Supplemental data

Efficient repair of human genetic defect by CRISRP/Cas9 mediated interlocus gene conversion

Fei Yang[1, 5, #], Yiyun Wang[4, #], Qiudao Wang[1], Jingtao Pang[5], Guolong Liu[1], Yang Yang[6], Shenguang Qin[5], Ying Zhang[1], Yongrong Lai[6], Bin Fu[7], Yating Zhu[2, 3], Mengyao Wang[2, 3], Ryo Kurita[8], Yukio Nakamura[9], Dan Liang[2, 3, *] & Yuxuan Wu[1,5, *]

Supplementary Figures

Supplementary Figure 1. GC in homozygous mutant cells via the nearby HBD during homology repair.

(A) Summary of the most frequent indels by deep sequencing following Cas9 RNP *HBD-4142* editing of homozygous β^0 4142 $CD34^+$ HSPCs (HBD^{KO}).

(B) Summary of deep sequencing data derived from Cas9 RNP (coupled with sgRNA-1)-edited homozygous β^0 4142 $CD34^+$ HSPCs (HBD^{KO}). The arrows indicate unedited alleles. The same genotypes with measurable conversions were in the same color.

Supplementary Figure 2. Indel spectra of engrafted BM correcting β^0 4142 through GC.

(A, B) Summary of the most frequent indels by deep sequencing of input cells and corresponding bone marrow cells from the primary recipient. (A) Restoration by deep sequencing following Cas9 RNP editing of Donor #5 (HT). (B) Restoration by deep sequencing following Cas9 RNP editing of Donor #1 (HM). HT, HM mock, heterozygote and homozygous β^0 4142-thalassemia patient $CD34^+$ HSPCs; Input, deep sequencing following Cas9 RNP editing of patient $CD34^+$ HSPCs before transplantation; BM, engrafted edited $CD34^+$ HSPCs analyzed 16 weeks after transplantation. -TCTT indicates unedited allele; WT remarks in red indicate normal wild-type alleles; Green indicates other most frequent indels.

(C) RT–qPCR analysis showed that the expression of β-globin was rescued in 4142 deletion HSPCs edited with sgRNA-1 that underwent GC. Error bars indicate the standard deviation (n= 3 replicates).

Supplementary Figure 3. GC- and HR-associated proteins.

(A-C) Comparing homologous recombination (HR) and GC when inhibiting HR-related genes (sgRNAs; see details in Table S1) in the homozygous Donor #1 and HUDEP-2 cell lines. After 72 hours, the cells were electroporated with RNP using sgRNA-1 targeting the mutant site with or without exogenous DNA templates. Data are plotted as the mean ± s.d. and analyzed using unpaired two-tailed Student's t tests. NS, not significant. Data are representative of three biologically independent replicates.

(D) The *HBD* genotype of the HUDEP-2 cell line.

(E) Summary of the most frequent indels by deep sequencing of sgRNA-1 RNP targeting the *HBB-4142* site in the HUDEP-2 cell line. Genotype of HBD^- remarks in black box indicates GC restorations.

Supplementary Figure 4. Off-target analysis edited by Cas9 RNP targeting sgRNA-1 conversion of *HBB*-4142 deletion to normal in homozygous and heterozygous patient donors.

(A) Using the CasOFFinder tool, 21 potential genomic off-target sites with 3 or fewer mismatches to the on-target sgRNA-1 sequence were evaluated by amplicon deep sequencing. The on-target sequence is at the *HBB-4142* site. The dotted line at 0.1% denotes the sensitivity of deep sequencing to detect indels

(B) .RT–qPCR analysis of *p21* expression after gene editing. Relative expression to β-actin is shown. Error bars indicate the standard deviation (n= 3 replicates).

Supplementary Figure 5. GC to correct multiple types of mutations in the coding region *HBB*.

(A) β-globin expression by RT–qPCR analysis in erythroid cells in vitro differentiated from RNP-edited β-thalassemia patient donors. Error bars indicate the standard deviation (n= 3 replicates).

(B) Summary of the restorations by deep sequencing following Cas9 RNP editing of CD7172. The genotypes with measurable conversion from *HBD* are labeled in red.

(C) Summary of the restorations by deep sequencing following Cas9 RNP editing of CD17.

(D) Summary of the restorations by deep sequencing following Cas9 RNP editing of IVS1-1. WT marked in red indicates normal wild-type allele; mutation sites marked in black indicate unedited allele.

Supplemental Table 1. Sequence of sgRNAs and DNA templates

sgRNAs	5'-3'	Target regions
sgRNA-1	UCCCCAAAGGACUCAACCUC	Targeting HBB-4142
sgRNA-2	CCCCAAAGGACUCAACCUCU	Targeting HBB-4142
sgRNA-3	GACCCAGAGGUUGAGUCCUU	Targeting HBB-4142
sgRNA-4	ACCCAGAGGUUGAGUCCUUU	Targeting HBB-4142
sgRNA-5	CCCAGAGGUUGAGUCCUUUG	Targeting HBB-4142
sgRNA-6	GGACUCAACCUCUGGGUCCA	Targeting HBB-4142
sgRNA-7	GACUCAACCUCUGGGUCCAA	Targeting HBB-4142
HBD4142 ^{KO}	CAAAGGACUCAAAGAACCUC	Targeting HBB/HBD WT4142 sites
HU-HBD^{-}	CCCAAAGGACUCAAAGAACC	Targeting HBD^{-} mutant site in HUDEP-2 cell line
BRCA1^{KO}	AGCAGUAUUUCAUUGGUACC	Targeting BRCA1 gene
BRCA2^{KO}	UCUACUAUAUUAGAAGAAUC	Targeting BRCA2 gene
ATM^{KO}	CCUUUAGGGCAGCUGAUAUU	Targeting ATM gene
PALB2^{KO}	CUUGGAUGAUGAUGCUUUCA	Targeting PALB2 gene
Rad54^{KO}	UAAGGCUUUUAAGUCUGUGU	Targeting Rad54 gene
sgIVS1-1	UGGUGAGGCCCUGGGCAGUU	Targeting IVS I-1 G>T in Donor #2
sgCD17	CGUUACUGCCCUGUGGGGCU	Targeting CD17 A>T in Donor #5
sgHbE	CGUGGAUGAAGUUGGUGGUA	Targeting HbE G>A in Donor #6
sgCD7172	UGAGCCAGGCCAUCACUUAA	Targeting CD7172 +A in Donor #7
sgSCD	GUAACGGCAGACUUCUCCAC	Targeting Sickle mutation A>T in SCD cell line
DNA templates	**5'-3'**	

ssODN	TCCCACCCTTAGGCTGCTGGTGGTCTACCCTTGGACCCAGAGGTTCTTTGAGTCCTTTGGGGATCTGTCCACTCCTGATGCTGTTATGGG
ssODN-R	CCCATAACAGCATCAGGAGTGGACAGATCCCCAAAGGACTCAAAGAACCTCTGGGTCCAAGGGTAGACCACCAGCAGCCTAAGGGTGGGA
ssODN-P	TCCCACCCTTAGGCTGCTGGTGGTCTACCCTTGGACACAGAGGTTCTTTGAGTCCTTTGGGGATCTGTCCACTCCTGATGCTGTTATGGG
ss-AmC6	TCCCACCCTTAGGCTGCTGGTGGTCTACCCTTGGACCCAGAGGTTCTTTGAGTCCTTTGGGGATCTGTCCACTCCTGATGCTGTTATGGG
ss-AmC12	TCCCACCCTTAGGCTGCTGGTGGTCTACCCTTGGACCCAGAGGTTCTTTGAGTCCTTTGGGGATCTGTCCACTCCTGATGCTGTTATGGG

Note：

Letters in Green represent 4 bp deletion sites

Lerrer in Red represents synonymous mutations in PAM of ssODN-P.

ss-AmC6 represents 5′ modifications including an amine group with a C6 linker (AmC6)

ss-AmC12 represents 5′ modifications including an amine group with a C12 linker (AmC12)

Supplemental Table 2. Genotypes of β-thalassemia patient donors and cell lines

Donor ID	Genotype information		Sequence information
1	$\beta^{4142}\beta^{4142}$	β^0 β^0	Homozygous for Codon 41/42 (-TCTT) CD4142: TGGTGGTCTACCCTTGGACCCAGAGGT- - - -TGAGTCCTTTGGGGATCTGTCCACTCCTGAT
2	$\beta^{4142}\beta^{IVS\ I-1}$	β^0 β^0	Compound heterozygous for Codon 41/42 (-TCTT) and IVS I-1 G>T CD4142: TGGTGGTCTACCCTTGGACCCAGAGGT- - - -TGAGTCCTTTGGGGATCTGTCCACTCCTGAT IVS I-1 G>T: GGATGAAGTTGGTGGTGAGGCCCTGGGCAG**T**TTGGTATCAAGGTTACAAGACAGGTTT
3	$\beta^{4142}\beta^{-28}$	β^0 β^+	Compound heterozygous for Codon 41/42 (-TCTT) and -28 (A>G) (promoter TATA box) CD4142: TGGTGGTCTACCCTTGGACCCAGAGGT- - - -TGAGTCCTTTGGGGATCTGTCCACTCCTGAT -28: GGAGGGCAGGAGCCAGGGCTGGGCATA**G**AAGTCAGGGCAGAGCCATCTATTGCTTACATTTG
4	$\beta^{4142}\beta^{-29}$	β^0 β^+	Compound heterozygous for Codon 41/42 (-TCTT) and -29 (A>G) (promoter TATA box) CD4142: TGGTGGTCTACCCTTGGACCCAGAGGT- - - -TGAGTCCTTTGGGGATCTGTCCACTCCTGAT -29: GGAGGGCAGGAGCCAGGGCTGGGCAT**G**AAAGTCAGGGCAGAGCCATCTATTGCTTACATTTG
5	$\beta^{4142}\beta^{CD17}$	β^0 β^0	Compound heterozygous for Codon 41/42 (-TCTT) and Codon 17 (A>T; AAG>TAG; Lys>stop codon) CD4142: TGGTGGTCTACCCTTGGACCCAGAGGT- - - -TGAGTCCTTTGGGGATCTGTCCACTCCTGAT CD17: GAAGTCTGCCGTTACTGCCCTGTGGGGC**T**AGGTGAACGTGGATGAAGTTGGTGGTGAGGCCC
6	$\beta^{Hb\ E}\beta^{-28}$	β^+ β^+	Compound heterozygous for Hb E (G>A; GAG>AAG) and -28 (A>G) (promoter TATA box) Hb E: GGCAAGGTGAACGTGGATGAAGTTGGTGGT**A**AGGCCCTGGGCAGGTTGGTATCAAGGTTACA -28: GGAGGGCAGGAGCCAGGGCTGGGCATAGAAGTCAGGGCAGAGCCATCTATTGCTTACATTTG
7	$\beta^{7172}\beta^{4142}$	β^0 β^0	Compound heterozygous for Codon 7172 (+A; frameshift mutation) and Codon 41/42 (-TCTT) CD4142: TGGTGGTCTACCCTTGGACCCAGAGGT- - - -TGAGTCCTTTGGGGATCTGTCCACTCCTGAT CD7172: CTCATGGCAAGAAAGTGCTCGGTGCCTTT**(A)**AGTGATGGCCTGGCTCACCTGGACAACCT
n/a	$HU4142^{del}HBD^-$	β^0	Homozygous HUDEP-2 cell line for Codon 41/42 (-TCTT), Homozygous for HBD indels (-CAGA +G)

		β^0	CD4142: TGGTGGTCTACCCTTGGACCCAGAGGT- - - -TGAGTCCTTTGGGGATCTGTCCACTCCTGAT HBD$^-$: CTGGTGGTCTACCCTTGGACC- - **G** - GGTTCTTTGAGTCCTTTGGGGATCTGTCC
n/a	SCD Cell line	β^S	Homozygous SCD cell line for SCD (A>T; GAG>GTG) Sickle mutation SCD: AAACAGACACCATGGTGCATCTGACTCCTG**T**GGAGAAGTCTGCCGTTACTGCCCTGTGGGGC
		β^S	

Supplemental Table 3. Off-target sites analyzed and primers used for sequencing.

Identified by Cas-OFFinder		GRCh38 reference genome				Primers for off-target detection
Target	**Gene Alias**	**HOMER Annotation**	**chromosome**	**Mismatches**	**Aligned sequence at site**	
On target	HBB		chr11	N/A	TCCCCAAAGGACTCAA CCTC	N/A
OT1	ZNF385D-AS2-U BE2E2-AS1	intergenic	chr3	3	ATTCCAAAGGACTCAA CCTC	F-**GGAGTGAGTACGGTGTGC**ACCTTATAGTCGC TGCTTGT R-**GAGTTGGATGCTGGATGG**TACTCTGCTATTT GAACATC
OT2	RP5-1119A7.17	exon	chr22	3	TCCCCAGAGAGCTCAA CCTC	F-**GGAGTGAGTACGGTGTGC**TGCTGACCTGGG AAAATAA R-**GAGTTGGATGCTGGATGG**AGGTGACTGAAG GTCAAGGT
OT3	BMPER	intron	chr7	4	TCCAGAAAGGACCTAA CCTC	F-**GGAGTGAGTACGGTGTGC**CTCCCCCTTATTT CCAAGTG R-**GAGTTGGATGCTGGATGG**CTGCTAACAGTG GCAGAGCT
OT4	MOCS1-RP11-55 2E20.1	intergenic	chr6	4	TCTAGAGAGGACTCAA CCTC	F-**GGAGTGAGTACGGTGTGC**GCTCTGTCCACTG CTAAGTT R-**GAGTTGGATGCTGGATGG**ATCTAGAGAGGC TCCAAGGA
OT5	SMYD4	exon	chr17	4	CCACCAAAAGAGTCAA CCTC	F-**GGAGTGAGTACGGTGTGC**AGGCCATAATTA AGTGTCTT R-**GAGTTGGATGCTGGATGG**TACCACAGGACA

						GAATGTCC
OT6	RNU6-649P-AC107057.1	intergenic	chr2	4	ACCCCAAAAGACACAATCTC	F-**GGAGTGAGTACGGTGTGC**CTTTCTAAAGATAAAAATGAG R-**GAGTTGGATGCTGGATGG**ATACAGACAATCAAATTCAAC
OT7	SNORA25-RP11-16D22.2	intergenic	chr13	4	TTTCCAAAGGACACAAACTC	F-**GGAGTGAGTACGGTGTGC**GAGAAGAACTTGGTACGGTC R-**GAGTTGGATGCTGGATGG**ACTCTGACAGGACTCAGTAT
OT8	TTTY2B	intron	chrY	2	TCCCCAAAAGTCTCAACCTC	F-**GGAGTGAGTACGGTGTGC**GCTCTTACAGCTGTCAGTCT R-**GAGTTGGATGCTGGATGG**AGAGTGGATTCCAGTGACAA
OT9	TTTY2	intron	chrY	2	TCCCCAAAAGTCTCAACCTC	F-**GGAGTGAGTACGGTGTGC**GCTCTTACAGCTGTCAGTCT R-**GAGTTGGATGCTGGATGG**AGAGTGGATTCCAGTGACAA
OT10	MAP2	exon	chr2	4	TCTCCAGTGAACTCAACCTC	F-**GGAGTGAGTACGGTGTGC**CGTGGTCTCAATTCTAGATC R-**GAGTTGGATGCTGGATGG**ACTCATCATGAACTCGGGTC
OT11	FAM117A-FAM117A/KAT7	intergenic	chr17	4	TTCCCAAACTAGTCAACCTC	F-**GGAGTGAGTACGGTGTGC**CTCAGAGGAACCCGTTCCTA R-**GAGTTGGATGCTGGATGG**AGTCAGTCTTGCCCACTATA

OT12	Gap-AL354822.1	intergenic	chrUn_GL000218v1	4	GCCCCCAAGGACCCAAACTC	F-**GGAGTGAGTACGGTGTGC**AAGGTCCCTCAGTGGCAGTT R-**GAGTTGGATGCTGGATGG**CTTCCTCCTGTTCTTGGTGT
OT13	GEMIN8-GLRA2	intergenic	chrX	4	TCCCCATAGAAATCAACATC	F-**GGAGTGAGTACGGTGTGC**TGTCAGTGGTTTGAAGTTCC R-**GAGTTGGATGCTGGATGG**TCACAATCCGGCTGAGAGGA
OT14	BARX2-RP11-237N19.3	intergenic	chr11	3	TCCTCATTGGACTCAACCTC	F-**GGAGTGAGTACGGTGTGC**CCAAGCTGAAATAAAGATGC R-**GAGTTGGATGCTGGATGG**TAGAAGACTAGAACCTGTTG
OT15	AC006227.1-AC096669.1	intergenic	chr2	4	CCCTCAAAGGTCACAACCTC	F-**GGAGTGAGTACGGTGTGC**GCTCAGAATAAATAGGATACA R-**GAGTTGGATGCTGGATGG**GTCTAGGACCCAGGTCCTTAC
OT16	TNIK	intron	chr3	3	TTCCCAAAGTACTCAGCCTC	F-**GGAGTGAGTACGGTGTGC**TCACTGTCTATATTCTGTCT R-**GAGTTGGATGCTGGATGG**CCAGAGGTTTTGTATGTGCC
OT17	RN7SKP256-AL591034.1	intergenic	chr6	4	TTCCCAAAGGACAAAACTTC	F-**GGAGTGAGTACGGTGTGC**ATGCAGATGGCTGTTGCATT R-**GAGTTGGATGCTGGATGG**CCTAGTCTTCCAGTCTCATT
OT18	LINC00700-RP11-69C17.2	intergenic	chr10	4	GGCACAAAGAACTCAACCTC	F-**GGAGTGAGTACGGTGTGC**TCAGTGCTGAGTCATTTGAC

						R-**GAGTTGGATGCTGGATGG**GTCACATGCATATTACGCC
OT19	ATP2B4	intron	chr1	3	TCCCCAAAAGACTTAACCCC	F-**GGAGTGAGTACGGTGTGC**ATGTCTCTCTTTATCCCTGC R-**GAGTTGGATGCTGGATGG**GAAACACATACCCCCAGACC
OT20	C18orf63	intron	chr18	4	ACCCCAAAGGACCAAAACTC	F-**GGAGTGAGTACGGTGTGC**TTAGGAAGGGTCCCTCGTAG R-**GAGTTGGATGCTGGATGG**CGTTTCTGTCCCAGGATTTC
OT21	HBD	exon	chr11	4	CAAAGGACTCAAAGAACCTC	F-**GGAGTGAGTACGGTGTGC**GTCCCTTGGGCTGTTTTCCT R-**GAGTTGGATGCTGGATGG**CACTCAGCTGAGAAAAAG

Supplemental Table 4. Buffer for protein purification

Nickel-NTA column buffer	Buffer A	20 mM TRIS + 500 mM NaCl + 20 mM imidazole, pH 8.0
	Buffer B	20 mM TRIS, 250 mM NaCl, 500 mM Imidazole, 10% glycerol, pH 8.0
SP, Heparin column buffer	Buffer C	20 mM HEPES +10% glycerol, pH 7.5
	Buffer D	20 mM HEPES+1 M NaCl+10% glycerol, pH 7.5
	Buffer E	20 mM HEPES + 250 mM NaCl + 1 mm EDTA +10% glycerol, pH 7.5
	Buffer F	20 mM HEPES +10% glycerol, pH 7.5
	Buffer G	20 mM HEPES +1 M NaCl+10% glycerol, pH 7.5
Q-HP column buffer	Buffer C	20 mM HEPES +10% glycerol, pH 7.5
	Buffer D	20 mM HEPES+1 M NaCl+10% glycerol, pH 7.5