGACCtcTAcCCTTGACAGGG	chr11	93500809	+	4

Table S3. T7EI primers for mouse off-target sites.

Name	Sequence (5' to 3')	Amplicon size (bp)	T7EI cut (bp)
mOTS1_1F	AGAATTCCCCTTCCTCTTCCTATCCC	324	215
mOTS1_1R	AGACGACACCAGAGACCTGCTGATGT		109
mOTS2_1F	GGGGCATGGCGGTCTGAGAATCCTCTT	475	213
mOTS2_1R	AGGCTTGGGTTTGTCCTCAAGGCACAA		262
mOTS3_2F	CTCTGGCATATCAAACATCTTGGGGTAT	418	251
mOTS3_2R	GAAGGCATAAAGGTGTCATGGAGAGTAG		167
mOTS4_1F	GCAACATCTTCAGGTGTCAGGACAAGGC	386	261
mOTS4_1R	CACCAACACGTGGAATCCAGGATCCACT		125
mOTS5_2F	AGGGAAAAGGGATTAACAAACTCAGGAATAGC	421	190
mOTS5_2R	AGGGGTGGGAAAGGTGGAAGGGACAG		231
mOTS6_2F	GAATACTGAGTGAGTGGTGCCAAATGAG	434	277
mOTS6_2R	ACTGACGAGAGAAGGAAAACCTAACAGG		157
mOTS7_2F	AACCCTAACATTTCAGCCAGGAACTT	372	145
mOTS7_2R	AGAGTGGTTTTGGATGTCAGCAGCAT		227
mOTS8_2F	TCCCTCCCCACTCTCTCCCCTACACT	380	193
mOTS8_2R	AATTCTGGCATTCACACATGATCTTCTG		187
mOTS9_2F	TCAATCAACTGTGAATGTAGAGCAAAAA	454	149
mOTS9_2R	CAAAAGCAAGTCTCATAAACTGGGAAAT		305
mOTS10_1F	TGCTTGCTGCCATGCCTTCATATTTTAA	340	145
mOTS10_1R	ACAAGATGTTGCCATCTACGTGGTTCCA		195

Table S4. Oligos used for cloning.

ID	Name	Sequence (5' to 3')	Notes
#1	BE4_delBsmBI	TAGAGATCCGCGGCCGCTAATACGACTCACTATAGGGAGAGCCGCCACCATGAGC TCAGAGACTGGCCCAGTGGCTGTGGACCCCACATTGAGGCGGCGGATCGAGCCCC ATGAGTTTGAGGTATTCTTCGATCCGAGAGAGCTCCGCAAGGAGACCTGCCTGCT TTACGAAATTAATTGGGGGGGCCGGCACTCCATTTGGC	For cloning pBE4-delBsmBI
#2	AncBE4max_delBsmBI	TGGACCCAACCCTGAGGAGGCGGATTGAGCCCCATGAATTTGAAGTGTTCTTTGA CCCAAGGGAGCTGAGGAAGGAGACATGCCTGCTGTACGAGATCAAGTGGGGCACA AGCCACAAGATCTGGCGCCACAGCTCCAAGAA	For cloning pBE4max-delBsmBI
#3FR	U6gRNA-F	CGTTAATTAAGGATCCGAGGGCCTATTTCCCATGATTCC	For cloning pBST-CRISPR
	U6gRNA-R2	AAATCGATTAGGATCCGGCGCGCCCACCGCGGAAAAAAAGCACCGACTCGGTG	
#4FR	pBST_del BSMBI_F1	CGTGGGTCTCGCGGTATCATTGCAGCA	
	pBST_del BSMBI_R1	GAATCTCACTCTGTCGCCCAGGCTGGAGTGTA	
#5FR	pBST_del BSMBI_F2	GACAGAGTGAGATTCAGTCTCAAAACAAAC	
	pBST_del BSMBI_R2	GATCCTTAATTAACGAAGGAGCCATC	
#6FR	pBS-U6_F1	GCACACAGATTAATGGATCCGAGGGCCTATTTCCCATGATTCC	
	pBS-Ef1a_R1	CGGCCGATCGATTACTAGCTCACGACACCTGAAATG	
#7FR	pBS-bac_F2	GCGCGCTCACTGGCCGTCGTTTTACAACGTCGTGACT	
	pBS-bac_R2	CATTAATCTGTGTGCTTAATTAATTCGAAGGTACCCAGCTTTTGTTCC	
#8	pBS-SV40-3FR	GTAATCGATCGGCCGCTAATACGACAAGAATTCCTCCTCCCTAGGGCCACCTAGA CTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGAGGAGA TAAAAGACCATCACCATTGAGTTTTCTTTATTTTCATTAGATCTGTGTGTTGGTT TTTTGTGTGCGGCCGCAAGCTTCAGACATGATAAGATACATTGATGAGTTTGGAC AAACCACAACTAGAATGCAGTGAAAAAAATGCTTTATTTGTGAAATTTGTGATGC TATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTAACAACAACAAT TGCATTCATTTTATGTTTCAGGTTCAGGGGGAGGTGTGGGAGGTTTTTTAAAGCA AGTAAAACCTCTACAAATGTGGTACTCGAGATTAAGTGGCAGGATTAATTAATTC GAAGAGCTCCAATTCGCCCTATAGTGAGTCGTATTACGCGCGCTCACTGGCC	For cloning pBS-U6-Ef1α
#9FR	pBS to max_F	GGAAAGTCTAACCGGTCATCATCACC	For cloning pBS-BE4max and pBS-ABEmax, etc.
	pBS to max_R	GTCGTATTAGCGGCCGATCGATTACT	
#10FR	bGHpA_HM_F	AGCGGCCGTTTGTTATGTACAAGCCATAGAGCCCACCGCATCC	For cloning pHM5-FI-PGK-mgmt-GFP
	PGK_HM_R	AGCGGCCGTTTGTTAGGCCGGCCAATTCCCACGGGGTTGGGGTTG	
#11FR	pHM to pHCA_PshAI_F7	GGGCATATTGGACAGGGGTCACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAG	For cloning pHCAS3-FI-PGK-mgmtGFP-MCS
	pHM to pHCA_PshAI_R7	GAGCATAAGTGACCCCTGTCTTAATGCAGACCCATAATACCCATAATGCCATTTC ATTACC	
#12FR	3UTRmiR_NotI_F	TTTTTGTGTGCGGCCGCCCCCGGGCGATA	For cloning pBS-ABEmax-sgRNA-miR
	3UTRmiR_HpaI_R	GCAATTGTTGTTGTTAACTTGTTTATTGCA	
#13FR	hPGK to ABEopti_F	TTTTTCCGCGGTGGGCGCGCCGGATCCTAATCGATTTCCACGGGGTTGGGGTTGC G	
	hPGK to ABEopti_R	CCGATCGATTACTAGCCCCTGGGGAGAGAGGTCGGTG	

#14	gBlock1_TadAwt	CTAGTAATCGATCGGCCGCTAATACGACTCACTATAGGGAGAGCCGCCACCATGA AACGGACAGCCGACGGAAGCGAGTTCGAGTCACCAAAGAAGAAGCGGAAAGTCTC TGAAGTCGAGTTCAGCCACGAGTATTGGATGAGGCACGCACTGACCCTGGCAAAA CGAGCCTGGGACGAAAGAGAAGTCCCCGTCGGCGCCGTGCTGGTGCACAACAATC GCGTGATCGGAGAAGGATGGAACAGGCCAATCGGACGCCACGACCCTACCGCACA CGCAGAGATCATGGCACTGAGGCAGGGAGGACTGGTCATGCAGAATTATCGCCTG ATCGATGCCACCCTGTATGTGACACTGGAGCCATGCGTGATGTGCGCAGGAGCAA TGATCCACAGCCGGATCGGAAGAGTGGTGTTCGGAGCACGAGACGCCAAGACCGG AGCAGCAGGCAGCCTGATGGATGTGCTGCACCACCCAGGCATGAACCACCGGGTG GAGATCACAGAGGGCATCCTGGCAGACGAGTGCGCCGCCCTGCTGAGCGATTTCT TTAGAATGAGGAGACAGGAGATCAAGGCCCAGAAGAAGGCACAGAGCTCCACCGA CTCTGGAGGCTCTAGCGGAGGATCCTCTGGAAGCGAGACACCAGGCACATCCGAG TCCGCCACACCAGAGTCTTCCGGCGGCTCCTCCGGAGGATCCTCTGAGGTGGAGT TTTCCCACGAGTACTGGATGAGACATGCCCTGAC	For cloning pBS-ABEopti-sgRNA-miR
#15	gBlock2_TadA_envol	GAGACATGCCCTGACCCTGGCCAAGAGGGCACGCGATGAGAGGGAGGTGCCTGTG GGAGCCGTGCTGGTGCTGAACAATAGAGTGATCGGCGAGGGCTGGAACAGAGCCA TCGGCCTGCACGACCCAACAGCCCATGCCGAAATTATGGCCCTGAGACAGGGCGG CCTGGTCATGCAGAACTACAGACTGATTGACGCCACCCTGTACGTGACATTCGAG CCTTGCGTGATGTGCGCCGGCGCCATGATCCACTCTAGGATCGGCCGCGTGGTGT TTGGCGTGAGGAATGCAAAAACTGGCGCCGCAGGCTCTCTGATGGACGTGCTGCA CTACCCCGGCATGAATCACAGAGTCGAAATTACCGAGGGAATCCTGGCAGATGAA TGTGCCGCCCTGCTGTGCTATTTCTTTCGGATGCCTAGACAGGTGTTTAATGCTC AGAAGAAGGCCCAGAGCTCCACCGATAGTGGCGGATCTAGCGGGGGCTCCTCTGG CTCTGAGACCCCTGGCACAAGCGAGAGCGCAACACCTGAAAGCAGCGGGGGCAGC AGCGGGGGGAGCGACAAGAAGTACAGC	
#16FR	Cas to ABEopti_F Cas to ABEopti_R	GACAAGAAGTACAGCATCGGCC CAGGGTCAGCACGATATCTTCC	
#17	bGHpA_gBlock	GAGACATGCCCTGACCCTGGCCAAGAGGGCACGCGATGAGAGGGAGGTGCCTGTG GGAGCCGTGCTGGTGCTGAACAATAGAGTGATCGGCGAGGGCTGGAACAGAGCCA TCGGCCTGCACGACCCAACAGCCCATGCCGAAATTATGGCCCTGAGACAGGGCGG CCTGGTCATGCAGAACTACAGACTGATTGACGCCACCCTGTACGTGACATTCGAG CCTTGCGTGATGTGCGCCGGCGCCATGATCCACTCTAGGATCGGCCGCGTGGTGT TTGGCGTGAGGAATGCAAAAACTGGCGCCGCAGGCTCTCTGATGGACGTGCTGCA CTACCCCGGCATGAATCACAGAGTCGAAATTACCGAGGGAATCCTGGCAGATGAA TGTGCCGCCCTGCTGTGCTATTTCTTTCGGATGCCTAGACAGGTGTTTAATGCTC AGAAGAAGGCCCAGAGCTCCACCGATAGTGGCGGATCTAGCGGGGGCTCCTCTGG CTCTGAGACCCCTGGCACAAGCGAGAGCGCAACACCTGAAAGCAGCGGGGGCAGC AGCGGGGGGAGCGACAAGAAGTACAGC	For cloning pHM-FI-Ef1α-mgmt(P140K)-GFP-bGHpA

Supplementary materials and methods

Reagents for *in vivo* transduction and selection: G-CSF (Neupogen™) (Amgen, Thousand Oaks, CA), AMD3100 (MilliporeSigma, Burlington, MA), and Dexamethasone Sodium Phosphate (Fresenius Kabi USA, Lake Zurich, IL) were used. O^6-Benzylguanine (O^6-BG) and Carmustine (BCNU) were from Millipore/Sigma.

Generation of HDAd vectors: We used base editing systems developed by David R. Liu's lab at Harvard University[1]. pCMV_AncBE4max and pCMV_ABEmax plasmids were purchased from Addgene (Watertown, MA). The following plasmids from Addgene were also used: BE4, ABE7.10, pLenti-BE3RA-PGK-Puro, pLenti-FNLS-PGK-Puro and BE3RA (Fig.S1)[2]. The oligonucleotides and gBlocks described below were synthesized by Integrated DNA Technologies (IDT) (Coralville, IA) and listed in Table S4.

<u>*CBE and first version of ABE constructs:*</u> The cloning involved 3 steps. Step 1) The *BsmBI* site in BE4 was destroyed by replacing the *EagI-NaeI* fragment with gBlock #1. The *BsmBI* site in pCMV_AncBE4max was destroyed by replacing the *BsmBI-NarI* fragment with gBlock #2. A vector named pBST-CRISPR with a *BsmBI* sgRNA cloning site was generated by combining the following four fragments using infusion (Takara, Mountain View, CA): a 2.3kb U6-filler-gRNA scaffold fragment amplified from LentiCRISPRv2 (Addgene) using #3FR, two 1.4kb and 1.0kb fragments amplified from pBST-sgBCL11Ae1[3] using #4FR and #5FR, respectively, and a 9.6k fragment of pBST-sgBCL11Ae1 released by *BsaI-BamHI* digestion. An intermediate plasmid pBS-U6-Ef1α was composed by joining the following three fragments using infusion: a 3.6kb U6-filler-gRNA scaffold-Ef1α sequence and a 2.9kb vector backbone amplified from pBST-CRISPR using primers #6FR and #7FR, respectively, and a 0.5kb gBlock containing a *BseRI* cloning site (#8). This intermediate was digested with *BseRI* and recombined with the 5.5kb fragment of BE4-ΔBsmBI after *EagI-PmeI* treatment, generating pBS-BE4. A 6.6kb pBS backbone-U6-filler-gRNA scaffold-Ef1α sequence was PCR amplified from pBS-BE4 using #9FR, followed by infusion with *NotI-AgeI*-digested pCMV-ABEmax and pCMV_AncBE4max-ΔBsmBI, generating pBS-AncBE4max and pBS-ABEmax, respectively. Next, sgRNA oligos were synthesized, annealed, and inserted into the *BsmBI* site of pBS-BE4, pBS-AncBE4max and pBS-ABEmax, generating shutter plasmids with all-in-one base editing components, such as pBS-ABEmax-sgHBG#2. Step 2) A 21.0kb pHCAS3-MCS vector with *PacI* cloning site was generated similarly as described previously[4] except that the stuffer DNA was trimmed down by *EcoRI* restriction and re-ligation with the 1.8k *EcoRI* fragment. A 2.2kb PGK-mgmt^{P140K}-2A-GFP-bGHpolyA sequence was amplified from pHCA-Dual-mgmt-GFP[3] by #10FR and recombined with *PacI*-digested pHM5-FRT-IR-Ef1α-GFP[5], resulting in pHM5-FI-PGK-mgmt-GFP. Subsequently, the fragment between *I-CeuI* and *PI-SceI* sites was transferred from this construct to the *PshAI* site of pHCAS3-MCS by #11FR and infusion cloning, forming pHCAS3-FI-PGK-mgmt-GFP-MCS. Step 3) The shuttle plasmids from step 1 and the resultant vector from step 2 was treated with *PacI* and recombined to generate the final constructs, such as pHCA-ABEmax-sgHBG#2-FI-mgmt-GFP. Final pHCA contructs with different sgRNA sequences were generated similarly except that different sgRNA were used in step 1.

<u>*Second version of ABE constructs:*</u> The second version of ABE constructs differs from the first version in promoters, alternative codon usage, and miRNA-regulated gene expression. The cloning also involved 3 steps. Step 1) A 1.5kb 3' β-globin UTR with miR183/218 target sequence was amplified from pBST-sgHBG1-miR[3] using primers #12FR, followed by insertion into *NotI-HpaI* sites of pBS-ABEmax-sgHBG#2, generating pBS-ABEmax-sgHBG#2-miR. Shuttle plasmids for the second version of ABE constructs, for example, pBS-ABEopti-sgHBG#2-miR, were obtained by joining the following 4 fragments with *AscI-EcoRV*-digested pBS-ABEmax-sgHBG#2-miR by infusion cloning: a human PGK promoter amplified from pHM5-FI-PGK-mgmt-GFP using #13FR, two gBlocks (#14 and #15) containing the two *TadA* genes with alternative codon usage to reduce sequence repetitiveness, and a 1.9kb sequence amplified from pBS-ABEmax-sgHBG#2 using #16FR. Step 2) The SV40 polyA sequence between *PshAI-NotI* sites of pHM-FRT-IR-Ef1α-mgmt(P140K)-

2A-GFP-pA was replaced with a bGH polyA sequence (gBlock #17), getting pHM-FI-Ef1α-mgmt(P140K)-GFP-bGHpA. Then, the whole 4.9kb transposon between *I-CeuI* and *PI-SceI* sites was transferred to the *PshAI* site of pHCAS3-MCS using #11FR, generating pHCAS3-FI-Ef1α-mgmt-GFP-MCS. Step 3) The resultant constructs from step 1 and 2 were combined by infusion cloning following *PacI* treatment, generating pHCA-ABEopti-sgHBG#2-FI-mgmt-GFP. Final pHCA constructs with different sgRNA sequences were generated similarly.

The Phusion Hot Start II High-Fidelity DNA Polymerase (New England Biolabs) was used in all PCR amplifications involved in cloning. Final constructs were screened by several restriction enzymes (*HindIII*, *EcoRI* and *PmeI*) and confirmed by sequencing the whole region containing transgenes.

For the production of HDAd5/35++ vectors, corresponding pHCA plasmids were linearized with *PmeI* and produced in 116 cells[6] with AdNG163-5/35++, an Ad5/35++ helper vector containing chimeric fibers composed of the Ad5 fiber tail, the Ad35 fiber shaft, and the affinity-enhanced Ad35++ fiber knob[5]. HD-Ad5/35++ vectors were amplified in 116 cells as described in detail elsewhere[6]. Helper virus contamination levels were found to be <0.05%. Titers were 1-5x10^{12} viral particles (vp)/mL and yields were listed in Fig. S2A.

The HDAd-HBG-CRISPR and HDAd-SB vectors have been described previously[3,7].

Transfection of cell lines: 293FT (Thermo Fisher Scientific) and K562 cells (ATCC) were cultured according to the vendors' instructions. 293FT cells pre-seeded in 6-well plate were transfected with 4μg plasmids (3μg base editor or CRISPR/Cas9 + 1μg pSP-sgBCL11AE1[8]) using lipofectamine 3000 (Thermo Fisher Scientific) per the manufacturer's protocol. K562 cells were transfected with 2.66μg plasmids (2μg base editor or CRISPR/Cas9 + 0.6μg pSP-sgBCL11AE1) using nucleofection (Catalog # V4XC-2024) (Lonza, Basel, Switzerland) according to the provider's protocol. Genomic DNA was isolated at 4 days after transfection for analyses.

HUDEP-2 cells and erythroid differentiation: HUDEP-2 cells[9] were cultured in StemSpan SFEM medium (STEMCELL Technologies) supplemented with 100 ng/mL SCF, 3 IU/mL EPO, 10^{-6} M dexamethasone and 1 μg/mL doxycycline (DOX). Erythroid differentiation was induced in IMDM containing 5% human AB serum, 100 ng/mL SCF, 3 IU/mL EPO, 10 μg/mL Insulin, 330 μg/mL transferrin, 2 U/mL Heparin and 1 μg/mL DOX for 6 days.

CD34$^+$ cell culture. CD34$^+$ cells from G-CSF-mobilized adult donors were recovered from frozen stocks and incubated overnight in StemSpan H3000 medium (STEMCELL Technologies, Vancouver, Canada) supplemented with penicillin/streptomycin, Flt3 ligand (Flt3L, 25 ng/ml), interleukin 3 (10 ng/ml), thrombopoietin (TPO) (2 ng/ml), and stem cell factor (SCF) (25 ng/ml). Cytokines and growth factors were from Peprotech (Rocky Hill, NJ). CD34$^+$ cells were transduced with HDAd5/35++ vectors in low- attachment 12-well plates.

Differentiation of human HSPCs into erythroid cells was done based on the protocol developed by Douay et al.[10] In brief, in step 1, cells at a density of 10^4 cells/ml were incubated for 7 days in IMDM supplemented with 5% human plasma, 2 IU/ml heparin, 10μg/ml insulin, 330 μg/ml transferrin, 1 μM hydrocortisone, 100 ng/ml SCF, 5 ng/ml IL-3, 3 U/ml erythropoietin (Epo), glutamine, and penicillin/streptomycin. In step 2, cells at a density of 1x10^5 cells/ml were incubated for 3 days in IMDM supplemented with 5% human plasma, 2 IU/ml heparin, 10 μg/ml insulin, 330 μg/ml transferrin, 100 ng/ml SCF, 3 U/ml Epo, glutamine, and Pen/Strep. In step 3, cells at a density of 1x10^6 cells/ml cells were incubated for 12 days in IMDM supplemented with 5% human plasma, 2 IU/ml heparin, 10 μg/ml insulin, 330 μg/ml transferrin, 3 U/ml Epo, glutamine, and Pen/Strep.

Animal studies: β-YAC$^{+/-}$/CD46$^{+/+}$ mice were described in the main text. The following primers were used for genotyping of mice: *CD46* forward, 5'-GCCAGTTCATCTTTTGACTCTATTAA-3', and reverse, 5'-AATCACAGCAATGACCCAAA-3'; β-YAC (γ-globin promoter) forward, 5'-AAACGGTCCCTGGCTAAACT-3', and reverse, 5'-GCTGAAGGGTGCTTCCTTTTT-3'.

HSPC mobilization and in vivo transduction: HSPCs were mobilized in mice by subcutaneous (SC) injections of human recombinant G-CSF (5μg/mouse/day, 4 days) followed by an SC injection of AMD3100 (5mg/kg) on day 5. In addition, animals received Dexamethasone (10 mg/kg, IP) 16 h and 2 h before virus injection. Thirty and sixty minutes after AMD3100, animals were intravenously injected with virus vectors through the retro-orbital plexus with two doses of viruses ($4x10^{10}$ vp/dose × 2 doses). The base editing and SB viruses were co-delivered at a 1:1 ratio.

In vivo selection: Selection was started at 4 weeks after transduction. Mice were injected with O^6-BG (15mg/kg, IP) two times, 30 minutes apart. One hour after the second injection of O^6-BG, mice were injected (IP) with 5mg/kg BCNU. At week 6, 8, and 10, three more rounds were performed with BCNU doses at 7.5, 10, and 10mg/kg, respectively.

Secondary bone marrow transplantation: Recipients were female C57BL/6J mice, 6 – 8 weeks old from the Jackson Laboratory. On the day of transplantation, recipient mice were irradiated with 1000 Rad. Bone marrow cells from *in vivo* transduced CD46tg mice were isolated aseptically and lineage-depleted cells were isolated using MACS as described above. Six hours after irradiation cells were injected intravenously at $1x10^6$ cells per mouse. The secondary recipients were kept for 16 weeks after transplantation for terminal point analyses.

Transplantation of human CD34$^+$ cells: The immunodeficient NOD-*scid IL2rγ^{null}* (NSG) mice were obtained from the Jackson Laboratory (Bar Harbor, ME). NSG recipient mice received 300 Rad whole-body irradiation. $2.5×10^5$ whole bone marrow cells of non-irradiated NSG mice were mixed with $3x10^5$ human CD34$^+$ cells and injected intravenously into recipient mice at 4 hours post irradiation.

Tissue analyses: Spleen and liver tissue sections of 2.5 μm thickness were fixed in 4% formaldehyde for at least 24 hours, dehydrated and embedded in paraffin. Staining with hematoxylin-eosin was used for histological evaluation of extramedullary hemopoiesis. Hemosiderin was detected in tissue sections by Perl's Prussian blue staining. Briefly, the tissue sections were treated with a mixture of equal volumes (2%) of potassium ferrocyanide and hydrochloric acid in distilled water and then counterstained with neutral red.

Blood analyses: Blood samples were collected into EDTA-coated tubes and analysis was performed on a HemaVet 950FS (Drew Scientific, Waterbury, CT). Peripheral blood smears were stained with Giemsa/May-Grünwald (Merck, Darmstadt, Germany) for 5 and 15 minutes, respectively. Reticulocytes were stained with Brilliant cresyl blue. The investigators who counted the reticulocytes on blood smears have been blinded to the sample group allocation. Only animal numbers appeared on the slides (4 random $1cm^2$ sections loaded on 2 slides per animal).

Colony-forming unit (CFU) assay: Lineage minus (Lin$^-$) cells were isolated by depletion of lineage-committed cells in bone marrow MNCs using the mouse lineage cell depletion kit (Miltenyi Biotec, San Diego, CA) according to the manufacturer's instructions. CFU assays were performed using ColonyGEL 1202 (Reachbio, Seattle, WA) with mouse complete medium according to the manufacturer's protocol. Colonies were scored 10 days after plating. For human CD34$^+$ cells, ColonyGEL 1102 (Reachbio, Seattle, WA) with human complete medium was used. Colonies derived from human HSPCs were counted at day 14.

T7EI mismatch nuclease assay: The T7EI assay is used for evaluating gene editing activities. Genomic DNA was isolated using PureLink Genomic DNA Mini Kit per provided protocol (Life Technologies, Carlsbad, CA)[11]. A genomic segment encompassing the target site of erythroid *BCL11A* enhancer was amplified by PCR primers: *BCL11A*_T7 forward, 5'-AGAGAGCCTTCCGAAAGAGG-3', reverse, 5'-GGCAGCTAGACAGGACTTGG-3'; *HBG*_T7 forward, 5'-CAGGGTTTCTCCTCCAGCATCTTCCACAT-3', reverse, 5'-AGCAGCAGTATCCTCTTGGGG-3'. Primers for off-target analysis were listed in Table S3. PCR products were hybridized and treated with 2.5 Units of T7EI (New England Biolabs) for 30 minutes at 37°C. Digested PCR products were resolved by 6% TBE PAGE (Bio-Rad) and stained with ethidium bromide. 100bp DNA Ladder (New England Biolabs) was used. Band intensity was analyzed using ImageJ software. % cleavage = (1-sqrt (parental band/(parental band + cleaved bands)) $\times$ 100%.

Measurement of base conversion by Sanger sequencing and Next-Generation Sequencing (NGS): For Sanger sequencing, genomic segments encompassing the target sites were amplified using primers: *HBG1* forward, 5'-CACACTCCACACTTTTTTGTTTAC-3', reverse, 5'-AAGTGTCTTTACTGCTTTTATTTGCT-3'; *HBG2* forward, 5'-TCCTTCTGTCATTTTGCCTCTGTT-3', reverse, 5'-CACTTCATTGTAGTTACCGTGGAAAGA-3'; *BCL11A* forward, 5'-AGAGAGCCTTCCGAAAGAGG-3', reverse, 5'-GGCAGCTAGACAGGACTTGG-3'. The amplicons were purified and sequenced with the following primers: *HBG1-seq*, 5'-TTTCCTTAGAAACCACTGCTAACTG-3'; *HBG2-seq*, 5'-CTTATTTGGAAACCAATGCTTACTA-3'; and *BCL11A-seq*, 5'-AGAGAGCCTTCCGAAAGAGG-3'. The base editing level was quantified based on Sanger sequencing (Eurofins Genomics) results by using EditR1.0.9[12].

For NGS, the HBG1/2 target site was amplified using the following primers: HBG-NGS forward, 5'-AGCCTTGTCCTCCTCTGTGA-3', reverse, 5'-AAACGGTCCCTGGCTAAACT-3'. After cleaning-up the amplicon using AMPure XP Beads (Beckman Coulter, Indianapolis, IN), libraries were prepared using the NEBNext Ultra II (New England Biolabs) workflow using End Repair/dA-Tailing, Ligation modules from IDT for Illumina TruSeq UD Indexes. Final libraries were purified using NEBNext sample purification beads, and quantified using Qubit (Invitrogen). Library sizes were confirmed on a gel. Purified libraries were pooled at equimolar concentration and deep sequenced on an Illumina NextSeq 550 machine. Approximately 500 thousand reads per amplicon were acquired to probe the types of mutations. Sequencing data were demultiplexed based on index sequence and filtered for high quality reads (q score > 30) using Trim Galore and aligned to the *HBG1/2* reference sequence using the CRISPResso2[13], a python-based genome editing analysis tool.

Flow cytometry: Cells were resuspended at 1x10^6 cells/100 µL in FACS buffer (PBS, 1%FBS) and incubated with FcR blocking reagent (Miltenyi Biotech, Auburn CA) for ten minutes on ice. Next the staining antibody solution was added in 100 µL per 10^6 cells and incubated on ice for 30 minutes in the dark. After incubation, cells were washed once in FACS buffer. For secondary staining, the staining step was repeated with a secondary staining solution. After the wash, cells were resuspended in FACS buffer and analyzed using a LSRII flow cytometer (BD Biosciences, San Jose, CA). Debris was excluded using a forward scatter-area and sideward scatter-area gate. Single cells were then gated using a forward scatter-height and forward scatter-width gate. Flow cytometry data were then analyzed using FlowJo (version 10.0.8, FlowJo, LLC). For analysis of LSK cells, cells were stained with biotin-conjugated lineage detection cocktail (catalog # 130-092-613) (Miltenyi Biotec, San Diego, CA), antibodies against c-Kit (clone 2B8, catalog # 12-1171-83) and Sca-1 (clone D7, catalog # 25-5981-82), followed by secondary staining with APC-conjugated streptavidin (catalog # 17-4317-82) (eBioscience, San Diego, CA). Other antibodies from eBioscience included anti-mouse CD3-APC (clone 17A2) (catalog # 17-0032-82), anti-mouse CD19-PE-Cyanine7 (clone eBio1D3) (catalog # 25-0193-82), and anti-mouse Ly-66 (Gr-1)-PE, (clone RB6-8C5) (catalog # 12-5931-82). Anti-mouse Ter-119-APC (clone Ter-119) (catalog # 116211) was from Biolegend (San Diego, CA). Anti-human CD45-APC (clone: 5B1) was from Miltenyi Biotec (Catalog # 130-108-020).

For intracellular flow cytometry detecting human γ-globin expression, the FIX & PERM™ cell permeabilization kit (Thermo Fisher Scientific) was used and the manufacture's protocol was followed. ~5×10^6 cells or 5μL blood was stained with 0.6μg anti-human γ-globin antibody (Clone 51-7, catalog# sc-21756-PE) (Santa Cruz Biotechnology, Dallas, TX). Mouse RBCs were stained with anti-Ter-119-APC antibody prior to γ-globin staining.

Globin HPLC: Lysates prepared from mouse RBCs or differentiated HUDEP-2 cells were used. Individual globin chain levels were quantified on a Shimadzu Prominence instrument with an SPD-10AV diode array detector and an LC-10AT binary pump (Shimadzu, Kyoto, Japan). Vydac 214TP™ C4 Reversed-Phase columns for polypeptides (214TP54 Column, C4 ,300 Å, 5 μm, 4.6 mm i.d. x 250 mm) (Hichrom, UK) were used. A 40%-60% gradient mixture of 0.1% trifluoroacetic acid in water/acetonitrile was applied at a rate of 1 mL/min.

Real-time reverse transcription PCR: Total RNA was extracted from 5×10^6 differentiated HUDEP-2 cells or 100μL blood by using TRIzol™ reagent (Thermo Fisher Scientific) followed by phenol-chloroform extraction. QuantiTect reverse transcription kit (Qiagen) and power SYBR™ green PCR master mix (Thermo Fisher Scientific) were used. Real time quantitative PCR was performed on a StepOnePlus real-time PCR system (AB Applied Biosystems). The following primer pairs were used: mouse *RPL10* (house-keeping) forward, 5'-TGAAGACATGGTTGCTGAGAAG-3', reverse, 5'-GAACGATTTGGTAGGGTATAGGAG-3'; human γ-globin forward, 5'-GTGGAAGATGCTGGAGGAGAAA-3', reverse, 5'-TGCCATGTGCCTTGACTTTG-3'; human β-globin forward, 5'-CTCATGGCAAGAAAGTGCTCG-3', reverse, 5'-AATTCTTTGCCAAAGTGATGGG-3'; mouse β-major globin forward, 5'- ATGCCAAAGTGAAGGCCCAT-3', reverse, 5'-CCCAGCACAATCACGATCAT-3', mouse α globin forward, 5'- CTGGGGAAGACAAAAGCAAC -3', reverse, 5'-GCCGTGGCTTACATCAAAGT -3.

Measurement of vector copy number: For absolute quantification of adenoviral genome copies per cell, genomic DNA was isolated from cells using PureLink Genomic DNA Mini Kit per provided protocol (Life Technologies), and used as template for qPCR performed using the power SYBR™ green PCR master mix (Thermo Fisher Scientific). The following primer pairs were used: mgmt forward, 5'-GCTGTCTGGTTGTGAGCAGGGTCT-3', reverse, 5'-CGGGCTGGTGGAAATAGGCATTC-3'.

Statistical analyses*:* Statistical significance was calculated by two-tailed Student's t-test. For comparisons of multiple groups, one-way and two-way analysis of variance (ANOVA) with Bonferroni post-testing for multiple comparisons were employed. Statistical analysis was performed using GraphPad Prism version 8.4.2 (GraphPad Software Inc., La Jolla, CA). P values <0.05 were considered statistically significant.

References

1. Koblan LW, Doman JL, Wilson C, et al. Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. *Nat Biotechnol*. 2018;36(9):843–846.
2. Zafra MP, Schatoff EM, Katti A, et al. Optimized base editors enable efficient editing in cells, organoids and mice. *Nat Biotechnol*. 2018;36(9):888-893.
3. Li C, Psatha N, Sova P, et al. Reactivation of gamma-globin in adult beta-YAC mice after *ex vivo* and *in vivo* hematopoietic stem cell genome editing. *Blood*. 2018;131(26):2915–2928.
4. Li C, Course MM, McNeish IA, Drescher CW, Valdmanis PN, Lieber A. Prophylactic *in vivo* hematopoietic stem cell gene therapy with an immune checkpoint inhibitor reverses tumor growth in syngeneic mouse tumor models. *Cancer Res*. 2020;80(3):549-560.

5. Richter M, Saydaminova K, Yumul R, et al. *In vivo* transduction of primitive mobilized hematopoietic stem cells after intravenous injection of integrating adenovirus vectors. *Blood*. 2016;128(18):2206-2217.
6. Palmer D, Ng P. Improved system for helper-dependent adenoviral vector production. *Mol Ther*. 2003;8(5):846-852.
7. Wang H, Richter M, Psatha N, et al. A Combined In Vivo HSC Transduction/Selection Approach Results in Efficient and Stable Gene Expression in Peripheral Blood Cells in Mice. *Mol Ther Methods Clin Dev*. 2018;8:52-64.
8. Li C, Psatha N, Gil S, Wang H, Papayannopoulou T, Lieber A. HDAd5/35++ adenovirus vector expressing anti-CRISPR peptides decreases CRISPR/Cas9-related toxicity in human hematopoietic stem cells. *Mol Ther Methods Clin Dev*. 2018;9:390-401.
9. Kurita R, Suda N, Sudo K, et al. Establishment of immortalized human erythroid progenitor cell lines able to produce enucleated red blood cells. *PLoS One*. 2013;8(3):e59890.
10. Douay L, Giarratana MC. Ex vivo generation of human red blood cells: a new advance in stem cell engineering. *Methods Mol Biol*. 2009;482:127-140.
11. Miller JC, Holmes MC, Wang J, et al. An improved zinc-finger nuclease architecture for highly specific genome editing. *Nat Biotechnol*. 2007;25(7):778-785.
12. Kluesner MG, Nedveck DA, Lahr WS, et al. EditR: A Method to Quantify Base Editing from Sanger Sequencing. *CRISPR J*. 2018;1:239-250.
13. Clement K, Rees H, Canver MC, et al. CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat Biotechnol*. 2019;37(3):224-226.