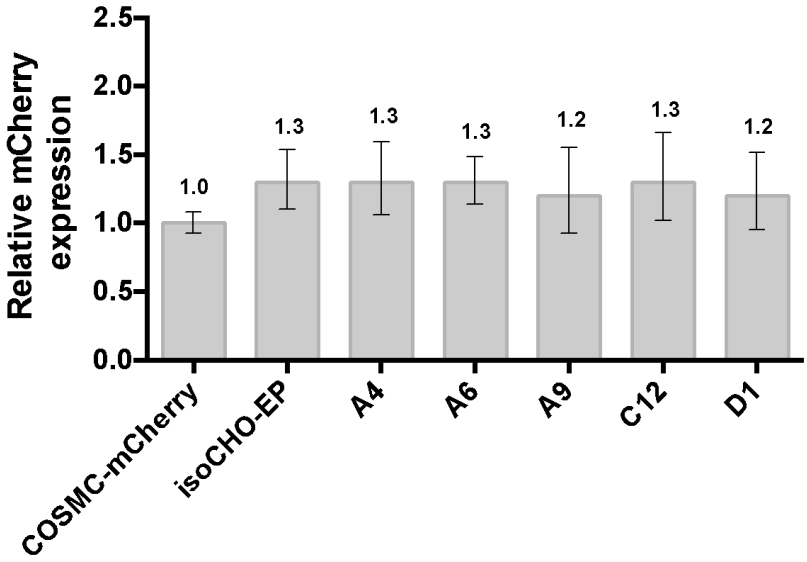
Supporting Information

Minimizing clonal variation during mammalian cell line engineering for improved systems biology data generation

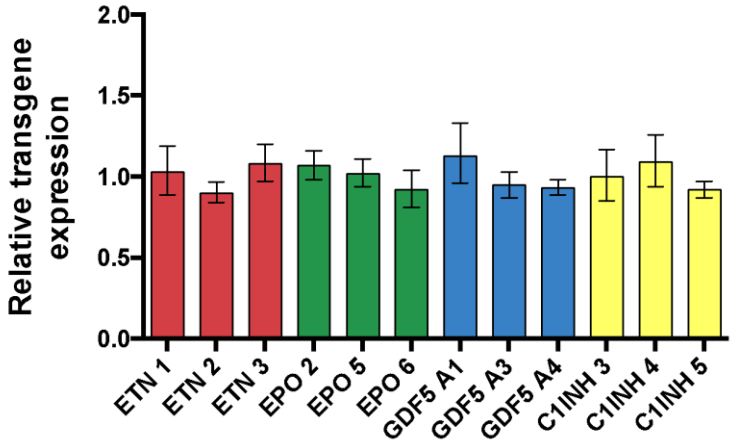
Lise Marie Grav[1], Daria Sergeeva[1], Jae Seong Lee[1,2], Igor Marin de Mas[1], Nathan E. Lewis[3, 4], Mikael Rørdam Andersen[5], Lars Keld Nielsen[1,6], Gyun Min Lee[1,7], Helene Faustrup Kildegaard[1]

[1]The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kgs. Lyngby, Denmark
[2]Department of Molecular Science and Technology, Ajou University, Suwon, Republic of Korea
[3]Department of Pediatrics, University of California, San Diego, United States
[4]The Novo Nordisk Foundation Center for Biosustatainability, University of California, San Diego, United States
[5]Department of Biotechnology and Biomedicine, Technical University of Denmark, Kgs. Lyngby, Denmark
[6]Australian Institute for Bioengineering and Nanotechnology, University of Queensland, Brisbane, Australia
[7]Department of Biological Sciences, KAIST, Daejeon, Republic of Korea

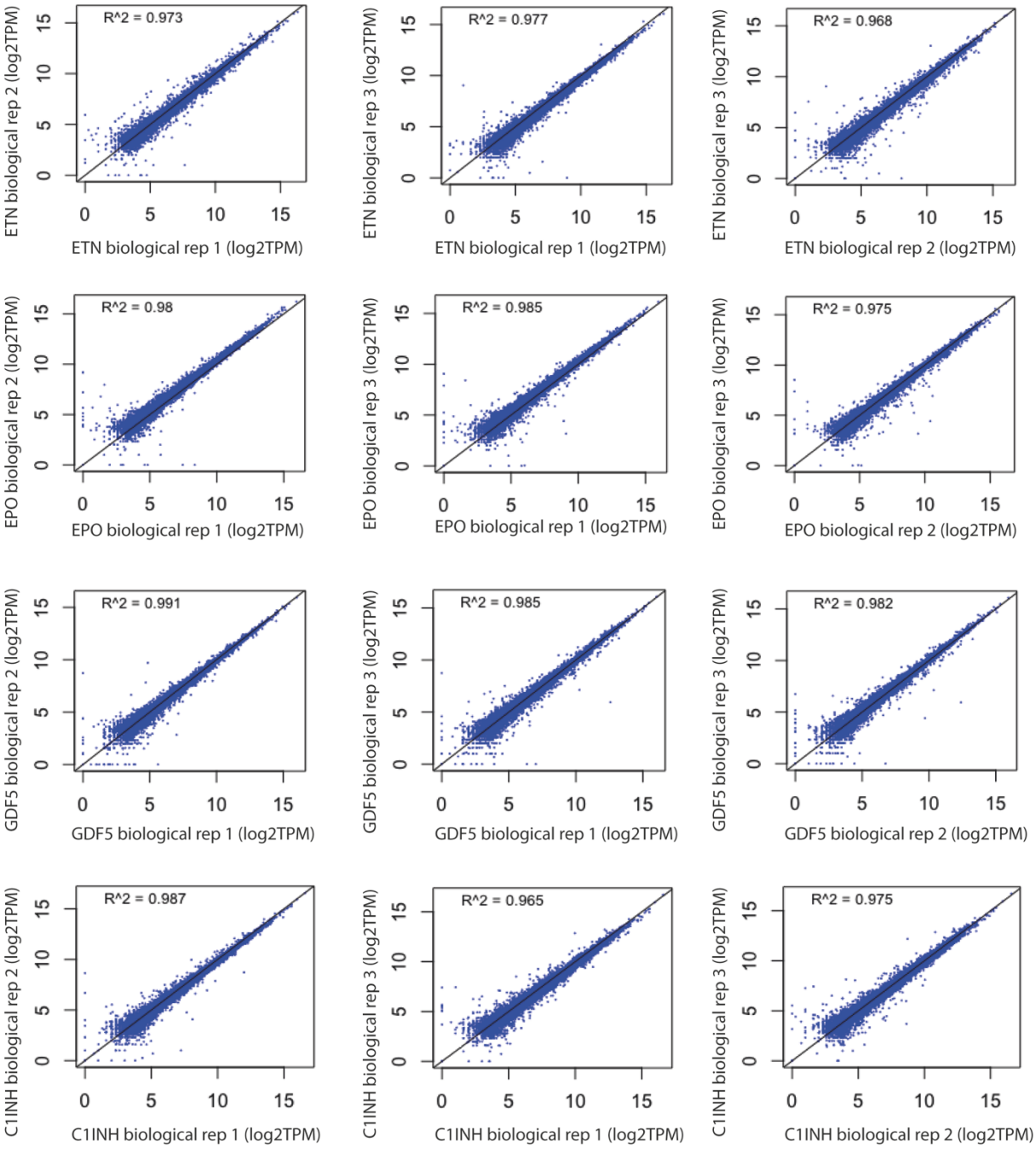
Supporting Figures



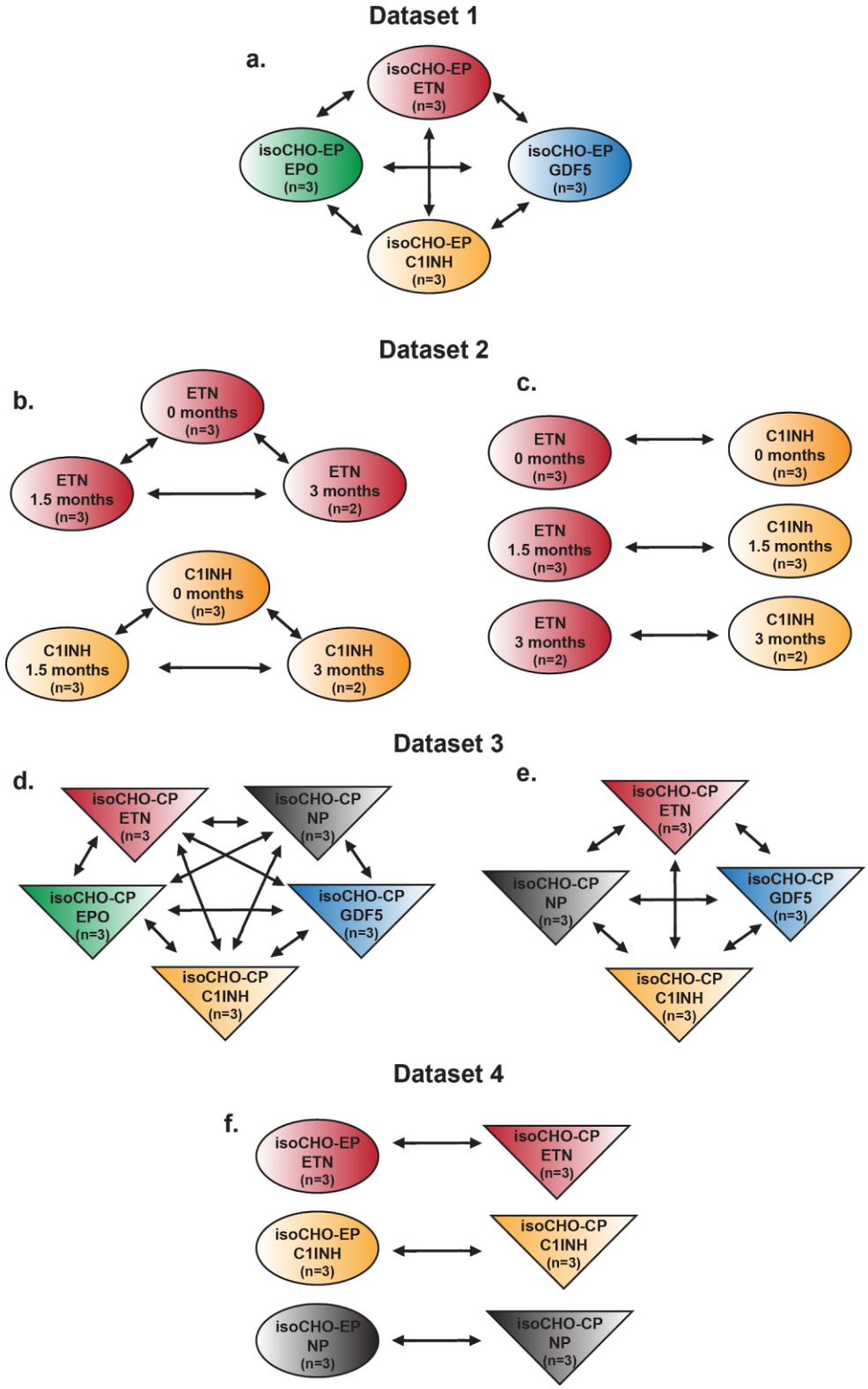
Supporting Figure S1 Relative mCherry expression levels of the panel of mCherry-EP clones, compared to COSMC-mCherry clone, generated in previous study[7]**.** Error bars represent the standard deviations of technical replicates (n≥3).



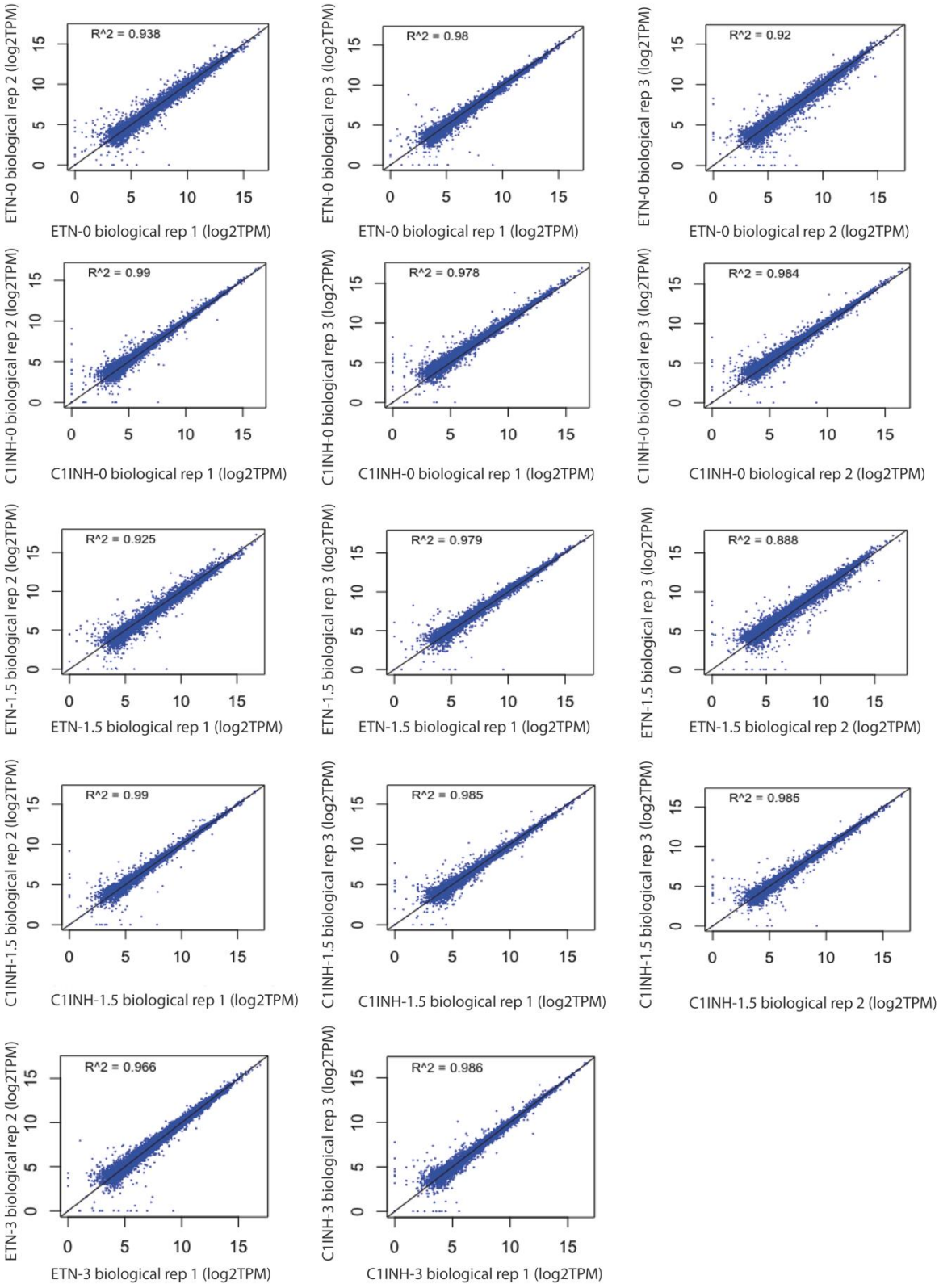
Supporting Figure S2 Relative levels of transgene expression in isoCHO-EP-derived subclones, measured by qRT-PCR, normalized to average value of corresponding isoCHO-EP subclones. The error bars represent the standard deviations of technical replicates (n=3).



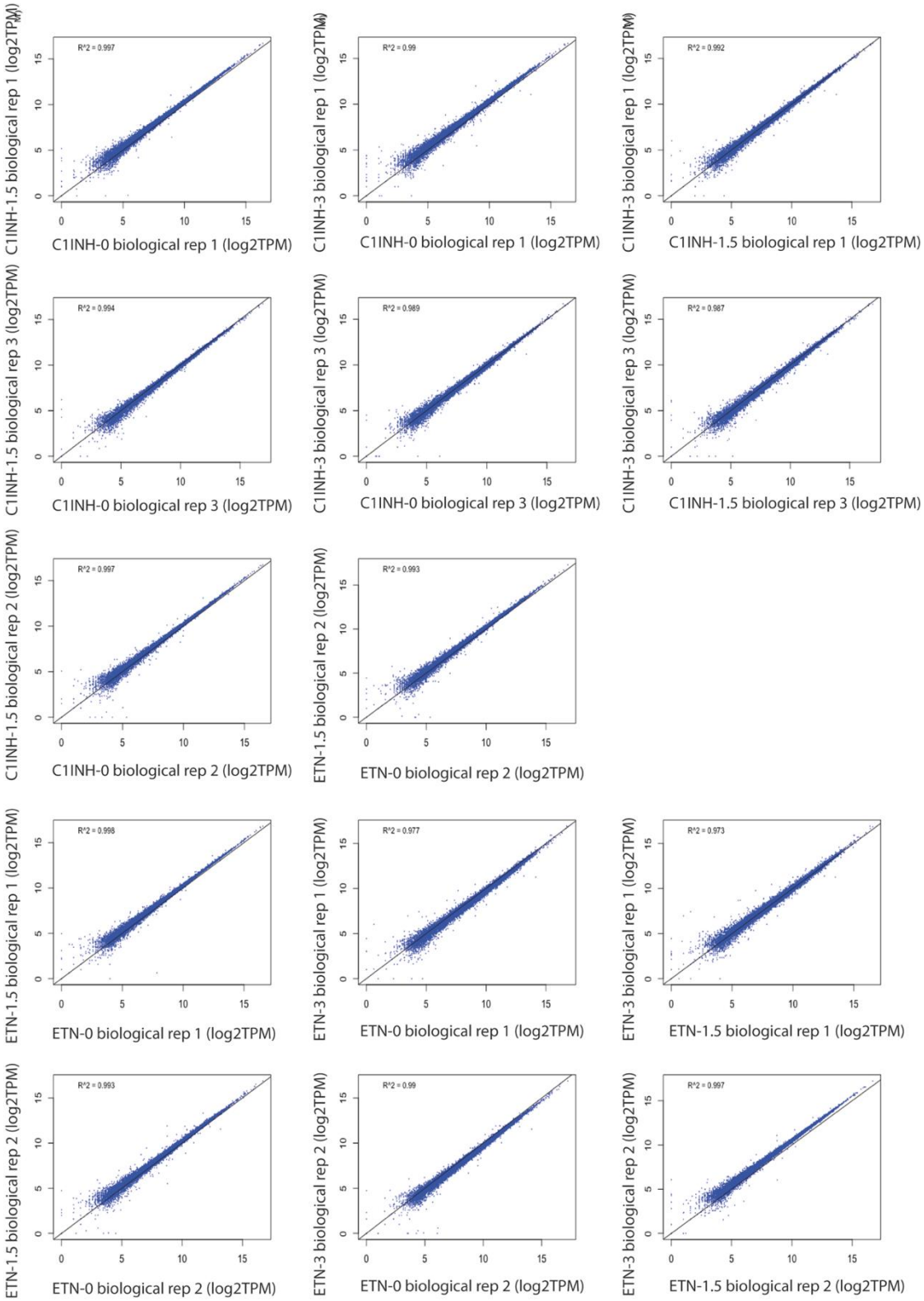
Supporting Figure S3 Comparison of expression values for pairs of replicates in dataset 1, measured in log2TPM, the spearman correlation coefficient (R^2) is denoted in the top left corner for each comparison.



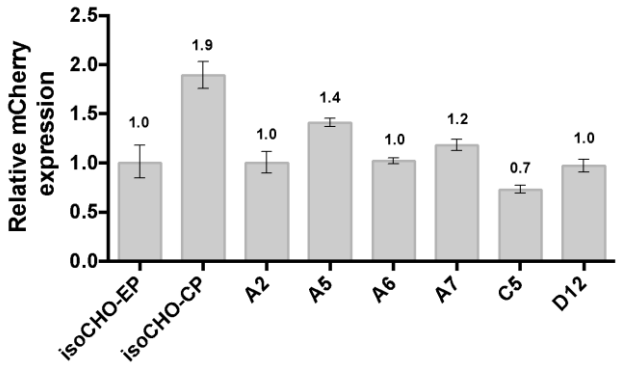
Supporting Figure S4 Overview of differential expression analysis datasets.



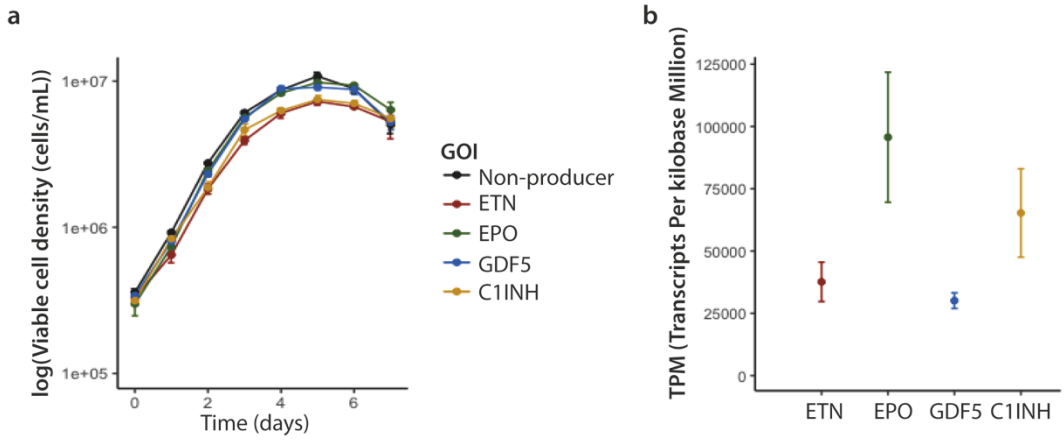
Supporting Figure S5 Comparison of expression values for pairs of replicates in dataset 2, measured in log2TPM, the spearman correlation coefficient (R^2) is denoted in the top left corner for each comparison. The number following the name of the gene of interest denotes the number of months in culture for the particular sample.



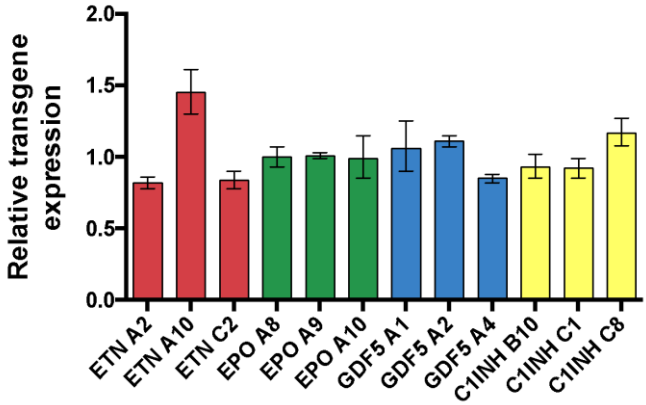
Supporting Figure S6 Comparison of expression values within biological replicates measured at different time points, measured in log2TPM, the spearman correlation coefficient (R^2) is denoted in the top left corner for each comparison. The number following the name of the gene of interest denotes the sample time point (in months).



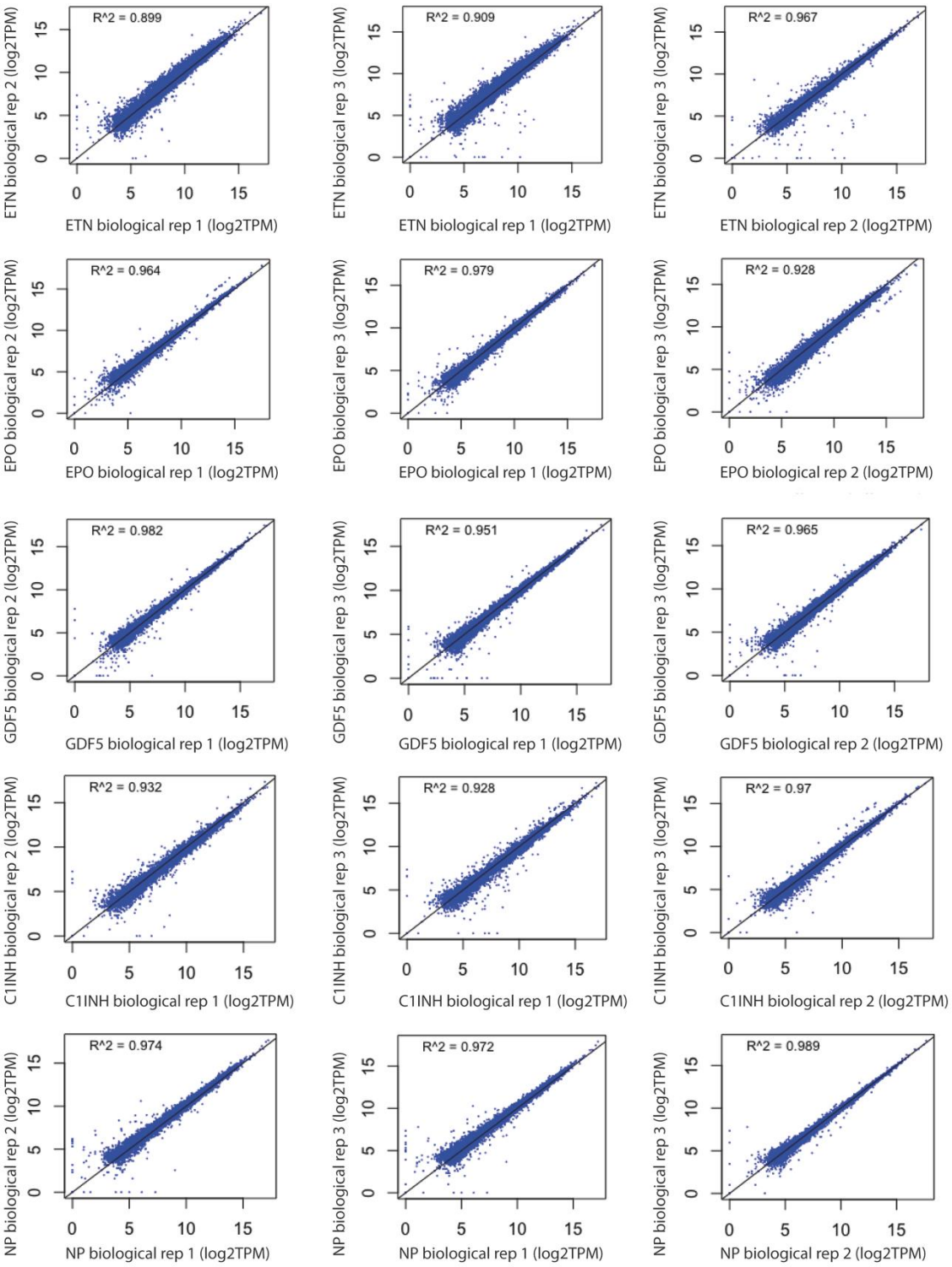
Supporting Figure S7. Relative mCherry expression levels of the panel of mCherry-CP clones, normalized to isoCHO-EP**.** Error bars represent the standard deviations of technical replicates (n≥3).



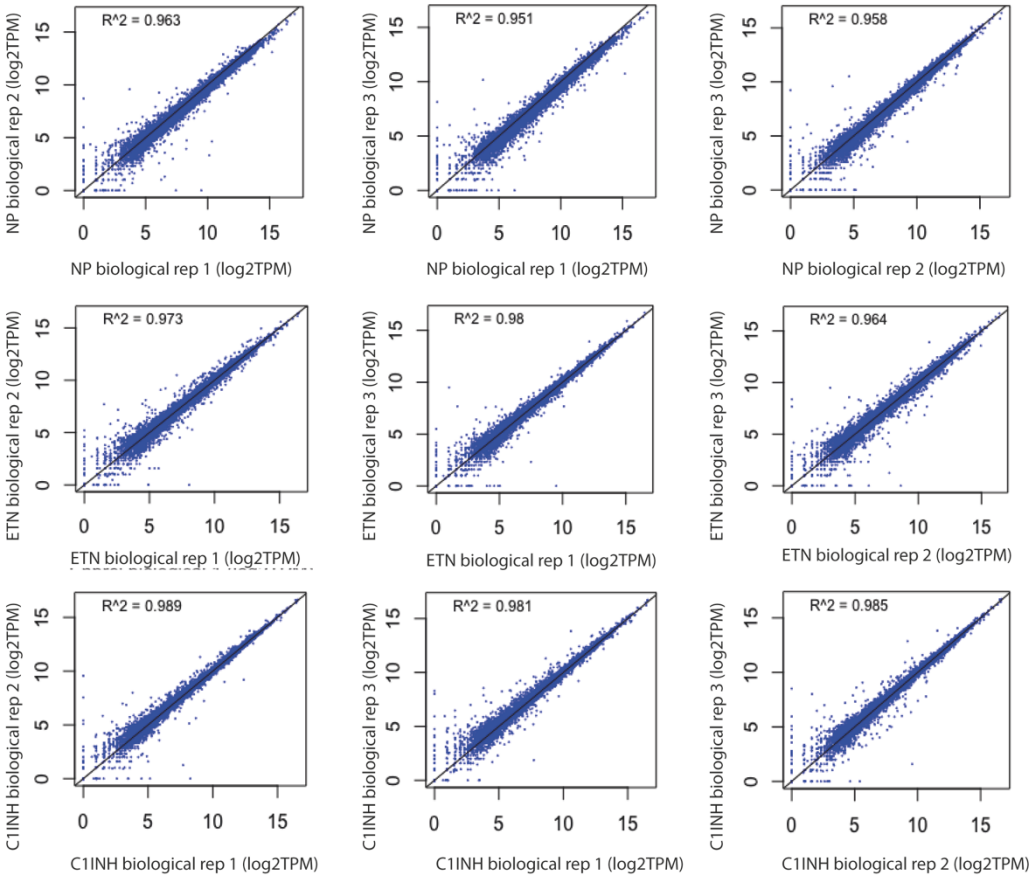
Supporting Figure S8. Phenotypes of isoCHO-CP subclones. (a) Viable cell densities of isoCHO-CP subclones expressing ETN, EPO, GDF5 or C1INH. The error bars of each line represent the standard deviations of three isogenic subclones expressing the same GOI (n=3). **(b)** Relative levels of transgene expression, as measured in transcripts per kilobase million (TPM) of ETN, EPO, GDF5 or C1INH. The error bars represent the standard deviations of three isogenic clones expressing the same GOI (n=3).



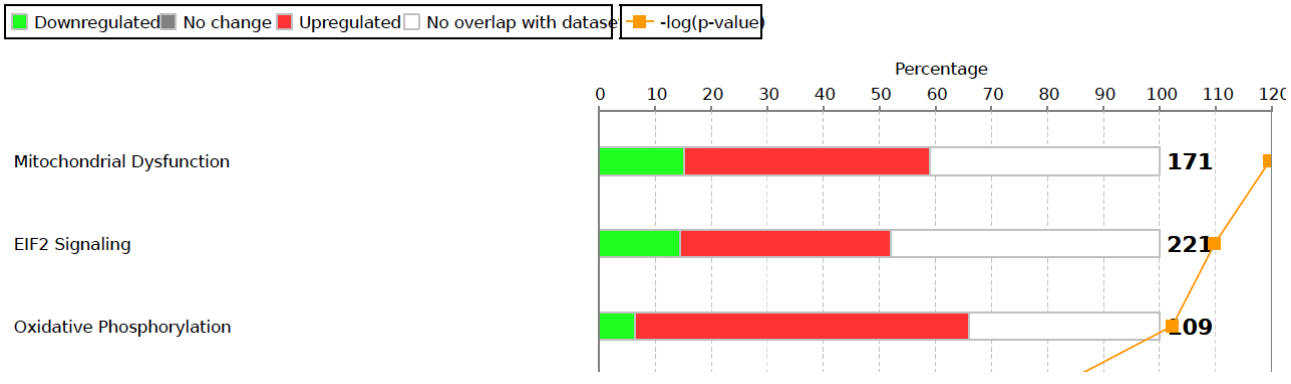
Supporting Figure S9 Relative levels of transgene expression in isoCHO-CP-derived subclones, measured by qRT-PCR, normalized to average value of corresponding isoCHO-CP subclones. The error bars represent the standard deviations of technical replicates (n=3).



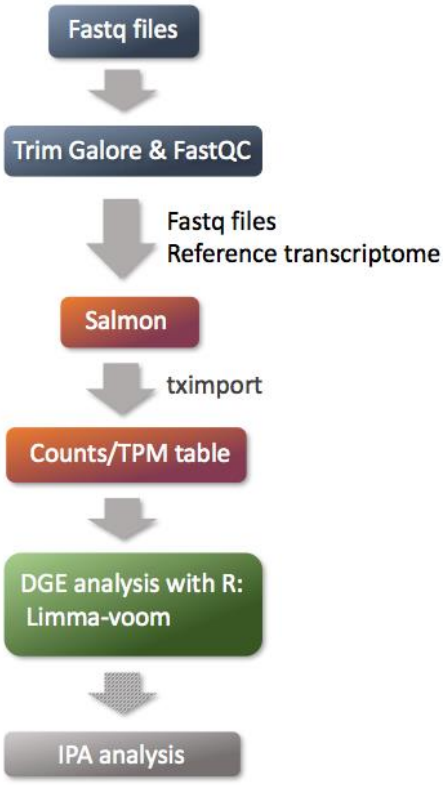
(continues on the next page)



Supporting Figure S10 Comparison of expression values for a pair of replicates in dataset 3 and 4, measured in log2TPM, the spearman correlation coefficient (R^2) is denoted in the top left corner for each comparison. Row 1-5 are isoCHO-CP clones, and row 6-8 are isoCHO-EP clones.



Supporting Figure S11 Top three pathways from the canonical pathway enrichment analysis.



Supporting Figure S12 Overview of the transcriptomic analysis pipeline.

Supporting Tables

Supporting Table S1. Fluorescence level analysis of mCherry-EP clones. Data shows mean value of two replicates.

Clone	CHO-S WT	A4	A6	A9	C12	D1	isoCHO-EP
mCherry positive population	0,0%	99,7%	100,0%	99,8%	99,8%	99,8%	100,0%
mCherry intensity	4	138	133	136	128	128	158

Supporting Table S2. Fluorescence level analysis of mCherry-CP clones. Data shows mean value of two replicates.

Clone	CHO-S WT	A2	A5	A6	A7	C5	D12	isoCHO-CP
mCherry positive population	0,2%	97,6%	92,0%	97,2%	87,4%	93,4%	94,4%	93,2%
mCherry intensity	6	139	165	160	148	188	136	205

Supporting Table S3. Nucleotide sequence of CHO codon-optimized human GDF5 gene.

ATGAGACTGCCCAAGCTGCTGACCTTCCTGCTGTGGTATCTGGCCTGGCTGGACCTGGAATTCATCTGCACCGTGCTGGGCGCTCCCGATCTGG
GACAGAGGCCTCAGGGAACCAGACCCGGACTGGCTAAGGCCGAGGCCAAAGAGAGGCCTCCCCTGGCCAGAAACGTGTTCAGACCTGGCGG
CCACTCTTACGGCGGAGGCGCCACCAATGCCAACGCCAGAGCTAAGGGCGGCACCGGACAGACAGGTGGCCTGACCCAGCCTAAGAAGGAC
GAGCCCAAGAAGCTGCCTCCTAGACCAGGCGGCCCTGAGCCTAAGCCTGGACATCCTCCACAGACCAGACAGGCCACCGCCAGAACCGTGAC
CCCTAAGGGACAGCTGCCTGGCGGAAAGGCCCCTCCTAAGGCTGGCTCTGTGCCCTCCAGCTTTCTGCTGAAGAAGGCCAGAGAGCCTGGCCC
CCCTAGAGAGCCCAAAGAGCCCTTCAGACCCCCCCCTATCACCCCCCACGAGTACATGCTGTCCCTGTACCGGACCCTGTCTGACGCCGATCGG
AAGGGCGGAAACTCCTCCGTGAAGCTGGAAGCCGGCCTGGCCAACACCATCACCAGCTTCATCGACAAGGGCCAGGACGACAGGGGTCCCGT
CGTGCGGAAGCAGAGATACGTGTTCGACATCTCCGCCCTGGAAAAGGACGGCCTGCTGGGAGCCGAGCTGCGGATCCTGAGAAAGAAGCCTT
CCGACACCGCCAAGCCTGCTGCTCCTGGCGGAGGTAGAGCTGCCCAGCTGAAGCTGTCCAGCTGCCCTTCTGGCAGACAGCCTGCCGCTCTGC
TGGATGTGCGATCTGTGCCAGGACTGGACGGCTCCGGATGGGAGGTGTTCGATATCTGGAAGCTGTTCCGCAACTTCAAGAACTCCGCCCAGC
TGTGCCTGGAACTGGAAGCTTGGGAGAGGGGCAGAGCCGTGGATCTGAGAGGCCTGGGCTTCGACAGAGCCGCTAGACAGGTGCACGAGA
AGGCCCTGTTTCTGGTGTTCGGCCGGACCAAGAAGCGGGACCTGTTCTTCAACGAGATCAAGGCCAGATCCGGCCAGGATGACAAGACCGTG
TACGAGTACCTGTTCTCCCAGCGGCGGAAGCGGAGAGCCCCTCTGGCTACAAGACAGGGCAAGCGGCCCTCCAAGAACCTGAAGGCCCGGTG
CTCTAGAAAGGCCCTGCACGTGAACTTCAAGGACATGGGCTGGGACGACTGGATCATTGCCCCCCTGGAATACGAGGCCTTCCACTGCGAGG
GCCTGTGCGAGTTCCCTCTGAGATCCCACCTGGAACCCACCAACCACGCCGTGATCCAGACCCTGATGAACTCCATGGACCCCGAGTCCACCCC
CCCTACCTGTTGTGTGCCTACCCGGCTGTCCCCCATCTCCATCCTGTTCATCGACTCCGCCAACAACGTGGTGTACAAGCAGTACGAGGACATG
GTGGTGGAATCCTGCGGCTGCCGGTGA

Supporting Table S4. Nucleotide sequences of the 750 bp homology arms of the donor plasmid

Nucleotide sequence of 5' homology arm
GATAAACTATTCTTCATTTGCCAGGGGTGAACGAAACTCCAGTGCACATTTCTCAGCTCTATAGCTGAGAGATACCTCATAAAACATAGGATAC TTAGGTAAATTTGAGTTTCAGGTGAAGATTTTTTAAGTATAATCATGTCTTGCACAATGTTTGGGATTCATTTTTATTGATTAAATCTGACAGTCC TTTTTTTTTCTTTCCTTTAAAAAACTTTTTTTTTTTTTTTAAAGACAGGGTCTTATGTATCCTAAGGTGACCTTGAACTTGCCTTGTAGATGAGTCTG GCTTTGAGCCCCTAATACCTACCTCCACCTCTCAAATCCTAGGATTACAGGTGTATACCACTCTACCCATTTGACAATTCTACTTTTGAGCCATCT TAAAGGTAAAAAGTACACTTATTTATCATATGCCTGTGTGTGGAAGTCAGAACAACTTGTGGTAGTTGGTTCTTTCCTTCATGGGGGCCCTGGG GGTTAGACTCCAATTGTCAGGCTTGCTGGTAAATGCCTTATTACCACCTTGATGTGGAGGTACAAATAATGATTCACAGGACTAGGTTCAAATA

TTAAATATAATTTATTCCGGGGAGAAGTCACTTACAGAAGAAGCAGACAGCTGCTGCAGGTTCCAGCTCTAAGGATCCGCACAGGCTACTCCC
TGTGGCCCAACCAGACGGAAGTGAGAGCCACCTGCTGGGACCCCAGACCTAAACTCACTCCCAGTCAGGTCCTCACCTGTAACCAC

Nucleotide sequence of 3' homology arm

TGGATATCTCACTACACTTTGATGGCCCAATCTTTTTTTAAAAGCATATATTTATTGAATAAAATATGAGTTCATATTGGCATCCTCATGCATGTC
TGGCAAGTACTTGGCTATTATTCACTTCCTCCTCTAAATCCCATCCATTCTTGGGTCCATAAACCTATCATCTTAGTTTCTCCTCATTTCAAAAGAA
ATTAGGAAACAGAAGTTTGTAACAGGAAATGGATTTATTTATAGTTTTGCCACTGGGGAGAAAAAATAACTCCATCCTGTTCTCTAACATGCTG
GTAGTCTTTTACAGTTTATTGAAAACTATGAGACGGGATGGAGAGGTGGTTCAGAGGTTAAGAGCAATGGCTGTTCTTCCAGAGGTCCTGAGT
TCAATTCCCAGCAACCACATAGTAGCTCACAGCTATCTGTAATGAGATCTGGTGTGCAAGCATACATGCCGGCAGAACACTGTATACAAAATAA
ATAAATCTTAAAAAAAAAAAGAAAACTATGAGACAAAGACAAGAAGGTATGTAGCTTTGCTGTGAGACAGTTCAACATGGCTTTTCTGGGGTT
GTGAGGAGGTCACTAGTTTCATATGACCTGCAGTATCTCAGTTCTCAGCAGTTCTTGTTTAGCATGGAGGGGGTCTCTCTTTACAATTTTGCCAT
GGTTCCAGCCCTTTCTCAAGCATGGAATCACTGCTATATTAGTGTCCTCTTTGTTATGGGGTAAAAACCCAGAGACCATTCAAGCAAC

Supporting Table S5. Primers for plasmids construction, junction and insert PCR.

sgRNA vector construction (for primer annealing)

Primer name	**Sequence (5' -3')**
sgRNA T9 fwd	GGAAAGGACGAAACACCGCCCACTACAGGTTCGGGCGGTTTTAGAGCTAGAAAT
sgRNA T9 rev	CTAAAACcgcccgaacctgtagtgggCGGTGTTTCGTCCTTTCCACAAGATAT

*Target sequence marked in blue

lox-mCherryOri vector construction (USER primers)

Primer name	**Sequence (5' -3')**
LoxP-kozak_mcherry_LA_fwd*	agtcggtgUATAACTTCGTATAGCATACATTATACGAAGTTATCGCCACCATGGTGAGCA
Lox2272-mcherry_O4_rev*	AGACTGTGUataacttcgtataaagtatcctatacgaagttatCTACTTGTACAGCTCGT
BGH pA_O4_fwd	ACACAGTCUCTGTGCCTTCTAGTTGCC
BGH pA_O5_rev	ACGCAAGUCCATAGAGCCCACCGCAT

*Lox sequences are marked in green.

Landing pad vectors construction (USER primers)

Primer name	**Sequence (5' -3')**
EF-1a_LB_fwd	aagcagcgUGTGAGGCTCCGGTGCCC
EF-1a_LC_rev	atgacgtcUTCACGACACCTGAAATGGAA
LinkB-mCMVenhancer-fwd	AAGCAGCGUGAGTCAATGGGAAAAACC

HTLV5'UTR-linkC-Rev	ATGACGTCUGTAGGCGCCGGTCACA
LoxP_LC_fwd	agacgtcaUATAACTTCGTATAGCATACAT
BGH pA_O2_rev	ATCGCACUccatagagcccaccgcatcc
Marker NeoR_O2_fwd	AGTGCGAUCTGTGGAATGTGTGTCAGTT
Marker NeoR_LD_rev	actcagaccUcagacatgataagatacattg
CMV_O1_fwd	ACGTCGCUGTTGACATTGATTATTGACT
BGH pA_O5_rev	ACGCAAGUccatagagcccaccgcatcc
pJ204 backbone_O5_fwd	ACTTGCGUAGTGAGTCGAATAAGGGCGACACAAA
pJ204 backbone_LA_rev	acaccgacUGAGTCGAATAAGGGCGACACCCCA
T9 5' arm_750bp_LA_fwd	agtcggtgUGATAAACTATTCTTCATTTGC
T9 5' arm_750bp_LB_rev	acgctgctUGTGGTTACAGGTGAGGACC
T9 3' arm_750bp_LD_fwd	aggtctgagUTGGATATCTCACTACACTTTG
T9 3' arm_750bp_O1_rev	AGCGACGUGTTGCTTGAATGGTCTCTGG

RMCE donor vectors construction (USER primers)

Primer name	**Sequence (5’ -3’)**
LoxP-kozak_EPO_LA_fwd	agtcggtgUATAACTTCGTATAGCATACATTATACGAAGTTATCGCCACCATGGGAGTGC
Lox2272-EPO_O5_rev	ACGCAAGUataacttcgtataaagtatcctatacgaagttatTCATCTATCGCCGGTCC
LoxP-kozak_ETN_LA_fwd	agtcggtgUATAACTTCGTATAGCATACATTATACGAAGTTATAGCACCATGGCGCCCGT
Lox2272-ETN_O5_rev	ACGCAAGUataacttcgtataaagtatcctatacgaagttatTTATCATTTACCCGGAG
LoxP-kozak_C1INH_LA_fwd	agtcggtgUATAACTTCGTATAGCATACATTATACGAAGTTATAGCACCATGGCCAGCAG
Lox2272-C1INH_O5_rev	ACGCAAGUataacttcgtataaagtatcctatacgaagttatTCAGGCTCTGGGGTCGTA
LoxP-kozak_GDF5_LA_fwd	AGTCGGTGUATAACTTCGTATAGCATACATTATACGAAGTTATGCCACCATGAGACTGCC
Lox2272-GDF5_O5_rev	ACGCAAGUataacttcgtataaagtatcctatacgaagttatTCACCGGCAGCCGCAG

Junction PCR

Primer name	**Sequence (5’ -3’)**
T9 5’ junction genomic fwd	GCATGCACAGAGAGGGACAT
T9 3’ junction genomic rev	CCCTCTGCAACTGCTAACCA
Neo(R) junction fwd	CTGGACGAAGAGCATCAGGG

EF1-a junction rev	ATCCTGGCCCGCATTTACAA
mCMV enh junction rev	TGGCTTACCTCCCATTGACC

Insert PCR

Primer name	**Sequence (5' -3')**
EF1a-mcherry junction fwd	CCTCAGACAGTGGTTCAAAGT
BGH pA rev	AGATGGCTGGCAACTAGAAG
min-hEF1a fwd	GGAGAACCGTATATAAGTGCAGTAG

Supporting Table S6. qRT-PCR primers and probes.

qRT-PCR primers (SYBR Green assay)

Gene	**Fwd primer**	**Rev primer**
mCherry	AGGACGGCGAGTTCATCTA	CCCATGGTCTTCTTCTGCATTA
GAPDH	TTGTCATCAACGGGAAGG	GTGAAGACGCCAGTAGATT

TaqMan assays

Gene	**Fwd primer**	**Rev primer**	**Probe**	**Dye**
Fkbp1a	CTCTCGGGACAGAAACAAGC	GACCTACACTCATCTGGGCTAC	ATGCTAGGCAAGCAGGAGGTGATC	VIC-MGB
Gnb1	CCATATGTTTCTTTCCCAATGGC	AAGTCGTCGTACCCAGCAAG	ACTGGTTCAGACGATGCTACGTGC	ABY-MGB
ETN	CAGCCGGAGAACAACTACAA	CATCACGGAGCATGAGAAGA	TACAGCAAGCTCACCGTGGACAAG	FAM-MGB
EPO	CTGGAAAGATACCTGCTGGAAG	AGGCGTAGAAGTTCACTTTGG	CCAAAGAGGCCGAGAACATCACCA	FAM-MGB
GDF5	GTGATCCAGACCCTGATGAAC	GTCGATGAACAGGATGGAGATG	TACCTGTTGTGTGCCTACCCGG	FAM-MGB
C1INH	GGATGGAGCCCTTTCACTTTA	GGATGACCAGGCTCAGATTATG	TCATCGACCAGACCCTGAAGGCTA	FAM-MGB

Supporting Table S7. RNA-seq datasets.

Dataset	**Samples**	**Sequencing kit**
1	**isoCHO-EP**: ETN (n=3), EPO (n=3), GDF5 (n=3), C1INH (n=3)	mid-output

2	**isoCHO-EP**: ETN-0 months (n=3), ETN-1.5 months (n=3), ETN-3 months (n=2), C1INH-0 months (n=3), C1INH-1.5 months (n=3), C1INH-3 months (n=2)	high-output
3	**isoCHO-CP**: ETN (n=3), EPO (n=3), GDF5 (n=3), C1INH (n=3), non-producer (n=3)	high-output
4	**isoCHO-EP**: ETN (n=3), C1INH (n=3), non-producer (n=3) **isoCHO-CP**: ETN (n=3), C1INH (n=3), non-producer (n=3)	high-output