

Supplementary Information

Multidimensional genome-wide analyses show accurate FVIII integration by ZFN in primary human cells

Jaichandran Sivalingam^{1,2,7,#}, Dimitar Kenanov^{3,#}, Hao Han^{3,#}, Ajit Johnson Nirmal¹, Wai Har Ng¹, Sze Sing Lee¹, Jeyakumar Masilamani⁵, Toan Thang Phan^{4,5}, Sebastian Maurer-Stroh^{3,6} & Oi Lian Kon^{1,2}.

¹ Division of Medical Sciences, Laboratory of Applied Human Genetics, Humphrey Oei Institute of Cancer Research, National Cancer Centre, Singapore 169610, Republic of Singapore.

² Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 117596, Republic of Singapore.

³ Bioinformatics Institute, Agency for Science, Technology and Research, Singapore 138671, Republic of Singapore.

⁴ Department of Surgery, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 119228, Republic of Singapore.

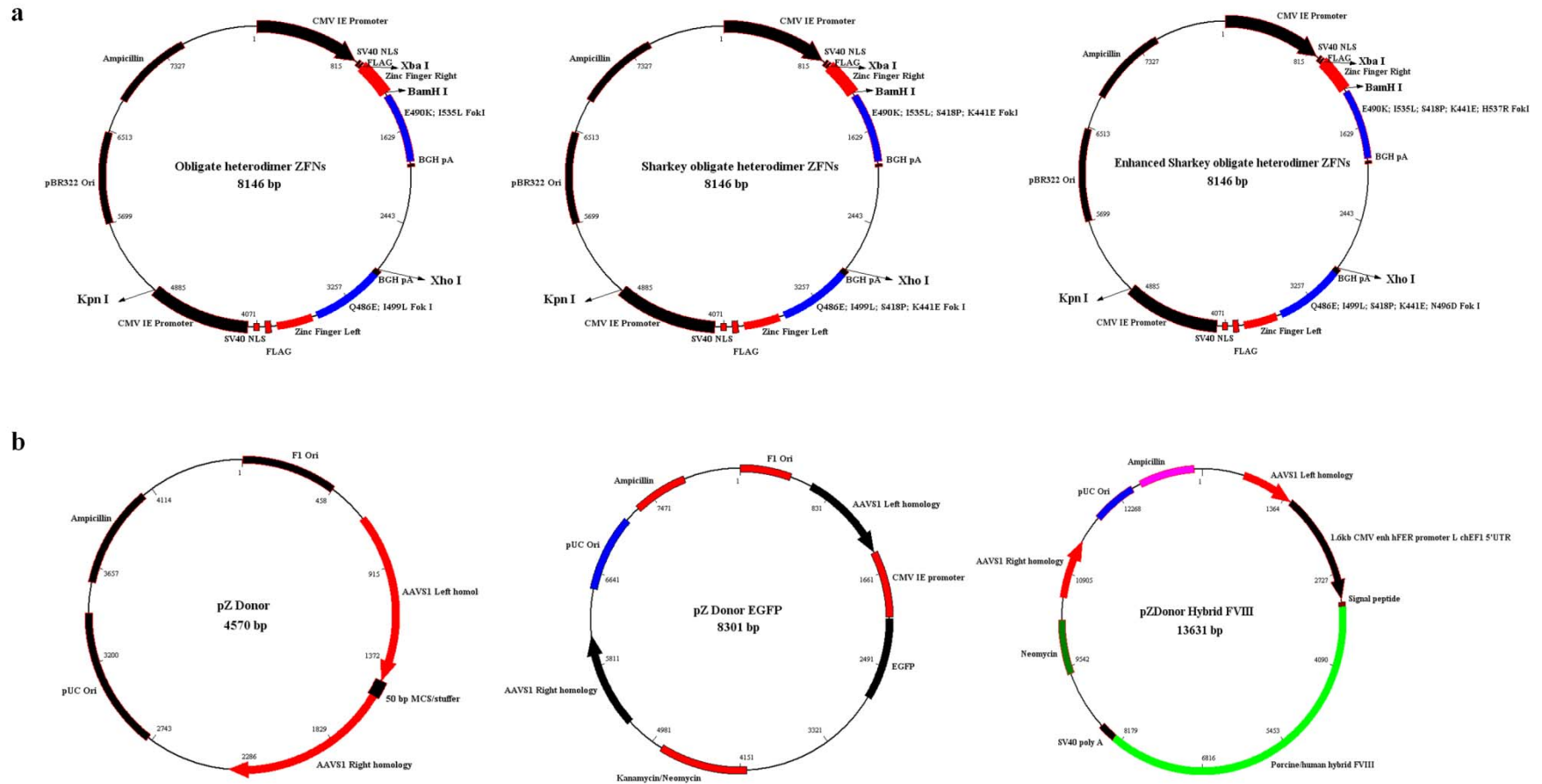
⁵ CellResearch Corporation, Singapore 117597, Republic of Singapore.

⁶ School of Biological Sciences, Nanyang Technological University, Singapore 637551[#]

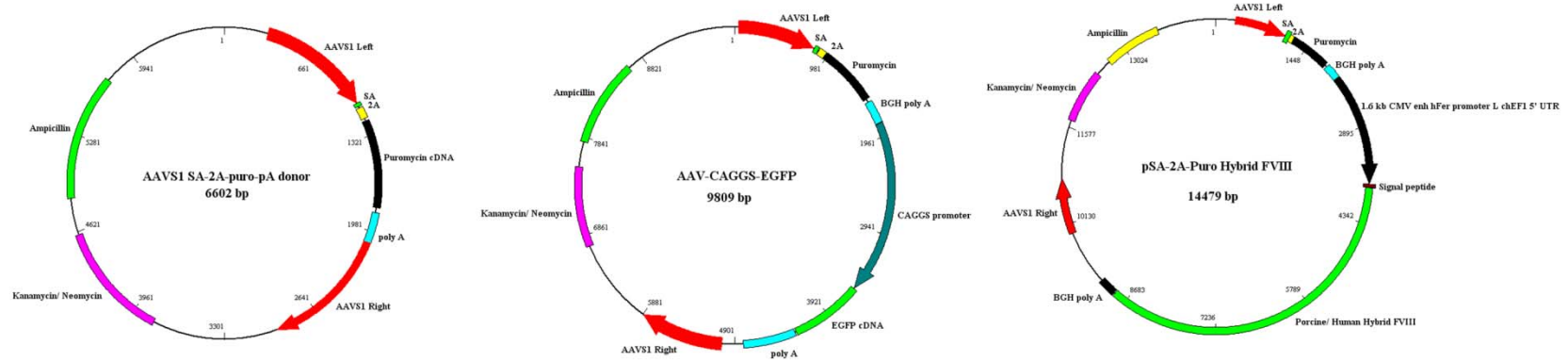
These authors contributed equally to this work.

⁷Current address: Bioprocessing Technology Institute, Agency for Science, Technology and Research, Singapore 138668, Republic of Singapore

Supplementary Figure 1. Plasmid DNA constructs used in this study.

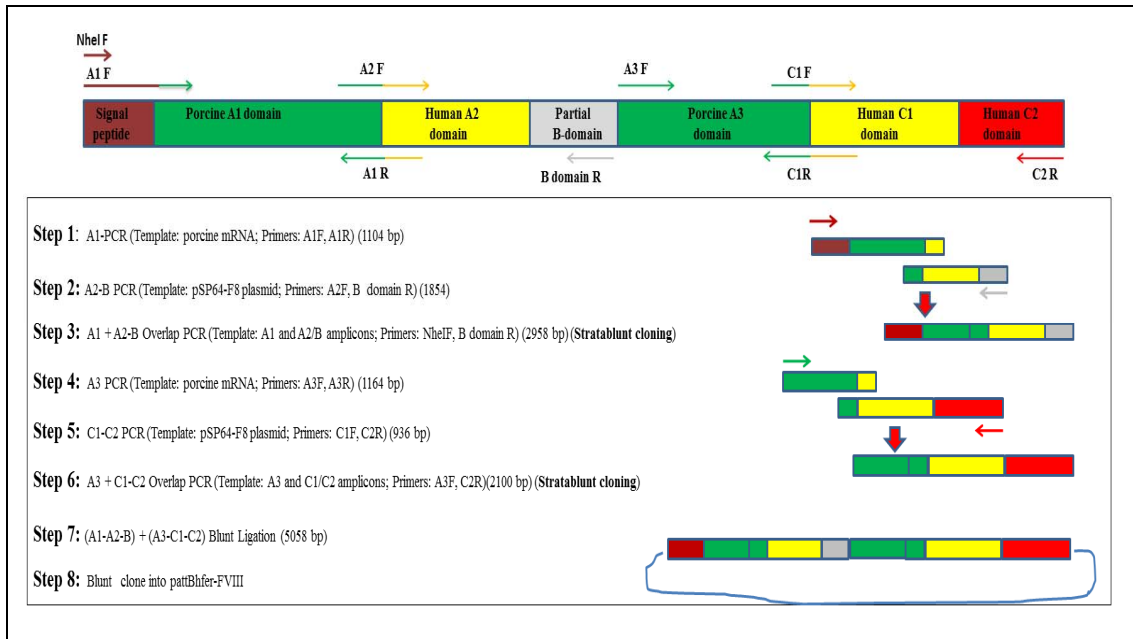


c



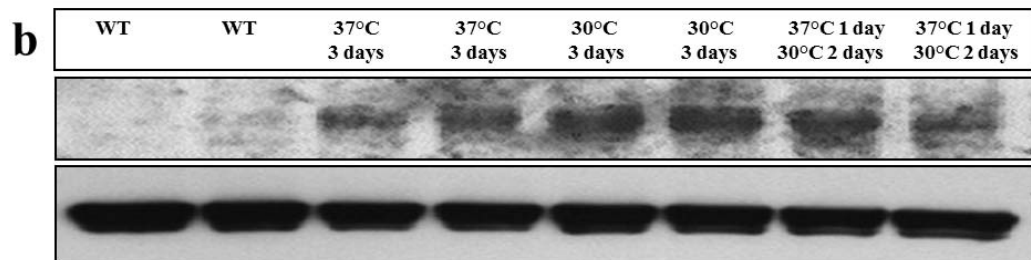
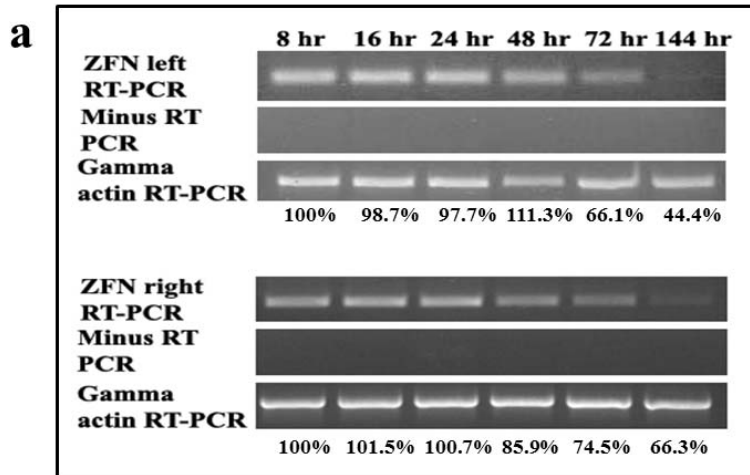
(a) AAVS1 ZFN variant constructs. Three variants were (left) obligate heterodimer (OH) ZFN (modified according to Miller, J.C., *et al.* (2007) *Nat Biotechnol* **25**: 778-785); (center) Sharkey obligate heterodimer ZFN (obligate heterodimer further modified according to Guo, J, *et al.* (2010). *J Mol Biol* **400**: 96-107); (right) Enhanced Sharkey obligate heterodimer ZFN (Sharkey obligate heterodimer ZFN variant further modified according to Doyon, Y, *et al.* (2011) *Nat Methods* **8**: 74-79). (b) Donor constructs with AAVS1 homology arms. Three donors were (left) pZDonor (50-bp insert and multiple cloning site); (center) pZDonor EGFP (encoding 3-kb EGFP); (right) pZDonor Hybrid FVIII (encoding 9-kb human-porcine FVIII cDNA). (c) Donor constructs for gene trap strategy. Donor constructs with AAVS1 homology arms and a splice acceptor sequence to express a promoterless puromycin resistance gene from the endogenous *PPP1R12C* promoter following integration at the AAVS1 locus were (left) AAVS1 SA-2A-puro-pA donor (Addgene plasmid #22075) which integrates a 1-kb puromycin resistance gene at the AAVS1 locus; (center) AAV-CAGGS-EGFP (Addgene plasmid #22212) which integrates a 4.2-kb fragment comprised of puromycin resistance gene and

an EGFP reporter gene expressed from CAGGS promoter; (right) pSA-2A-Puro Hybrid FVIII which integrates a 9-kb fragment comprised of puromycin resistance gene and human-porcine FVIII cDNA expressed from the human ferritin light chain promoter.



Supplementary Figure S2. Hybrid human-porcine B domain-truncated FVIII cDNA and assembly

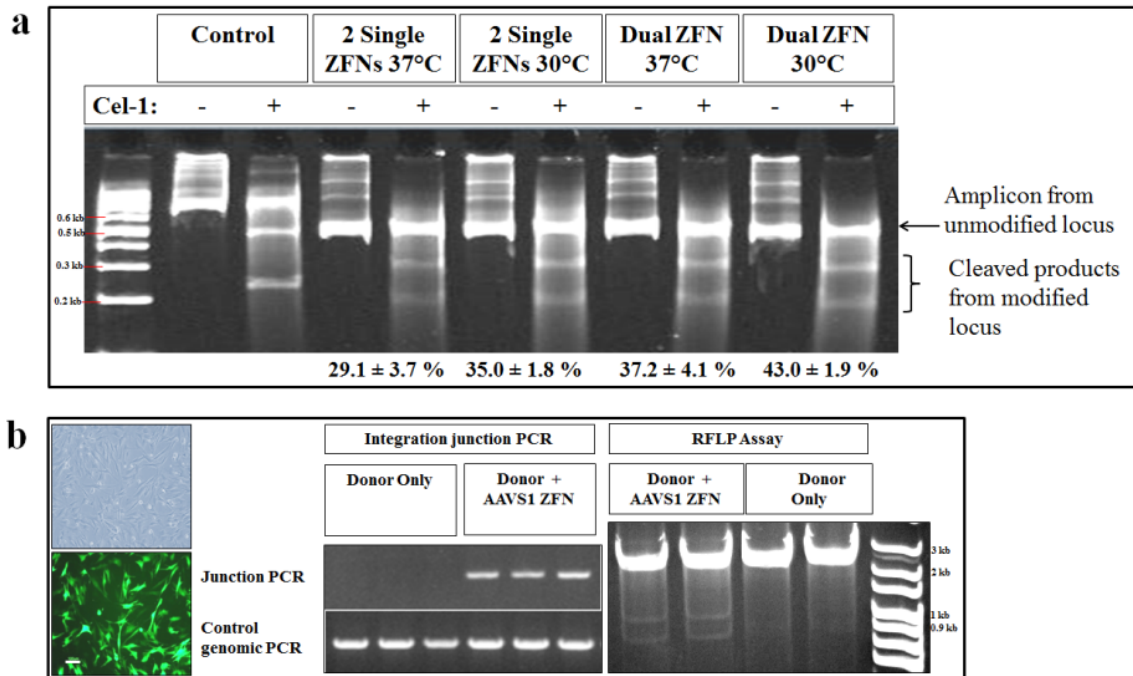
steps. Top: Schematic of B domain-truncated human-porcine FVIII cDNA consisting of porcine A1 and A3 domains, human signal peptide, A2, residual B (comprising the first 247 amino acids and six glycosylation sites), C1 and C2 domains. **Bottom:** Overlap PCR steps used in domain assembly. A1 and A3 domains were obtained by RT-PCR of total pig liver RNA based on the reference porcine cDNA sequence (NM_214167.1). Human domains were amplified from complete human FVIII cDNA in pSP64-VIII (American Type Culture Collection



Supplementary Figure S3. Time-course of ZFN transcription and effect of mild hypothermia on ZFN protein levels in transiently electroporated CLECs.

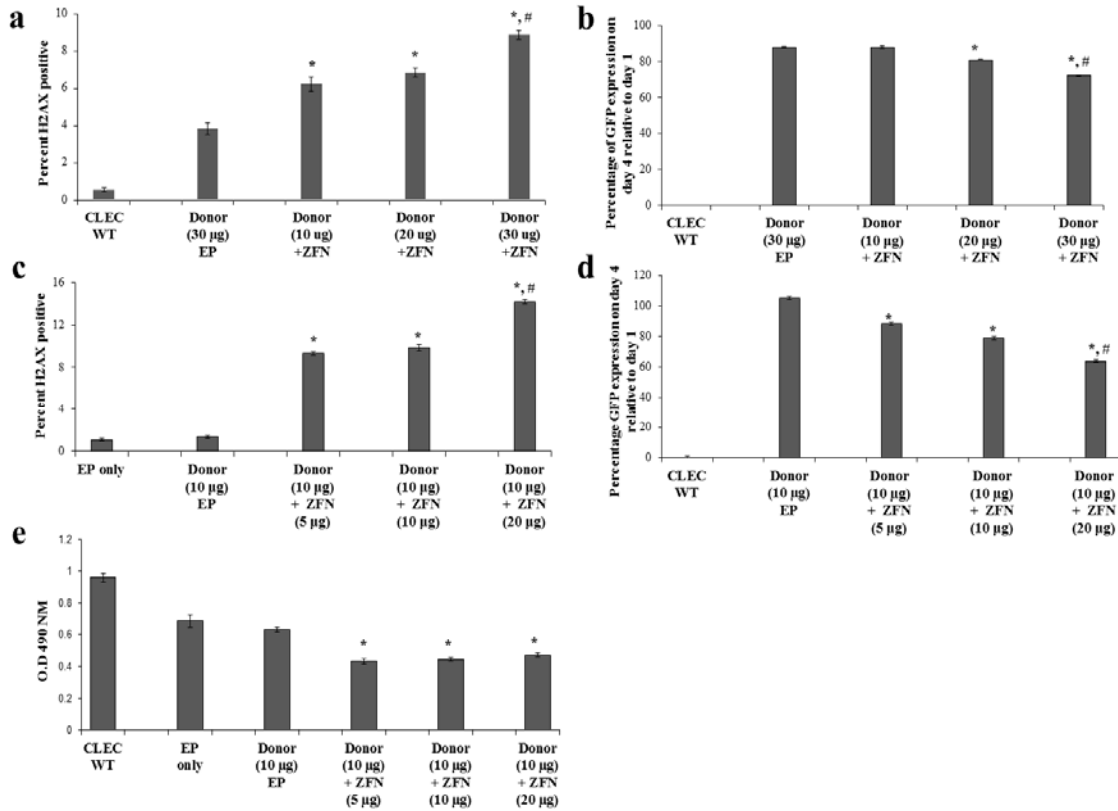
(a) Time course of transcription of ZFN constructs. RT-PCR was performed on total RNA from transfected CLECs at the time points indicated after electroporation with plasmid DNA encoding either the left ZFN (top) or right ZFN (bottom) using homology arm-specific primers. Negative controls were reactions performed without reverse transcription (Minus RT PCR). RT-PCR of γ -actin mRNA was the positive control. Densitometric measurements of ZFN transcript bands were normalized to their respective actin levels and expressed as a percentage of ZFN mRNA transcript levels at 8 hours (indicated against each lane in both gels). (b) Time course of ZFN protein expression. Upper panel: Protein immunoblot of FLAG-tagged ZFN proteins in CLECs transfected with a single plasmid encoding both left and right ZFNs and incubated at either 37°C or 30°C for the indicated number of days showed higher

abundance of ZFN proteins when cells were exposed to mild hypothermia. Untransfected CLECs (WT) were negative for expression of FLAG-tagged ZFN protein. Lower panel: β -Actin served as loading control.



Supplementary Figure S4. Site-specific double-strand DNA cleavage and homology-directed repair in primary human CLECs. (a) Comparison of AAVS1 ZFN monomer delivery constructs. CLECs electroporated with two separate plasmids encoding left or right AAVS1 ZFNs (2 single ZFNs) or a single plasmid encoding both left and right AAVS1 ZFNs (Dual ZFN) were incubated at either 37°C for 3 days (37°C) or 37°C for 1 day followed by 30°C for 2 days (30°C). The genomic region spanning the AAVS1 ZFN target site was amplified and digested with CEL-1 nuclease (+) or left undigested (-). PCR amplicons were resolved by 10% polyacrylamide gel electrophoresis, imaged and quantified using BioRad® Gel Doc 2000 transilluminator and QuantityOne software. PCR amplicon from the Surveyor™ mutation detection kit (Transgenomic) was the positive control for CEL-1 nuclease digest. Estimates of the proportion of modified genomic DNA (ZFN cleaved and repaired with indels) in the bulk treated population based on densitometry are reported below the respective lanes in gel images for each treatment condition. **(b)** AAVS1 ZFN-dependent donor DNA integration. **Left:** CLECs electroporated with pZDonor alone or with dual AAVS1 ZFN plasmid, both in the presence of pEGFP, were evaluated for gene transfer efficiency

by fluorescence microscopy (original magnification x100). Scale-bar = 100 μ m. **Center:** These cells were analysed for site-specific integration of a pZDonor by integration junction PCR. Integration junction PCR was performed with a vector-specific and a genome-specific primer to amplify a 1-kb region spanning the integration junction. Control genomic PCR amplified a 900-bp region of the AAVS1 locus. **Right:** RFLP assay was performed by digesting PCR amplicons that spanned the integration site with *Hind* III followed by 5% polyacrylamide gel electrophoresis.

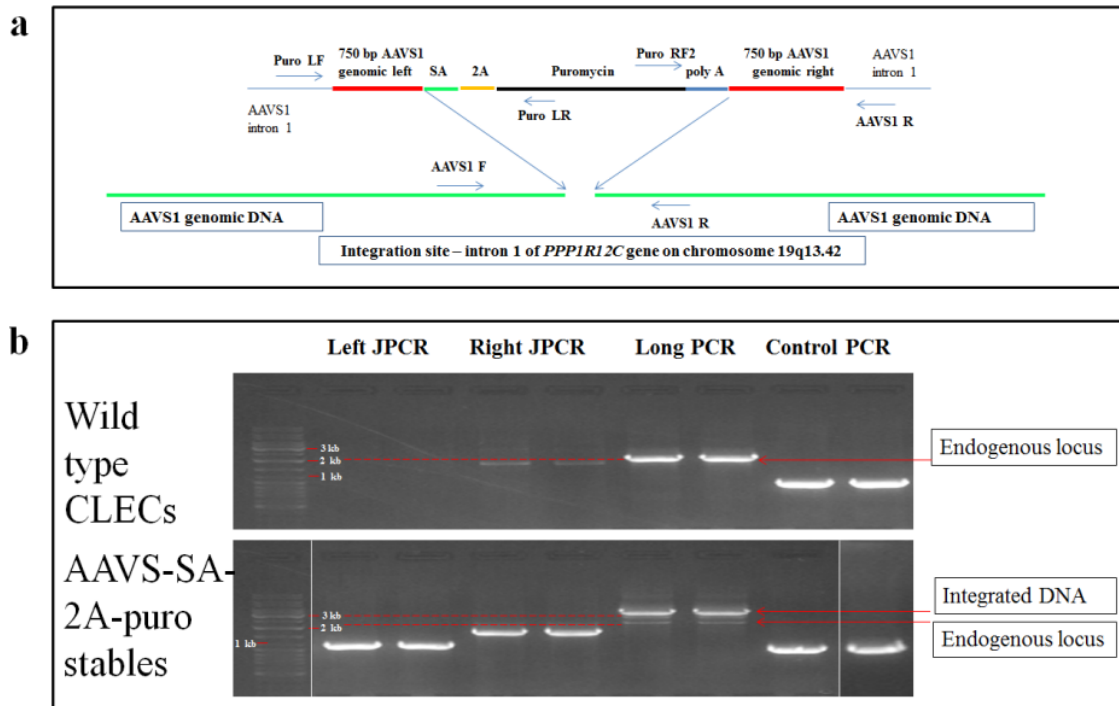


Supplementary Figure S5. Cellular toxicity and genotoxicity induced by donor DNA or AAVS1

ZFN. (a) Dose titration of donor DNA for genotoxicity. Effects of increased donor DNA dosage on inducing DNA double-strand breaks. Two million CLECs were electroporated with 2 µg pEGFP (reporter gene), a fixed dose of AAVS1 ZFN (5 µg) and increasing doses of pZDonor as indicated. CLEC WT denotes unmodified cells which were not electroporated. Genotoxicity was evaluated by the percentage of phosphorylated H2AX-positive cells on day 4 post-electroporation.

*Indicates $P < 0.01$ compared to Donor (30 µg) EP and # indicates $P < 0.01$ compared to Donor (10 µg) + ZFN and Donor (20 µg) + ZFN. Data are mean \pm SEM; $n = 3$. (b) Dose titration of donor DNA for cytotoxicity. Cellular toxicity was evaluated by comparing the decline in percentage of GFP-positive cells on day 4 post-electroporation relative to day 1 after electroporation under the same experimental conditions as (a) above. Data are presented as percentage GFP-expressing cells on day 4 relative to day 1. * Indicates $P < 0.01$ compared to Donor (10 µg) + ZFN and #

indicates $P < 0.01$ compared to Donor (20 μg) + ZFN. Data are mean \pm SEM; $n = 3$. (c) Dose titration of AAVS1 ZFN for genotoxicity. Effects of increased ZFN DNA dosage on inducing DNA double-strand breaks. Two million CLECs were electroporated with 2 μg pEGFP (reporter gene), a fixed dose of pZDonor (10 μg) and increasing doses of AAVS1 ZFN as indicated. Genotoxicity was evaluated by the percentage of phosphorylated H2AX-positive cells on day 4 post-electroporation. EP only denotes CLECs which were electroporated without any added plasmid construct. * Indicates $P < 0.05$ compared to Donor (10 μg) EP and # indicates $P < 0.05$ compared to Donor (10 μg) + ZFN (5 μg) and Donor (10 μg) + ZFN (10 μg). Data are mean \pm SEM; $n = 3$. (d) Dose titration of AAVS1 ZFN for cytotoxicity. Cellular toxicity was evaluated by comparing the decline in percentage of GFP-positive cells on day 4 post-electroporation relative to day 1 after electroporation under the same experimental conditions as (c) above. Data are presented as percentage GFP-expressing cells on day 4 relative to day 1. *Indicates $P < 0.05$ compared to Donor (10 μg) EP and # indicates $P < 0.01$ compared to Donor (10 μg) + ZFN (10 μg). Data are mean \pm SEM; $n = 3$. (e) Dose titration of AAVS1 ZFN for cell proliferation. MTS assay of viable CLECS that were untreated (CLEC WT), that received electroporation only (EP only), received 10 μg pZDonor donor DNA only or pZDonor DNA and increasing doses of ZFNs as indicated, 1 day post-electroporation. *Indicates $P < 0.01$ compared to Donor (10 μg) EP. Data are mean \pm SEM; $n = 4$.

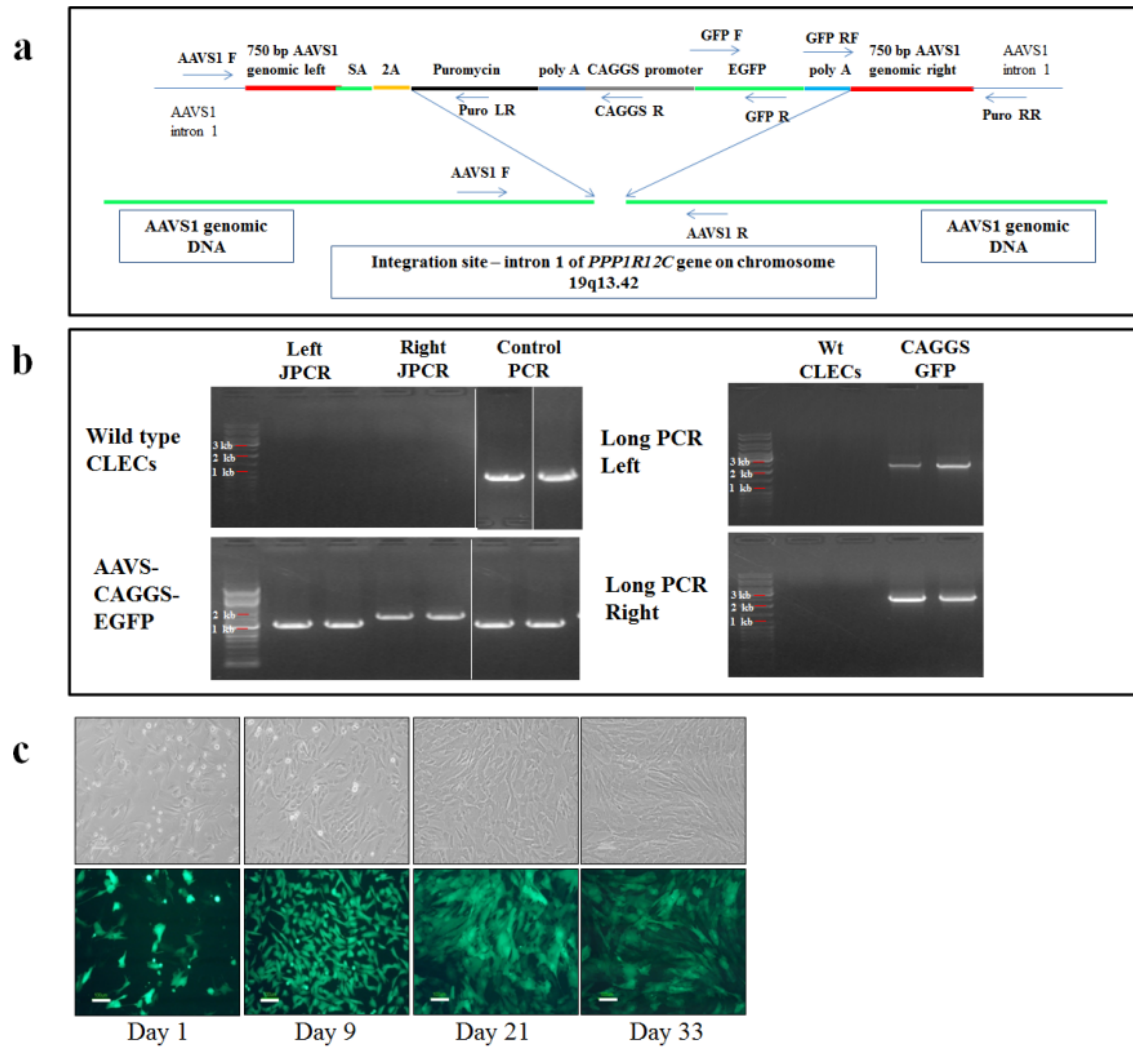


Supplementary Figure S6. AAVS1 site-specific integration of 1.3-kb donor DNA in CLECs. (a)

Schematic (not drawn to scale) showing homologous recombination-mediated integration of pAAVS-SA-2A-puro-pA donor (Addgene plasmid #22075) into the AAVS1 locus. The primer pairs for integration junction PCR were Puro LF; Puro LR (left junction); and Puro RF2; AAVS1 R (right junction). Primers for long PCR anchored beyond the integration site were Puro LF and AAVS1 R.

(b) Accurate integration of donor DNA. Left and right integration junction PCR (Left JPCR and Right JPCR) and Long PCR (spanning the integrated transgene) performed on 200 ng genomic DNA (equivalent to 30440 cells) stably integrated puromycin-resistant CLECs co-electroporated with pAAVS-SA-2A-puro-pA donor (1-kb puromycin resistance gene) and AAVS1 ZFN. Amplicons of the predicted sizes (Left JPCR amplicon, 1.1 kb; Right JPCR amplicon, 1.6 kb) were evidence of donor DNA integration at the AAVS1 locus. Control PCR amplified a 900-bp region of the AAVS1 locus 2 kb away from the integration site. Long PCR performed with genome-specific primers

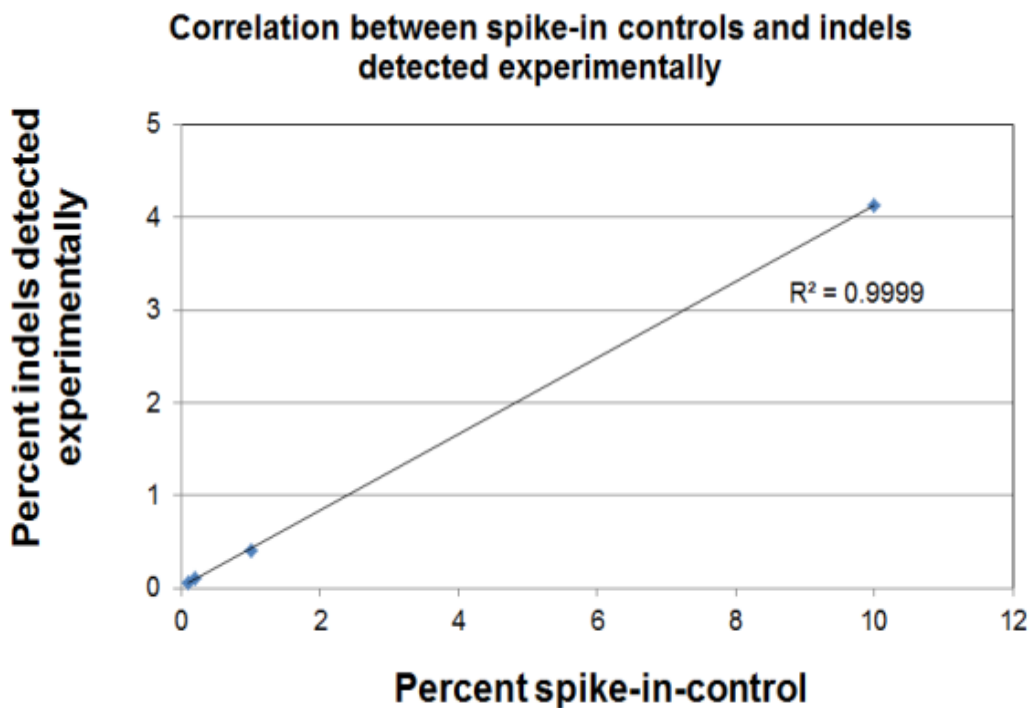
anchored beyond the integration site yielded a 2-kb amplicon in the absence of donor DNA integration and a dominant 3-kb amplicon in donor-integrated cells. Control PCR reactions on unmodified wild-type CLECs showed only the 2-kb amplicon. The identities of integration junction PCR and long PCR products were confirmed by sequencing. White vertical lines demarcate regions of gel images that were merged.



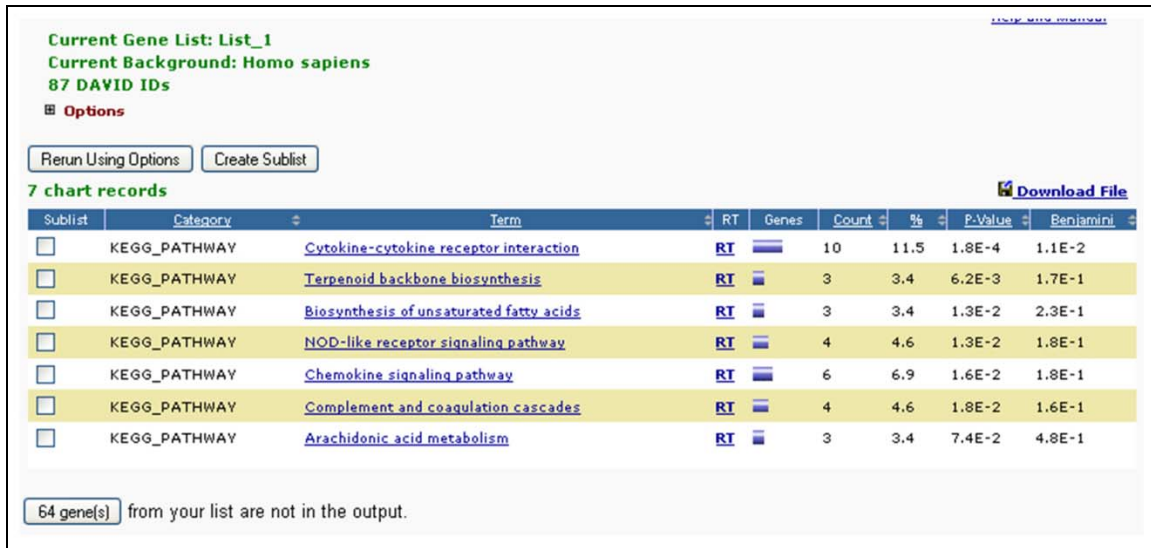
Supplementary Figure S7. AAVS1 site-specific integration of 4.2-kb donor DNA in CLECs. (a) Schematic (not drawn to scale) showing homologous recombination-mediated integration of pAAV-CAGGS-EGFP (Addgene plasmid #22212) into the AAVS1 locus. For integration junction PCR, one primer was specific to the vector (donor DNA) and the other primer was specific to genomic DNA beyond the integration site (left junction: AAVS1F, Puro LR; right junction: GFP RF, Puro RR). Long PCR amplified two overlapping segments to cover the entire length of donor DNA using outer primers specific to genomic DNA beyond the integration site (Long PCR Left: AAVS1

F, CAGGS R; Long PCR Right: GFP F, Puro RR). **(b)** Accurate integration of donor DNA. Integration junction PCR (Left JPCR and Right JPCR) and overlapping long PCRs (spanning the integrated transgene) were performed on 200 ng genomic DNA (equivalent to 30440 cells) puromycin-resistant CLECs co-electroporated with pAAV-CAGGS-EGFP (4.2 kb donor DNA consisting of promoterless puromycin resistance gene cDNA and CAGGS promoter-EGFP cDNA) and a plasmid construct encoding both left and right AAVS1 ZFNs showed donor DNA integration at the AAVS1 locus. Products of left and right integration junction PCR (1 kb and 1.3 kb, respectively) were of the correct predicted size. Control PCR amplified a 900-bp region of the AAVS1 locus 2 kb away from the integration site. Overlapping long PCR encompassing the integrated transgene amplified products of the predicted sizes (Long PCR Left; 2.3 kb, Long PCR Right; 2.5 kb). No amplicons were generated in parallel PCRs performed on unmodified wild-type CLECs. **(c)** Durable transgene expression integrated in AAVS1 locus. Brightfield and fluorescence images of CLECs on day 1 (before puromycin selection) and on days 9, 21 and 33 (after puromycin selection) post-electroporation with pAAV-CAGGS-EGFP show that the AAVS1 locus supports durable EGFP expression of donor DNA. Scale-bar = 100 μ m. White vertical lines demarcate regions of gel images that were merged.

Ratio of mutant: wt amplicon	Percent spiked-in mutant amplicon	Total reads mapped	Number of indel reads	Percent indels (%)
1:10	10	60562	2504	4.13
1:100	1	55315	222	0.40
1:500	0.2	36029	40	0.11
1:5000	0.1	79659	45	0.056

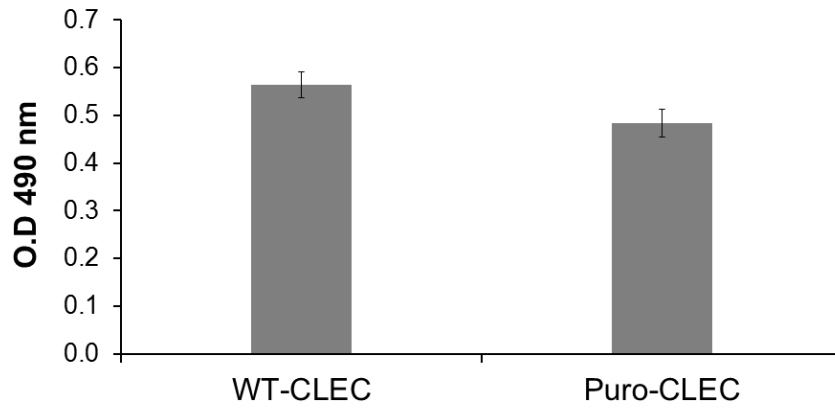


Supplementary Figure S8. Sensitivity of indel detection by targeted deep sequencing. Table and graph show the ratios and percentages of mutant amplicons to wild type amplicons, the number of experimentally retrieved mapped reads, the number and percentage of indels detected in these mapped reads for each spike-in concentration. A highly linear correlation ($R^2 = 0.999$) was observed between the known percentage of indels in the spike-in controls and the percentage of indels experimentally determined by deep sequencing.

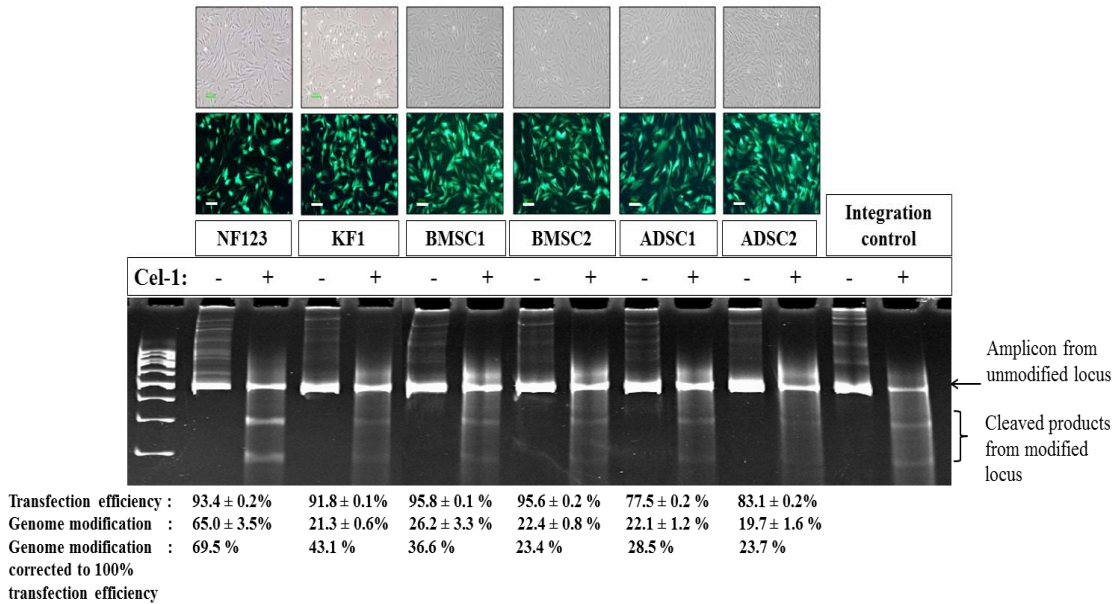


Supplementary Figure S9. Pathway analysis of 90 dysregulated transcripts in puro-CLECs.

DAVID analysis of genes which were over-expressed ($n = 57$) or under-expressed ($n = 33$) by ≥ 2 -fold in puro-CLECs compared to wild-type CLECs were mapped only to chemokine-chemokine receptor interaction.



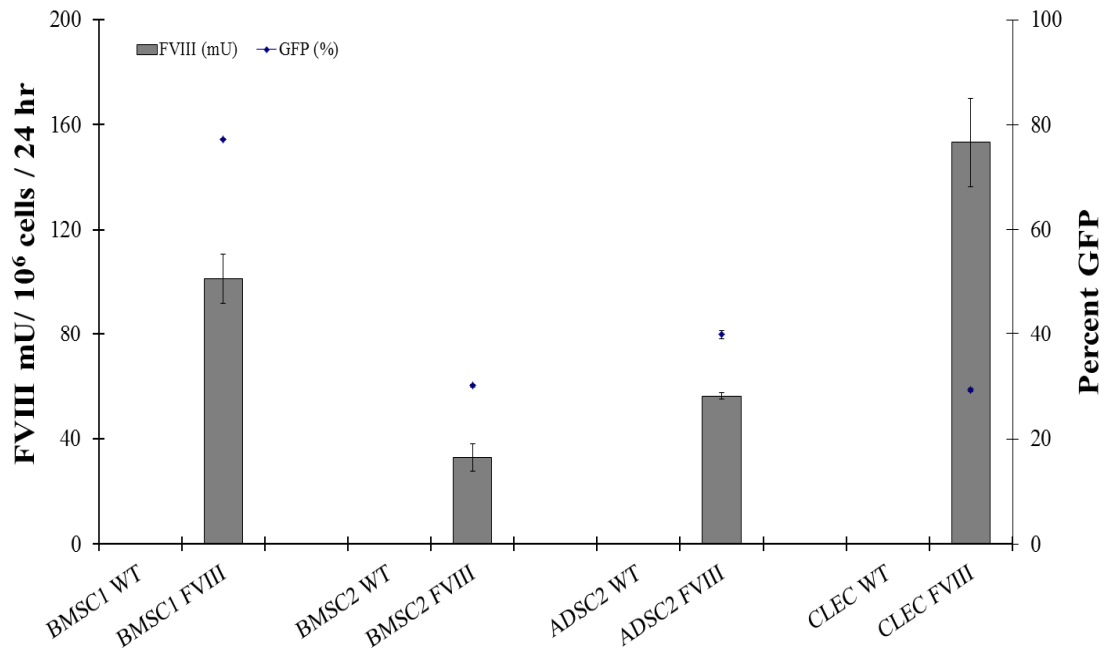
Supplementary Figure S10. Comparison of wt-CLEC and puro-CLEC cell proliferation. MTS assay was performed 7 days after wt-CLECs and puro-CLECs were plated at an initial density of 100 cells per well (96-well plate). Data are mean absorbance readings \pm SD; $n = 6$ per group. $P = 0.303$.



Supplementary Figure S11. Site-specific genome modification in different adult primary human cells.

(Top) Transfection of adult primary human cells. Brightfield and fluorescence images of primary human cells 1 day post-electroporation with pmaxGFP (Lonza). NF123 and KF1 are dermal fibroblasts; BMSC1 and BMSC2 are bone marrow-derived stromal cells; ADSC1 and ADSC2 are adipose tissue-derived stromal cells. Scale-bar = 100 μ m. **(Bottom)** Site-specific AAVS1 ZFN activity in adult primary human cells. A region spanning the AAVS1 integration site was amplified from genomic DNA extracted from BMSC1, BMSC2, ADSC1, ADSC2, NF123 and KF1 electroporated with pZDonor. CEL-1 nuclease digested (+) or undigested (-) PCR amplicons were resolved by 10% polyacrylamide gel electrophoresis. Positive integration control was a PCR amplicon from genomic DNA provided in the CompoZr[®] targeted integration kit (Sigma-Aldrich). Transfection efficiency determined by flow cytometry analysis of GFP-positive cells and genome modification efficiency determined by densitometry of cleaved and uncleaved amplicons are shown. **Per cent genome modification normalized to 100% transfection efficiency is also shown.**

Data are mean \pm SD; $n = 3$ per group.



Supplementary Figure S12. FVIII transgene secretion by different primary human cell types.

Primary human bone marrow-derived stromal cells (BMSC1 and BMSC2), human adipose tissue-derived stromal cells (ADSC2) and human cord lining epithelial cells (CLEC) secrete high levels of hybrid FVIII following co-electroporation with pmaxGFP (Lonza) and plasmid DNA encoding hybrid FVIII cDNA expressed from a CMV promoter. WT denotes untransfected control cells. Graph shows the percentage of GFP-positive cells (diamonds; right axis) determined by flow cytometry and FVIII activity (bars; left axis) in overnight conditioned media determined using the Coamatic® Factor VIII kit assay (Chromogenix). Data are mean \pm SEM; $n = 3$.

Supplementary Table S1. Sequence of the complete 9-kb FVIII donor DNA integrated in AAVS1 locus of CLECs.

AAVS1 site-specific amplicons of puro-CLEC genomic DNA were sequenced to ascertain

integration of the complete FVIII donor DNA. Key to colour coding:

Black lower case, endogenous sequence of chromosome 19 at the AAVS1 locus. Green, AAVS1

homology arms. Blue, prokaryotic sequences in donor DNA. Red, hybrid B domain-truncated

human-porcine FVIII cDNA. Integrated sequences are in bold upper case. Underscored

sequences around the ZFN cleavage site were used for analysis of integrated donor DNA.

tctggccactggctgtttaagatacgcctggtgcccagtggcaggccggccaggccttcagtgctcagtggaaaccacgaaa
ggactcctggctatctgcaaacaggaagtgaacggggaagggaggggcttctcatctgggtgcgggaacccacatggtac
ctgttagacacggcaaaacccccgtcaccaccacaggtggcgcttcagtgctcagactaggaagaggtccagccctcc
tcctcagagccaggagtctgccccagccccctcctgccttaaaccagccaggctcctcaagggtcaagctcggaaacca
ccccagcagatactctgcaggaacgaagccgtgggcccagggtatgcaggggtggaggaaggccaccctgtgctgggacag
actcaggggctggcgggactcccagaggggtgagacagctgcacacctgtgtcctgggccccaggctgtcacactcca
gttactgaggccccctctgcacggggcctgcagccaggggctgacacgggcccaccgtttctcattctcccttaggggtcca
aaacttggggggacaaaagccgaagtccagggggtcgaggaggactgcccaggccttgggacactgggtgggctcc
gggacctgaactggagctgaggaaggagtgaagctaaactcctagatccacgggataaattacccccaaagtcctcacctctc
caaagctgcccctctggagagggcgggagggagctacagggccaagagcatgaggtcatggaactcgggctgtgaagg
ggccgcacgtgcctgggaacgggatgaactggctcgtttattccaccagttgtcatggcgatagggaggggggcaagg
agagcaatgggctttcccttcaaggacctgccagtagcagcatcctgtgaaagatgectgaggcctgggaccaggggact
ccagagtcaggcccaacccctccccattcaaccaggaggccaggccccagccctccgacctagatgaaggagtccagg
ccccagcctctccccattcagaccaggggtccaggccccagccccgctccctaaagaccagaagtcaggccccagccc
ctctccctcagaccacgagtcaggccccagccccctcctcctcctggaccagggagtccaggccccagtcctccaccctc
agaccagggagtccaggccccagccccctcctcctcctcggaccagggagtccaggccccagccccctcctctcctaaaccagga
gcccaggccccagctctctctgttcagccctaagaatcctggctccagccccctcactctagccccaacccccagccact
aaggcaattggggtgcaggaatggggcaggggtaccagcctaccaagtgggtgataaaccacgtggggtaccctaagaac
ttgggaacagccacagcaggggggctgatgcttggggacctgctgggagaaggatgcaggacgagaaacacagccccaggt
ggagaaactggccgggaatcaagagtcaccagagacagtgaccaaccatccctgtttcctaggactgagggtttcagtgcta
aaactaggctgtcctgggcaaacagcataagctgggtacccacaccagacctgaccaaaccagctcccctgcttcttggc
cacgtaacctgagaagggaaatccctcctctctgaacccagccccccaatgctccaggcctcctgggataccccgaagagt
gagtttgcaagcagtcacccacagttggagggagaatccaccaaaggcagcctggtagacagggtggggtggcctctc
gtgggtccaggccaagtaggtggcctggggcctctgggggatgcaggggaagggggatgcaggggaacggggatgcag
gggaacggggctcagctgaa**GAGCAGAGCCAGGAACCCCTGTAGGGAAGGGGCAGGA**
GAGCCAGGGGCATGAGATGGTGGACGAGGAAGGGGGACAGGGAAGCCTGAG
CGCTCTCCTGGGCTTGCCAAGGACTCAAACCCAGAAGCCCAGAGCAGGGCC
TTAGGGAAGCGGGACCCTGCTCTGGGCGGAGGAATATGTCCCAGATAGCACT
GGGACTCTTTAAGGAAAGAAGGATGGAGAAAGAGAAAGGGAGTAGAGGCG
GCCACGACCTGGTGAACACCTAGGACGCACCATTCTCACAAGGGAGTTTTC
CACACGGACACCCCTCCTCACCACAGCCCTGCCAGGACGGGGCTGGCTAC

TGGCCTTATCTCACAGGTAAAAGTACGACGACGGAGGAACAATATAAATTGGG
GACTAGAAAGGTGAAGAGCCAAAGTTAGAACTCAGGACCAACTTATTCTGAT
TTTGTTTTTCCAAACTGCTTCTCCTCTTGGGAAGTGTAAGGAAGCTGCAGCAC
CAGGATCAGTGAAACGCACCAGACGGCCGCGTCAGAGCAGCTCAGGTTCTGG
GAGAGGGTAGCGCAGGGTGGCCACTGAGAACCAGGGCAGGTCACGCATCCCC
CCCTTCCCTCCCACCCCTGCCAAGCTCTCCCTCCCAGGATCCTCTCTGGCTCC
ATCGTAAGCAAACCTTAGAGGTTCTGGCAAGGAGAGAGATGGCTCCAGGAA
ATGGGGGTGTGTACCAGATAAGGAATCTGCCTAACAGGAGGTGGGGGTTAG
ACCAATATCAGGAGACTAGGAAGGAGGAGGCCTAAGGATGGGGCTTTTCTG
TCACCAATCCTGTCCCTAGTGACCCACAGTGGGGCAAGCTTCTGACCTCT
TCTCTTCCCTCCACAGGGCCTCGAGAGATCTGGCAGCGGAGAGGGCAGA
GGAAGTCTTCTAACATGCGGTGACGTGGAGGAGAATCCCGGCCCTAGGC
TCGAGATGACCGAGTACAAGCCCACGGTGCGCCTCGCCACCCGCGACGA
CGTCCCCAGGGCCGTACGCACCCCTCGCCGCGCGTTTCGCCGACTACCC
GCCACGCGCCACACCGTCGATCCGGACCGCCACATCGAGCGGGTACC
AGCTGCAAGAACTCTTCCCTCACGCGCGTCCGGCTCGACATCGGCAAGGT
GTGGGTGCGGGACGACGGCGCCGCGGTGGCGGTCTGGACCACGCCGA
GAGCGTCGAAGCGGGGGCGGTGTTCCCGAGATCGGCCCGCGCATGGC
CGAGTTGAGCGGTTCCCGGCTGGCCGCGCAGCAACAGATGGAAGGCCTC
CTGGCGCCGACCCGGCCCAAGGAGCCCGCGTGGTTCCTGGCCACCGTCG
GCGTCTCGCCCGACCCAGGGCAAGGGTCTGGGCAGCGCCGTCGTGCT
CCCCGGAGTGGAGGCGGCCGAGCGCGCCGGGGTGCCCGCCTTCTGGA
GACCTCCGCGCCCCGCAACCTCCCCTTCTACGAGCGGCTCGGCTTACC
GTCACCGCCGACGTCGAGGTGCCCGAAGGACCGCGCACCTGGTGCATGA
CCCGCAAGCCCGGTGCCTGATCTAGAGGGCCCGTTAAACCCGCTGATC
AGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGGCCCTCCC
CCGTGCCTTCCCTGACCCCTGGAAGGTGCCACTCCCCTGCTTCTTCTAA
TAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCT
GGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAA
TAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGGAATTGTAACCTAC
GGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCGCCATTGACG
TCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCATTG
ACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACAT
CAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAA
ATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTA
CTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGATGATGCGG
TTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGAT
TTCCAAGTCTCCACCCCAATTGACGTCAATGGGAGTTTGTGTTTACTAGTC
AGGGCCCCAACCCCCCAAGCCCCATTTCACAACACGCTGGCGCTACA
GGCGCGTGAATCCCCTTGCTTTGGGGCGGGGGGCTGAGACTCCTATGT
GCTCCGGATTGGTCAGGCACGGCCTTCGGCCCCGCTCCTGCCACCGCA
GATTGGCCGCTAGGCCTCCCCGAGCGCCCTGCCTCCGAGGGCCGGCGCA
CCATAAAAGAAGCCGCCCTAGCCACGTCCCCTCGCAGTTCGGCGGTCCC
GCGGGTCTGTCTCAAGCTTGCCGCCAGAACACAGGTAAGTGCCGTGTGT
GGTCCCAGGGGCTGGCCTCTTACGGGTTATGGCCCTTGGCTGCCTT
GAATTACTTCCATGCCCTGGCTGCAGTACGTGATTCTTGATCCCGAGCT
TCGGGTTGGAAGTGGGTGGGAGAGTTCGAGGCCTTGCCTTAAGGAGCC

CCTTCGCCTCGTGCTTGAGTTGAGGCCTGGCTTGGGGCGCTGGGGCCGCC
GCGTGCTAATCTGGTGGCACCTTCGCGCCTGTCTCGCTGCTTTTCGCTAA
GTCTCTAGCCATTTAAAATTTTTGATAACCAGCTGCGACGCTTTTTTTCT
GGCGAGATAGTCTTGTAATGCGGGCCAGGATCTGCACACTGGTATTTC
GGTTTTTGGGGCCGCGGGCGGGCAGGGGCCCGTGCGTCCCAGCGCAC
ATGTTCCGGCGAGGCGGGGCTGCGAGCGCGGCCACCGAGAATCGGACG
GGGGTAGTCTCAAACCTGGCCGGCCTGCTCTGGTGCCTGGCCTCGCGCCG
CCGTGTATCGCCCCGCCCTGGGCGGCAAGGCTGGCCCCGGTCGGCACCA
GTTGCGTGAGCGGAAAGATGGCCGCTTCCC GGCCCTGCTGCAGGGAGCT
CAAAATGGAGGACGCGGGC GCCCGGGAGAGCGGGCGGGTGAGTCACCCA
CACAAAGGAAAAGGGCCTTTCCTTCTCATCCGTGCTTCATGTGACTCC
ACGGAGTACCGGGCGCCGTCCAGGCACCTCGATTAGTTGTCGAGCTTTT
GGAGTACGTCGTCTTTAGGTTGGGGGGAGGGGTTTTATGCGATGGAGTT
TCCCCACACTGAGTGGGTGGAGACTGAAGAGTTAGGCCAGCTTGGCACT
TGATGTAATTCTCCTTGGAAATTTGCCCTTTTTGAGTTTGGATCTTGCCCT
ATTCTCAAGCCTCAGACAGTGGTTCAAAGTTTTTTTTCTTCCATTT CAGGT
GTCGTGAAAAC TACCCCTAAAAGCCACCGGCGTGGCTAGCATGCAAATA
GAGCTCTCCACCTGCTTCTTTCTGTGCCTTTTGCGATTCTGCTTTAGTGC
CATCAGGAGATACTACCTGGGCGCAGTGGAACTGTCCTGGGACTACCGG
CAAAGTGAACCTCCTCCGTGAGCTGCACGTGGACACCAGATTTCTTGCTA
CAGCGCCAGGAGCTCTTCCGTTGGGCCCGTCAGTCCTGTACAAAAAGAC
TGTGTTTCGTAGAGTTCACGGATCAACTTTTCAGCGTTGCCAGGCCCAGG
CCACCATGGATGGGTCTGCTGGGTCTTACCATCCAGGCTGAGGTTTACG
ACACGGTGGTCGTTACCCTGAAGAACATGGCTTCTCATCCCGTTAGTCTT
CACGCTGTCCGGCGTCTCCTTCTGGAATCTTCCGAAGGCGCTGAATATG
AGGATCACACCAGCCAAAGGGAGAAGGAAGACGATAAAGTCCTTCCCGG
TAAAAGCCAAACCTACGTCTGGCAGGTCCTGAAAGAAAATGGTCCAACA
GCCTCTGACCCACCATGTCTCACCTACTCATACTGTCTCACGTGGACCT
GGTGAAAGACCTGAATTCGGGCCCTCATTGGAGCCCTGCTGGTTTGTAGA
GAAGGGAGTCTGACCAGAGAAAGGACCCAGAACCTGCACGAATTTGTAC
TACTTTTTGCTGTCTTTGATGAAGGGAAAAGTTGGCACTCAGCAAGAAAT
GACTCCTGGACACGGGCCATGGATCCCGCACCTGCCAGGGCCCAGCCTG
CAATGCACACAGTCAATGGCTATGTCAACAGGTCTCTGCCAGGTCTGAT
CGGATGTCATAAGAAATCAGTCTACTGGCACGTGATTGGAATGGGCACC
AGCCCGGAAGTGC ACTCCATTTTTCTTGAAGGCCACACGTTTCTCGTGAG
GCACCATCGCCAGGCTTCTTGGAGATCTCGCCACTAACTTTCCTCACTG
CTCAGACATTCTGATGGACCTTGGCCAGTTCCTACTGTTTTGTCATATC
TCTTCCCACCACCATGGTGGCATGGAGGCTCACGTCAGAGTAGAAAGCT
GCGCCGAGGAGCCCCAGCTGCGGAGGAAAGCTGATGAAGAGGAAGATT
ATGATGACAATTTGTACGACTCGGACATGGACGTGGTCCGGCTCGATGG
TGACGACGTGTCTCCCTTTATCCAAATCCGCTCAGTTGCCAAGAAGCATC
CTAAAACCTTGGGTACATTACATTGCTGCTGAAGAGGAGGACTGGGACTA
TGCTCCCTTAGTCTCTCGCCCCGATGACAGAAGTTATAAAAAGTCAATATT
TGAACAATGGCCCTCAGCGGATTGGTAGGAAGTACAAAAAAGTCCGATT
TATGGCATAACAGATGAAACCTTTAAGACTCGTGAAGCTATT CAGCATG
AATCAGGAATCTTGGGACCTTTACTTTATGGGGAAGTTGGAGACACACT
GTTGATTATATTTAAGAATCAAGCAAGCAGACCATATAACATCTACCCTC

ACGGAATCACTGATGTCCGTCCTTTGTATTCAAGGAGATTACCAAAAAGGT
GTAAAACATTTGAAGGATTTTCCAATTCTGCCAGGAGAAATATTCAAATA
TAAATGGACAGTGACTGTAGAAGATGGGCCAACTAAATCAGATCCTCGG
TGCTGACCCGCTATTACTCTAGTTTCGTTAATATGGAGAGAGATCTAGC
TTCAGGACTCATTGGCCCTCTCCTCATCTGCTACAAAAGAATCTGTAGATC
AAAGAGGAAACCAGATAATGTCAGACAAGAGGAATGTCATCCTGTTTTTC
TGTATTTGATGAGAACCGAAGCTGGTACCTCACAGAGAATATAACAACGC
TTTCTCCCAATCCAGCTGGAGTGCAGCTTGAGGATCCAGAGTTCCAAG
CCTCCAACATCATGCACAGCATCAATGGCTATGTTTTTGTAGTTTTGCAG
TTGTCAGTTTGTGGCATGAGGTGGCATACTGGTACATTCTAAGCATTGG
AGCACAGACTGACTTCCTTTCTGTCTTCTCTCTGGATATACCTTCAAAC
ACAAAATGGTCTATGAAGACACACTCACCTATTCCATTCTCAGGAGAA
ACTGTCTTCATGTCGATGGAAAACCCAGGTCTATGGATTCTGGGGTGCC
ACAACCTCAGACTTTCGGAACAGAGGCATGACCGCCTTACTGAAGGTTTC
TAGTTGTGACAAGAACACTGGTGATTATTACGAGGACAGTTATGAAGAT
ATTTCAGCATACTTGCTGAGTAAAAACAATGCCATTGAACCAAGAAGCTT
CTCCCAGAATTCAAGACACCCTAGCACTAGGCCAAAAGCAATTTAATGCCA
CCACAATTCCAGAAAATGACATAGAGAAGACTGACCCTTGGTTTGCACA
CAGAACACCTATGCCTAAAATACAAAATGTCTCCTCTAGTGATTTGTTGA
TGCTCTTGCGACAGAGTCTACTCCACATGGGCTATCCTTATCTGATCTC
CAAGAAGCCAAATATGAGACTTTTTCTGATGATCCATCACCTGGAGCAAT
AGACAGTAATAACAGCCTGTCTGAAATGACACACTTCAGGCCACAGCTC
CATCACAGTGGGGACATGGTATTTACCCCTGAGTCAGGCCTCCAATTAA
GATTAATGAGAACTGGGGACAACCTGCAGCAACAGAGTTGAAGAACT
TGATTTCAAAGTTTCTAGTACATCAAATAATCTGATTTCAAACAATCCATC
AGACAATTTGGCAGCAGGTACTGATAATACAAGTTCCTTAGGACCCCCA
AGTATGCCAGTTCATTATGATAGTCAATTAGATACCACTCTATTTGGCAA
AAAGTCATCTCCCCTTACTGAGTCTGGTGGACCTCTGAGCTTGAGTGAA
GAAAATAATGATTCAAAGTTGTTAGAATCAGGTTTAAATGAATAGCCAAGA
AAGTTCATGGGGAAAAAATGTATCGTCAACAGAGAGTGGTAGGTTATTT
AAAGGGAAAAGAGCTCATGGACCTGCTTTGTTGACTAAAGATGGGAGGA
CTGAAAGGCTGTGCTCTCAAACCCACCAGTCTTGAAACGCCATCAACG
GGAAATAACTCGTACTACTCTTCAGTCAGATCAAGAGGAAATTGACTATG
ATGATACCATATCAGTTGAAATGAAGAAGGAAGATTTTGACATTTATGAT
GAGGATGAAAATCAGAGCCCCCGCAGCTTTCAGAAGAGAACCCGACACT
ATTTCAATTGCTGCGGTGGAGCAGCTCTGGGATTACGGGATGAGCGAATC
CCCCGGGCGCTAAGAAACAGGGCTCAGAACGGAGAGGTGCCTCGGTTT
AAGAAGGTGGTCTTCCGGGAATTTGCTGACGGCTCCTTACGCAGCCGT
CGTACCGCGGGGAACCTCAACAAACACTTGGGGCTCTTGGGACCCTACAT
CAGAGCGGAAGTTGAAGACAACATCATGGTAACTTTCAAAAACCAGGCG
TCTCGTCCCTATTCCTTCTACTCGAGCCTTATTTCTTATCCGGATGATCA
GGAGCAAGGGGCAGAACCTCGACACAACCTTCGTCCAGCCAAATGAAACC
AGAACTTACTTTTGGAAAGTGCAGCATCACATGGCACCCACAGAAGACG
AGTTTGACTGCAAAGCCTGGGCCTACTTTTCTGATGTTGACCTGGAAAAA
GATGTGCACTCAGGCTTGATCGGCCCCCTTCTGATCTGCCGCGCCAACA
CCCTGAACGCTGCTCACGGTAGACAAGTGACCGTGCAAGAATTTGCTCT
GTTTTTCACTATTTTTGATGAGACAAAGAGCTGGTACTTCACTGAAAATG

TGGAAGGAACTGCCGGGCCCCCTGCCACCTGCAGATGGAGGACCCAC
TCTGAAAGAAAATATCGCTTCCATGCAATCAATGGCTATGTGATGGATA
CACTCCCTGGCTTAGTAATGGCTCAGAATCAAAGGATCCGATGGTATCT
GCTCAGCATGGGCAGCAATGAAAATATCCATTTCGATTCATTTTAGCGGAC
ACGTGTTTCAAGTGTACGGAAAAAGGAGGAGTATAAAATGGCCGTGTACAA
TCTCTATCCGGGTGTCTTTGAGACAGTGGAAATGCTACCGTCCAAAGTTG
GAATTTGGCGAATAGAATGCCTGATTGGCGAGCACCTGCAAGCTGGGAT
GAGCACGACTTTCCTGGTGTACAGCAAGAAGTGTGAGACTCCCCTGGGA
ATGGCTTCTGGACACATTAGAGATTTTCAGATTACAGCTTCAGGACAATA
TGGACAGTGGGCCCCAAAGCTGGCCAGACTTCATTATTCCGGATCAATC
AATGCCTGGAGACCAAGGAGCCCTTTTCTTGGATCAAGGTGGATCTGT
TGGCACCAATGATTATTCACGGCATCAAGACCCAGGGTGCCCGTCAGAA
GTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTCTTGATG
GGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGT
CTTCTTTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTAAACC
CTCCAATTATTGCTCGATACATCCGTTTGCACCCAACCTCATTATAGCATT
CGCAGCACTCTTCGCATGGAGTTGATGGGCTGTGATTTAAATAGTTGCA
GCATGCCATTGGGAATGGAGAGTAAAGCAATATCAGATGCACAGATTAC
TGCTTCATCCTACTTTACCAATATGTTTGGCACCTGGTCTCCTTCAAAG
CTCGACTTCACCTCCAAGGGAGGAGTAATGCCTGGAGACCTCAGGTGAA
TAATCCAAAAGAGTGGCTGCAAGTGGACTTCCAGAAGACAATGAAAGTC
ACAGGAGTAACTACTCAGGGAGTAAAATCTCTGCTTACCAGCATGTATGT
GAAGGAGTTCCTCATCTCCAGCAGTCAAGATGGCCATCAGTGGACTCTC
TTTTTTCAGAATGGCAAAGTAAAGGTTTTTTCAGGGAAATCAAGACTCCTT
CACACCTGTGGTGAACCTCTCTAGACCCACCGTTACTGACTCGCTACCTTC
GAATTCACCCCCAGAGTTGGGTGCACCAGATTGCCCTGAGGATGGAGGT
TCTGGGCTGCGAGGCACAGGACCTCTACTGAGCTCGAGTCTAGAGGGCC
CGTTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCAT
CTGTTGTTTGGCCCTCCCCCGTGCCTTCCCTGACCCTGGAAGGTGCCACT
CCCCTGTCCTTTCCTAATAAAAATGAGGAAATTGCATCGCATTGTCTGAG
TAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGG
GGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTC
TATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCTAGGGGGTATCCC
CACGCGCCCTGTAGCGGCGCATTAAAGCGCGGGCGGGTGTGGTGGTTACGC
GCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGC
TTTCTTCCCTTCCTTCTCGCCACGTTTCGCCGGCTTTCCCCGTCAAGCTC
TAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTC
GACCCCAAAAATTTGATTAGGGTGTGGTTACGTTAGTGGGCCATCGC
CCTGATAGACGGTTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAAT
AGTGGACTCTTGTTCCAAACCTGGAACAACACTCAACCCTATCTCGGTCTA
TTCTTTTGGATTTATAAGGGATTTTGGCCGATTTTCGGCCTATTGGTTAAAAA
ATGAGCTGATTTAACAAAAATTTAACGCGAATTAATTCTGTGGAATGTGT
GTCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTAT
GCAAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTGGAAAGTCCCCA
GGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAG
CAACCATAGTCCCGCCCCCTAACTCCGCCCATCCCGCCCCCTAACTCCGCC
AGTTCCGCCCATCTCCGCCCATGGCTGACTAATTTTTTTTATTATGC

AGAGGCCGAGGCCGCCTCTGCCTCTGAGCTATTCCAGAAGTAGTGAGGA
GGCTTTTTTGGAGGCTCGACAGTACTAAGCTTTGCCCACTGTGGGGTGG
AGGGGACAGATAAAAGTACCCAGAACCAGAGCCACATTAACCGGCCCTGGG
AATATAAGGTGGTCCCAGCTCGGGGACACAGGATCCCTGGAGGCAGCAAAC
ATGCTGTCCTGAAGTGGACATAGGGGCCCCGGGTTGGAGGAAGAAGACTAGCT
GAGCTCTCGGACCCCTGGAAGATGCCATGACAGGGGGCTGGAAGAGCTAGC
ACAGACTAGAGAGGTAAGGGGGGTAGGGGAGCTGCCAAATGAAAGGAGTG
AGAGGTGACCCGAATCCACAGGAGAACGGGGTGTCCAGGCAAAGAAAGCAA
GAGGATGGAGAGGTGGCTAAAGCCAGGGAGACGGGGTACTTTGGGGTTGTC
CAGAAAACGGTGATGATGCAGGCCTACAAGAAGGGGAGGCGGGACGCAAG
GGAGACATCCGTCGGAGAAGGCCATCCTAAGAAACGAGAGATGGCACAGGC
CCCAGAAGGAGAAGGAAAAGGGAACCCAGCGAGTGAAGACGGCATGGGGTT
GGGTGAGGGAGGAGAGATGCCCGGAGAGGACCCAGACACGGGGAGGATCCG
CTCAGAGGACATCACGTGGTGCAGCGCCGAGAAGGAAGTGCTCCGGAAAGA
GCATCCTTGGGCAGCAACACAGCAGAGAGCAAGGGGAAGAGGGAGTGAGG
AAGACGGAACCTGAAGGAGGCGGCAGGGAAGGATCTGGGCCAGCCGTAGAG
GTGACCCAGGCCACAAGCTGCAGACAGAAAGCGGCACAGGCCAGGGGAGA
GAATGCAGGTCAGAGAAAGCAGGacctgcctgggaaggggaacagtgggccagagcgggcgcaga
agccagtagagctcaaagtggccgactcaggagagagacggcagcgtagaggccagagttccggcgccacagcaagg
gcactcgggggagagaggaggcagcgcaaatggccagggccaggcagatagaccagactgagctatgggag
ctggctcaggttcaggagagggcagggcaggggaaggagacaaagtccaggaccggctggaggggctcaacatcggaaga
ggggaagtcgagggagggatggaaggaggactgcatgggtcagcacaggctgcaaagccagggccagttaaagcgact
ccaatgcggaagagagtaggtcgaaggggaatggaaggagcctggggcagagtggtcagcacagagtggttaagcca
gggcccagttgaagcgctccaattcggaagtggggtggtcgaaggggaatggaaggggactgggacgggggtgcagcat
agggtggcaaacggccagggccaggaacgacggggcggatcgagactggcaacgggggaaggagatgccccaggtggcg
cagcagaggggtggacctggccccgggagacccggggcggggggcgctgacctggtgcagggcgctgataccgtcgcgct
ggtggagtccagcacggcgggcgggcgggcgggcgggcggggctgagctcggcgccggggccaggtcgggcgggcg
gcagcatcagacgcgctcgtccaggtcggcgccgacagggccaggaactcggcgggcgctcgaagcggacgggtg
cggggcgggcgctcctccggggccaggtcggcgcccggcgccccactgcccagctgctcccgtcggcgctcccg
ggcagccggcgccgccccgggcccagccggggccatcctcggacatcgaccggcgcccggcagcgagc
gagcgagcggcgagcccaaccggcgccaccaccggcgcccggcgccccggggggccggggaactgcccgtggc
ccccaccgccccaaagatctcccgtccccggcggtgctgacgtcacggcgctgccccaggggtgctgggacaggtc
gcggggagcgtgggaaatggagtccattagcagaagtggccctggccacttcaggagtcgctgtgccccgatgcactg
ggaagtccgcagctccgagggcccagtggaatcgccagatgagggcctcctccggggaatgctgggaaatggagtctac
aggccggaggggtgccccacggcactaggaagtgtgtagcaccgggtaaggggatgaatagcagactgccccggggc
agttaggaattcgactggacagccgctgggaggagtgcggggagaggcagagttgtttgtattgtttatgttttttt
gtttgagacggagctcgtctgtcggcagctggagttcagtgggcgatcggctcgtgcaacctccgctcccaggttca
agcaattctggctcagttcccagagtagctgggataacagggcgcgccaccggcctgctaattttatatttttagtagac
gggattccacatgttgccatgatggtcttgcactcattgatccggcgctcggcctgtaactctgctgggatgacga
gcgtaagccacatgcccagctgggttttattttgtttttctgacccttaactagaataagctccacgagagcgggatc
ttttgtcttctgtgactactgtcctcggttcttagaacagaacctgagagaacctgatcgaatatttttgaatgaatgaat
gggttcaccagggcaccatgggaaactgagtcgcaacctagaagccatgaaagacagtccactccaagcttc

Supplementary Table S2. *In silico* predicted ten most likely AAVS1 ZFN off-target sites

Potential off-target sites, OT1 to OT10 were ranked according to their similarities to the consensus ZFN target site determined by SELEX. Score for each SELEX predicted sequence was derived from experimentally determined base-frequency matrices (Hockemeyer, D, *et al.* (2009) *Nat Biotechnol* **27**: 851-857). Chromosomal locations (hg19), primers used for amplifying these genomic regions and amplicon sizes are listed.

Site	Score	Chromosomal location	SELEX predicted sequence	Forward primer	Reverse primer	Amplicon size (bp)
OT1	1.47E-09	chr8:141507033-141507064	GcTCCTGGCCCagTG CTGGCCACTGTGGG TGC	ttaagaactgtaacctatttccaa agtgtttg	cag cct ggc caa cat ggt gaa ac	389
OT2	8.39E-10	chr10:47635366-47635397	ACACCCACAGgGGC AGGGGcAGGGCCAG GAcT	aaggtgtaagtggagccacaagg ct	tgtggtccttgctggatcag gaa	308
OT3	5.76E-10	chr4:3303431-3303461	TTTCCTGTCCtTtACC TGCCACTGTGGGTtT	ttgaaataagaccatttgtgat gaga	ctggctcattccaacgtcca tgt	389
OT4	2.67E-10	chr10:117758694-117758724	TCACCCACAGafTTG TAATAGGGACAGGA TT	gacttggtggtggcagaatacac c	gggtaaggtcagataggg ctgtaagactc	601
OT5	1.93E-10	chr9:138563409-138563439	GCACCCACAGcGcA GTGCCcAGGGCCAGG AAC	ggaacaaggcacctggctcc	ccattcccgggagaaatct c	353
OT6	1.24E-10	chr14:101033117	GgTCCTGTCCCTgTG	tgagttgggctgaggtcatc	ggcttggaacaccaggt	320

		-101033148	GGACCCACaGTGGG gGC		g	
OT7	1.13E-10	chr7:50670960- 50670990	GTcCCTGTCCCTATA TCCACACTGTGGcTG G	cttgagtttagcagctccaggaa cc	gtttatcttcataaggtagt gggcagatgg	631
OT8	9.84E-11	chr16:28996789- 28996820	CATCCTGGCCaTgTT GATGgCACTGTGtGT GC	ggtcctcacccatcttcac	aaagagagggctggtgag gc	375
OT9	8.89E-11	chr19:1224727- 1224758	TTTCaTGaCCCTgCTA AGCCCACTGTGGGT GG	gttgcgagagtcctactgg	agcctgaagttgagcctgt c	348
OT10	6.35E-11	chr12:50284986- 50285017	CCACCCACAGgGcA GCCAGgAGGGACAG GATG	cacagagtcaggggatcgt	gccactttgtattgggtggt	650

Supplementary Table S3

Gene Ontology classification of 90 transcripts whose levels were ≥ 2 -fold different in puro-CLECs compared to wt-CLECs. RNA-Seq was performed on wt-CLECs and puro-CLECs 6 weeks after electroporation of donor DNA and AAVS1 ZFN. Potential oncogenes and tumour suppressor genes (underlined and italicized) did not map to any pathway in puro-CLEC transcriptome analysed by DAVID.

GENE ONTOLOGY CLASSIFICATIONS	OVEREXPRESSED GENES
Apoptosis	ZC3H12A, SERPINB2
Cell growth	TNFRSF12A
Cellular transport	HEPHL1
Cytoskeleton	ITAG2, KRT34, TUBA1A
Differentiation and development	IL8
Immune response	DPP4, IL1A, IL1B, SLPI, SAA1, SAA2, SAA4, C3, <u>CXCL1</u> , <u>CXCL2</u> , <u>CXCL3</u> , CXCL4, CXCL5, CXCL6
Metabolism	HMGCS1, DHCR7, ALDH1A3, <u>ENCI</u> , FDPS, FADS1, FADS2, FASN, GPAM, HAS1, IDI1, LDLR, SCD, VNN3
Others	C4orf26, F8, PI3
Signalling	IQGAP2, MRGPRX3, BDKRB1, BEX1, DUSP6, INSIG1, LIF, OR10J1, EGR2
Transcription/ translation	NRK, DIO2, EHF
Unknown	KIAA1199, FAM84B, KLHL5, PLD5, TMEM154, TMEM158
UNDEREXPRESSED GENES	
Cell cycle	HSPA2

Cell growth	APOE, CLEC11A, <u>IGFBP5</u> , SPP1
Cytoskeleton	PALM, TNC
Cellular transport	KCNE3
Differentiation and development	<u>FGF9</u> , <u>GPC3</u> , PPL
Immune response	CFD, SERPINA3
Metabolism	<u>BGN</u> , CKB, GPX3, IMPA2, PTGDS, PTGS1
Others	H19, VWA1
Signalling	CNR1, IFITM1, NOTCH3, NRP3
Transcription/ translation	CRABP2, CYTL1, HR, PDLIM1, TCEA3
Unknown	C1orf21, SCRG1, TMEM119

Supplementary Table S4

Relative transcript levels of 196 selected genes from RNA-Seq data showing their expressions in puro-CLEC compared to wt-CLEC. Altered expression was defined as ≥ 2 -fold difference.

Gene symbol	Expression compared to wt-CLEC
Neighboring genes within 1 Mb of AAVS1 locus	
BRSK1	No change
C19ORF51 (DNAAF3)	Not expressed
COX6B2	No-change
EPS8L1	No-change
FAM71E2	Not expressed
FCAR	No change
FIZ1	No change
GP6	Not expressed
HSPBP1	No change
ISOC2	No change
KIR2DL1	Not expressed
KIR2DL3	Not expressed
KIR2DL4	Not expressed
KIR2DS4	Not expressed
KIR3DL1	Not expressed
KIR3DL2	Not expressed
KIR3DL3	Not expressed
LILRA1	Not expressed
LILRA2	Not expressed
LILRB1	Not expressed
LILRB4	Not expressed
LILRP2	Not expressed

NAT14	No change
NCR1	Not expressed
NLRP2	Not expressed
NLRP7	Not expressed
PPP6R1	No change
PTPRH	No change
RDH13	No change
SBK2	Not expressed
SGK110	Not expressed
SHISA7	Not expressed
SSC5D	No change
SUV420H2	No change
SYT5	Not expressed
TMEM150B	Not expressed
TMEM190	No change
TMEM238	No change
TMEM86B	No change
TNNI3	Not expressed
TNNT1	No change
ZNF524	No change
ZNF579	No change
Predicted interacting partners of PPP1R12C as determined by Gene Network Central™ and STRING 10	
BRAF	No change
CAMKK1	No change
CCNA2	Not expressed
CCND3	No change
DUSP1	No change
DUSP16	No change

DUSP6	Up-regulated 4.9-fold
MAPK3	No change
MPRIP	No change
MYL2	No change
MYLPF	Not expressed
MYL5	No change
MYLK	No change
MYL7	Not expressed
MYH9	No change
MYL9	No change
MYL10	Not expressed
MYH11	No change
MYL12A	No change
MYL12B	No change
MYH14	Not expressed
PPP1R12A	No change
PRKCE	No change
PRKCI	No change
RAF1	No change
RHOA	No change
ROCK 1	No change
ROCK2	No change
WIPF1	No change
Predicted downstream effector genes from publications^{1,2}	
CDC42	No change
CDC42BPA	No change
CDC42BPB	No change
CDC42BPG	Not expressed
CDC42EP1	Not expressed

CDC42EP2	No change
CDC42EP3	No change
CDC42EP4	No change
CDC42EP5	No change
CDC42SE1	No change
CDC42SE2	No change
EZR	No change
MYH1	Not expressed
MYH10	No change
MYH13	No change
MYH15	No change
MYH16	Not expressed
MYH2 /// MYH4	Not expressed
MYH3	No change
MYH6	Not expressed
MYH7	Not expressed
MYH7B	Not expressed
MYH8	Not expressed
MYL1	Not expressed
MYL3	Not expressed
MYL4	Not expressed
MYL6	No change
MYL6B	No change
MYLIP	No change
MYLK2	Not expressed
MYLK3	No change
MYLK4	No change
RDX	No change
RHO	Not expressed
RHOB	No change

RHOBTB1	No change
RHOBTB2	No change
RHOBTB3	No change
RHOC	No change
RHOD	No change
RHOF	No change
RHOG	No change
RHOH	No change
RHOJ	No change
RHOQ	No change
RHOT1	No change
RHOT2	No change
RHOU	No change
RHOV	No change
ROCK2	No change
Members of the protein phosphatase family	
PPP1CA	No change
PPP1CB	No change
PPP1CC	No change
PPP1R10	No change
PPP1R11	No change
PPP1R12B	No change
PPP1R12C	Not expressed
PPP1R13B	No change
PPP1R13L	No change
PPP1R14A	No change
PPP1R14B	No change
PPP1R14C	No change
PPP1R14D	Not expressed
PPP1R15A	No change

PPP1R15B	No change
PPP1R16A	No change
PPP1R16B	Not expressed
PPP1R17	Not expressed
PPP1R18	No change
PPP1R1A	Not expressed
PPP1R1B	Not expressed
PPP1R1C	No change
PPP1R2	No change
PPP1R21	No change
PPP1R26	No change
PPP1R27	Not expressed
PPP1R2P9	Not expressed
PPP1R32	No change
PPP1R35	No change
PPP1R36	No change
PPP1R37	No change
PPP1R3A	Not expressed
PPP1R3B	No change
PPP1R3C	No change
PPP1R3D	No change
PPP1R3E	No change
PPP1R3F	No change
PPP1R7	No change
PPP1R8	No change
PPP1R9A	Not expressed
PPP1R9B	No change
PPP2CA	No change
PPP2CB	No change
PPP2R1A	No change

PPP2R1B	No change
PPP2R2A	No change
PPP2R2B	Not expressed
PPP2R2C	No change
PPP2R2D	No change
PPP2R3A	No change
PPP2R3B	Not expressed
PPP2R3B-AS1	Not expressed
PPP2R3C	No change
PPP2R4	No change
PPP2R5A	No change
PPP2R5B	No change
PPP2R5C	No change
PPP2R5D	No change
PPP2R5E	No change
PPP3CA	No change
PPP3CB	No change
PPP3CC	No change
PPP3R1	No change
PPP3R2	Not expressed
PPP4C	No change
PPP4R1	No change
PPP4R1L	No change
PPP4R2	No change
PPP4R4	No change
PPP5C	No change
PPP6C	No change
PPP6R1	No change
PPP6R2	No change
PPP6R3	No change

1. Surks HK, Mendelsohn ME (2003) Dimerization of cGMP-dependent protein kinase 1alpha and the myosin-binding subunit of myosin phosphatase: role of leucine zipper domains. *Cell Signal* **15**: 937 - 944.
2. Mulder J, Ariaens A, van den Boomen D, Moolenaar WH (2004) p116Rip targets myosin phosphatase to the actin cytoskeleton and is essential for RhoA/ROCK-regulated neuriteogenesis. *Mol Biol Cell* **15**: 5516 - 5527.

Supplementary Table S5. List of PCR primers used in this study

Primer sequences, amplicon sizes, PCR annealing temperatures, the intended purpose of the primers and the figures corresponding to their use are listed as Reference figure number.

Purpose	Forward sequence (5' → 3')	Reverse sequence (5' → 3')	Amplicon size (bp)	Annealing temperature (°C)	Reference figure number
Primers for AAVS1 ZFN study					
<u>Primers for mutagenesis of <i>Fok1</i> endonuclease</u>					
E490K mutation	gca acg ata tgt caa aga aaa tca aac acg	cgt gtt tga ttt tct ttg aca tat cgt tgc	Not applicable	55	1
I538K mutation	cac gat taa atc ata aga cta att gta atg gag c	gct cca tta caa tta gtc tta tga ttt aat cgt g	Not applicable	55	1
Q468E mutation	cca agc aga tga aat gga acg ata tgt cga ag	ctt cga cat atc gtt cca ttt cat ctg ctt gg	Not applicable	55	1
I499L mutation	cac gaa aca aac atc tca acc cta atg aat gg	cca ttc att agg ggt gag atg ttt gtt tcg tg	Not applicable	55	1
S418P mutation	att gaa att gcc aga aat ccc act cag gat aga att ctt	aag aat tct atc ctg agt ggg att tct ggc aat ttc aat	Not applicable	55	1
K441E mutation	gtt tat gga tat aga ggt gaa cat ttg ggt gga tca agg	cct tga tcc acc caa atg ttc acc tct ata tcc ata aac	Not applicable	55	1
H537R mutation	cag ctt aca cga tta aat cgt aag act aat tgt	tcc att aca att agt ctt acg att taa tcg tgt aag	Not applicable	55	1

	aat gga	ctg			
N496D mutation	gaa gaa aat caa aca cga gac aaa cat ctc aac cct aat	att agg gtt gag atg ttt gtc tcg tgt ttg att ttc ttc	Not applicable	55	1
Primers for RFLP assay, CEL-1 assay and integration junction PCR					
AAVS1 ZFN left RT-PCR	gcagacaggccc tggaca	cccaggtgtttctct g	180	62	S1
AAVS1 ZFN right RT-PCR	agaaacttcaccc tgcaga	gtcagctggctgtgt ct	180	62	S1
Integration junction PCR to detect integration of 50-bp donor DNA	agcttgaattctct agaaatattctcg aggtttaaacgtc gacgc	ggaacggggctca gtctg	1026	62	S2B, 3A
RFLP PCR to detect integration of 50-bp donor DNA	ggcctggccatt gtcactt	ggaacggggctca gtctg	1880	62	2, 3A, S2B
Surveyor mutation detection PCR (CEL-1 assay)	ttcgggtcacctct cactcc	ggctccatcgtgaag caaacc	514	60	S2A
Control PCR to amplify AAVS1 genomic region	aagaagegcacc acctccaggttct c	atgacctatgtctct tggcctcgta	976	60	3, 5, 10, S2B, S4B, S5B
Left integration junction PCR to detect integration of 4-kb GFP donor DNA	ggcctggccatt gtcactt	cgtaaataggggg cgtaactggcatatg atac	1237	62	3B

Right integration junction PCR to detect integration of 4-kb GFP donor DNA	gacatagcgttgg ctaccctgatatt gctgaagagc	ggaacggggctca gtctg	1356	62	3B
Left integration junction PCR to detect integration of 9-kb hybrid FVIII donor DNA	ggccctggccatt gtcactt	cgtaaataggggg cgtaactggcatatg atac	1237	62	3C
Right integration junction PCR to detect integration of 9-kb hybrid FVIII donor DNA	gacatagcgttgg ctaccctgatatt gctgaagagc	ggaacggggctca gtctg	1356	62	3C
Left integration junction PCR to detect AAVS1-SA-2A-Puro-pA integration	Puro LF: ggccctggccatt gtcactt	Puro LR: cggatcatctcgagc ctaggg	1103	62	S4B
Right integration junction PCR to detect AAVS1-SA-2A-Puro-pA integration	Puro RF2: tcaccgagctgc aagaact	AAVS1 R: ggaacggggctca gtctg	1621	62	S4B
PCR to detect complete integration of AAVS1-SA-	AAVS1 F: ctggccattgtca ctttgcg	AAVS1 R: ggaacggggctca gtctg	3084	62	S4B

2A-Puro-pA					
Left integration junction PCR to detect AAVS-CAGGS-EGFP integration	AAVS1 F: ctggccattgtca ctttgcg	Puro LR: cggtcatctcgagc ctaggg	1099	62	S5B
Right integration junction PCR to detect AAVS-CAGGS-EGFP integration	GFP RF: ctactcccagtca tagctgtc	Puro RR: cttgccacgtaac ctgaga	1370	62	S5B
Genomic PCR to detect left half of AAVS-CAGGS-EGFP integration	AAVS1 F: ctggccattgtca ctttgcg	CAGGS R: gatggggagagtg aagcagaacg	2382	60	S5B
Genomic PCR to detect right half of AAVS-CAGGS-EGFP integration	GFP F: aaacggccacaa gttcagcg	Puro RR: cttgccacgtaac ctgaga	2517	60	S5B
Left integration junction PCR to detect SA-2A Puro hybrid FVIII integration	AAVS1 F: ctggccattgtca ctttgcg	Puro LR: cggtcatctcgagc ctaggg	1099	62	5B
Right integration	C1F: acttctcgtgtga	Puro RR: cttgccacgtaac	3187	60	5B

junction PCR to detect SA-2A Puro hybrid FVIII integration	cagcaagaagtg tcagactcccctg gga	ctgaga			
Genomic PCR to detect left half of SA-2A Puro hybrid FVIII integration	AAVS1 F: ctggccattgtca ctttgcg	A2R: gcgggggctctgat tttcatcctc	6859	60	5B
Genomic PCR to detect right half of SA-2A Puro hybrid FVIII integration	A3F: agctttcagaaga gaaccgacac	Puro RR: cttgccacgtaac ctgaga	4156	60	5B
Genomic PCR to detect SA-2A Puro hybrid FVIII integration (Vector)	cgagatgaccga gtacaag	ctcgtagaagggg aggttgc	521	60	10
Left integration junction PCR to detect SA-2A Puro hybrid FVIII integration (Left)	ttcgggtcacctct cactcc	gacgcgcgtgaga ggaagagttc	602	60	10
Right integration junction PCR to detect SA-2A Puro hybrid FVIII integration (Right)	caaagcatgcatc tcaattagtc	ggctccatcgtaag caaacc	551	60	10

Control PCR to amplify AAVS1 genomic region (Control)	tcccctcccagaa agacctgc	gaccctcagccctg accagc	510	60	10
Primers to verify transcriptome data by RT-PCR					
<i>LILRB4</i>	ggacattgccc agagacag	gtcttcacgtgtgg gctct	197	60	9
<i>ISOC2</i>	tgacggagcagt accacaag	ggataagaacggg gtccaagatg	196	60	9
<i>PPP6R1</i>	gcgctacaagta ccccagtgc	tccgaagaaagga cacgagc	215	60	9
<i>NAT14</i>	gtcctccgaaacct tgtcgaa	ccttcacgccggcc ttc	175	60	9
<i>ZNF579</i>	aaggaggcgag gcatggat	gtaggggaaacgg aagaggc	184	60	9
<i>FIZ1</i>	cagagggaggtg gagagccc	ctgtgcttgaaccc ttgcc	235	60	9
<i>RDH13</i>	ggtecccttccca gctgaa	atgatgtgcctcct ctcctg	278	60	9
<i>DUSP1</i>	ggatacgaagcg tttcggc	ggccaccctgatcg tagagt	151	60	9
<i>DUSP6</i>	acctggaaggtg gcttcagta	accatccgagtctgt tgcac	197	60	9
<i>CDC6</i>	agaagggcccca tgattgtg	tgcagcattgtcca gaacct	296	60	9
<i>DUSP16</i>	ctgcttgcaggtg ggtttg	gcacacgcaggaa atgagac	271	60	9
Primers for analysis of potential off-target sites of AAVS1 ZFN					
OT1	ttaagaactgtaa cctatttccaaag tgtttg	cag cct ggc caa cat ggt gaa ac	389	62	Table 1

OT2	aagggtgaagtg gagccacaaggc t	tgtggtccttgctgg atcaggaa	308	60	Table 1
OT3	ttggaataagac ccatttggatga ga	ctggctcattccaac gtccatgt	389	58	Table 1
OT4	gacttggtggtg gcagaatacacc	gggtaaggtcagat agggtgtaagact c	601	60	Table 1
OT5	ggaacaaggcac ctggctcc	ccattcccgggaga aatctc	353	58	Table 1
OT6	tgagttgggct gaggtcatc	ggcttgaaacacc caggtg	320	60	Table 1
OT7	cttgagtttagca gctccaggaac c	gtttatcttcataag gtagtgggcagatg g	631	60	Table 1
OT8	ggtcctcacccca tcttcac	aaagagagggtg gtgaggc	375	60	Table 1
OT9	gttgcgagagtcc ctactgg	agcctgaagttgag cctgtc	348	60	Table 1
OT10	cacagagttcag gggatcgt	gccacttgtattgg gtggt	650	60	Table 1

Supplementary Table S6. Quality metrics of whole-genome sequence dataData statistics

	Wt-CLEC	Puro-CLEC
Number of reads	821345674	1730505362
Data size	73921110660	155745482580
N of fq1	676752	2187303
N of fq2	267401	1037535
GC % of fq1	39.87-39.89	35.36-36.25
GC% of fq2	39.91-39.96	35.32-36.29
Q20% of fq1	98.01-98.70	88.68-98.83
Q20% of fq2	96.32-96.81	92.67-96.77
Q30% of fq1	93.50-95.58	79.40-95-86
Q30% of fq2	90.57-91.87	84.15-91.86
Discard reads related to N	927398	3948294
Discard reads related to low quality	40193038	110992072
Discard reads related to adapter	308374	598522
Clean data/raw data	95.20%	93.74%

Alignment statistics

	Wt-CLEC	Puro-CLEC
Clean reads	903578022	1730505362
Clean bases (bp)	81322021980	155745482580
Mapped reads	876624167	1241671029
Mapped bases (bp)	78104534645	108632498375
Mapping rate	97.02%	71.75%
Uniq reads	841199183	1192115829
Uniq bases (bp)	74952412021	104325871251
Unique rate	95.9%	96.01%
Duplicate reads	98080273	465546767
Duplicate rate	11.19%	37.49%
Mismatch bases (bp)	261346135	498326197
Mismatch rate	0.33%	0.46%
Average sequencing depth	23.88	23.53
Coverage	99.69%	99.70%

SNP statistics

	Wt-CLEC	Puro-CLEC
Total	3518620	3493513
1000Genome and dbsnp135	3175559	3151218
1000Genome specific	18798	19080
dbsnp135 specific	174653	173505
dbSNP rate	95.21%	95.1%
Shared SNPs	94.2%	94.8%
Novel	149610	149710
Homozygous	1454531	1446386
Heterozygous	2064089	2047127
Synonymous	10549	10352
Missense	9346	9170
Stopgain	65	70
Stoploss	35	33
Exonic	19713	19345
Exonic and splicing	282	280
Splicing	144	141
ncRNA	92955	92102
5'UTR	3867	3641
5'UTR and 3'UTR	10	9
3'UTR	22668	22414
Intronic	1197882	1185485
Upstream	18258	17505
Upstream and downstream	588	567
Downstream	20176	19914
Intergenic	2142077	2132110
SIFT	1153	1138
Ti/Tv	2.0147	2.0157
dbSNP Ti/Tv	2.0676	2.0654
Novel Ti/Tv	1.2047	1.2468

Indel statistics

	Wt-CLEC	Puro-CLEC
Total	542077	533792
1000genome and dbsnp135	245282	243502
1000genome specific	66265	66235
dbSNP135 specific	108169	104510
dbSNP rate	65.20%	65.20%
Novel	122361	119545
Homozygous	254640	251241
Heterozygous	287437	282551
Frameshift insertion	108	109
Non-frameshift insertion	80	71
Frameshift deletion	95	89
Non-frameshift deletion	115	107
Frameshift block substitution	0	0
Non-frameshift block substitution	0	0
Stopgain	4	3
Stoploss	1	1
Exonic	398	375
Exonic and splicing	5	5
Splicing	67	65
ncRNA	13931	13670
5'UTR	410	369
5'UTR and 3'UTR	3	3
3'UTR	4427	4269
Intronic	192847	188567
Upstream	2853	2724
Upstream and downstream	95	86
Downstream	3646	3564
Intergenic	323395	320095

Copy number variant statistics

	Wt-CLEC	Puro-CLEC
Total	3793	3127
Exonic	813	757
Exonic and splicing	0	0
Splicing	191	184
ncRNA	150	140
5'UTR	1	0
5'UTR and 3'UTR	0	0
3'UTR	8	2
Intronic	740	559
Upstream	50	44
Upstream and downstream	0	3
Downstream	32	30
Intergenic	1808	1408
Amplification size	12882600	12651500
Deletion size	77031800	64108800