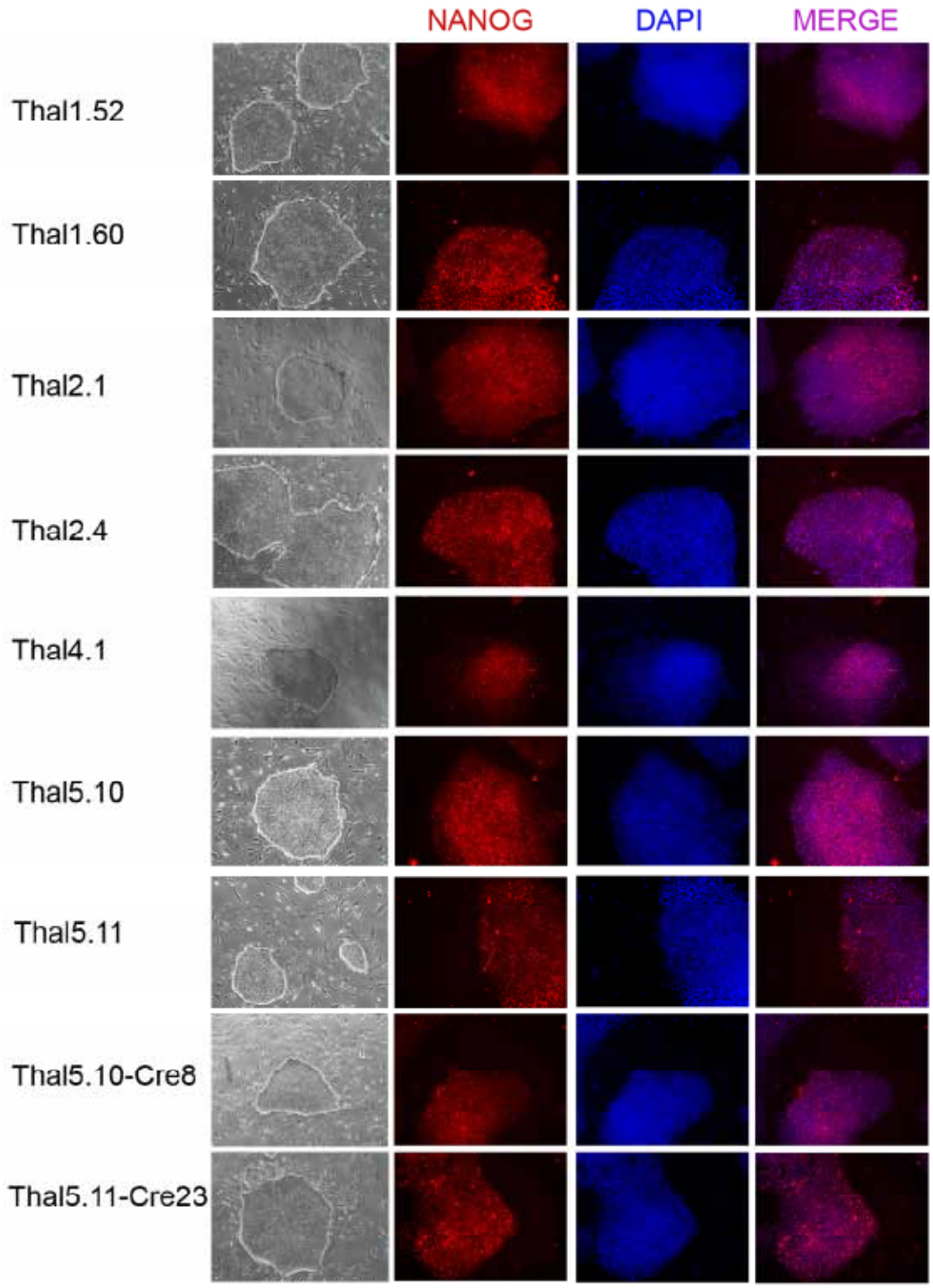
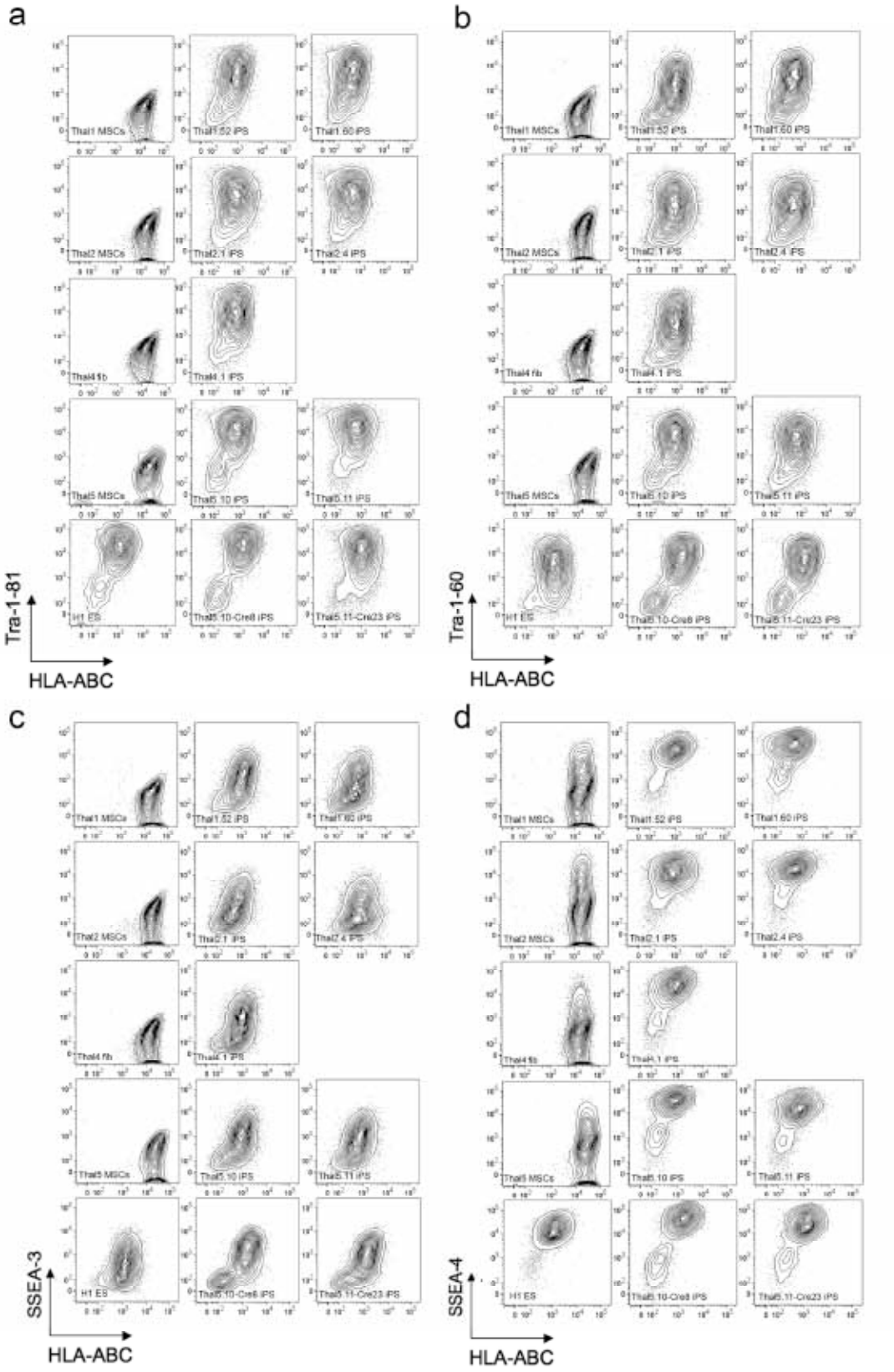


Supplementary Figure 1. Immunofluorescence for NANOG expression. NANOG expression in the iPS cell lines thal1.52, thal1.60, thal2.1, thal2.4, thal4.1, thal5.10, thal5.11, thal5.10-Cre8 and thal5.11-Cre23.

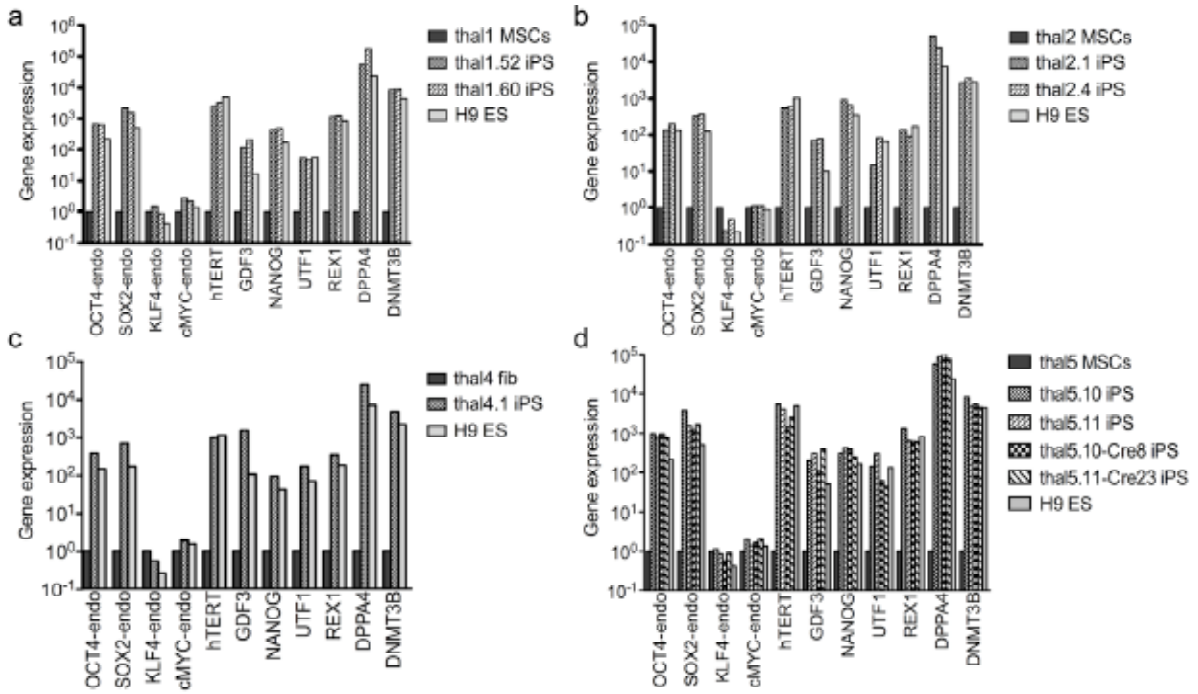
Nature Biotechnology: doi: 10.1038/nbt.1717

Supplementary Figure 2

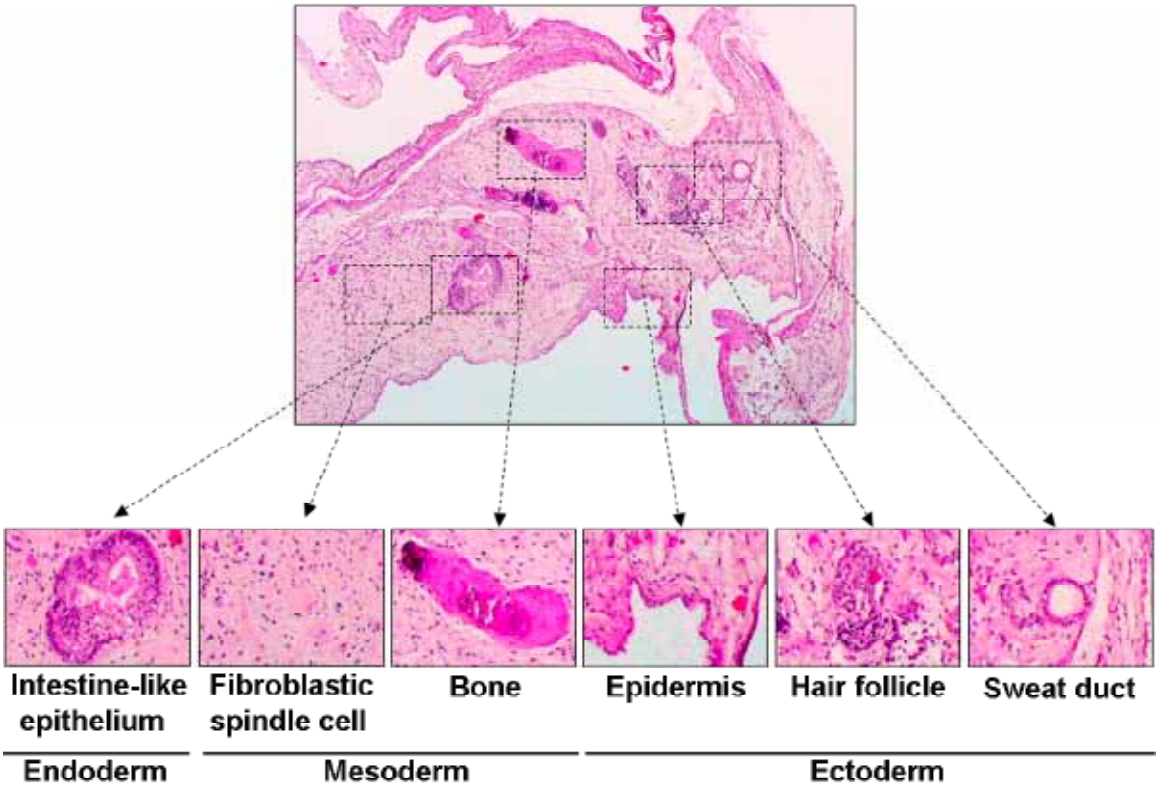


Nature Biotechnology: doi: 10.1038/nbt.1717

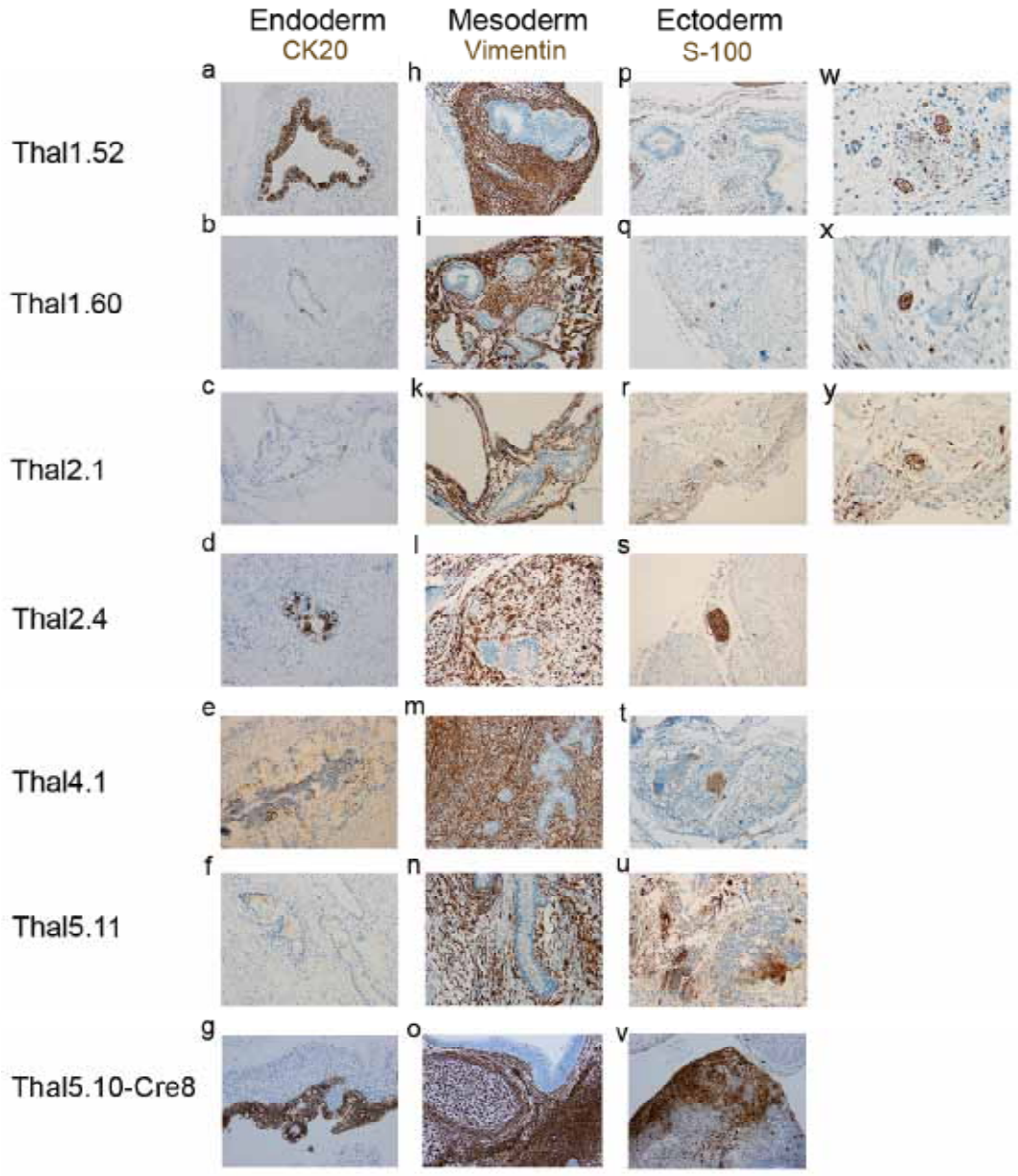
Supplementary Figure 2. Expression of surface pluripotency markers in thal-iPS cell lines. Expression of pluripotent cell markers Tra-1-81 (**a**), Tra-1-60 (**b**), SSEA-3 (**c**) and SSEA-4 (**d**) in the thal-iPS cell lines thal1.52, thal1.60, thal2.1, thal2.4, thal4.1, thal5.10, thal5.11, thal5.10-Cre8, thal5.11-Cre23 and in the hES cell line H1, as indicated. The corresponding MSCs (thal1, thal2, thal5) or skin fibroblasts (thal4 fib) are shown in the left panels as controls. Additionally, iPS cells are shown to characteristically down-regulate human leukocyte antigen (HLA)-ABC, similarly to hES cells, compared to the MSCs or fibroblasts they were derived from[15]. The pluripotency marker-negative/HLA-ABC-negative population corresponds to MEFs.



Supplementary Figure 3. Expression of pluripotency genes. Expression of pluripotency-associated genes (listed below the X axis) by qRT-PCR in the iPS cell lines thal1.52, thal1.60 (a), thal2.1, thal2.4 (b), thal4.1 (c), thal5.10, thal5.11, thal5.10-Cre8, thal5.11-Cre23 (d) and in the hES cell line H9 relative to expression in the corresponding MSCs (thal1, thal2, thal5) or skin fibroblasts (thal4 fib) they were derived from. The suffix “endo” refers to the endogenous, as opposed to the vector-encoded, transcript.

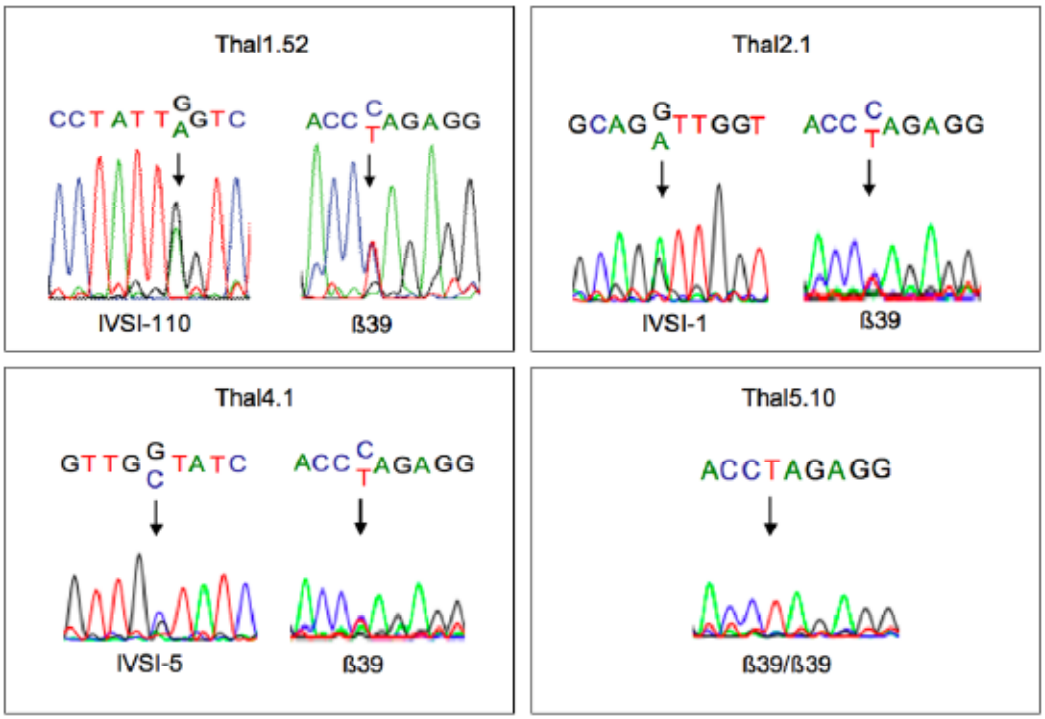


Supplementary Figure 4. Histological analysis of a representative teratoma. Hematoxylin and eosin staining of histological sections of a teratoma derived from line thal2.1 comprising tissues of all three germ layers.



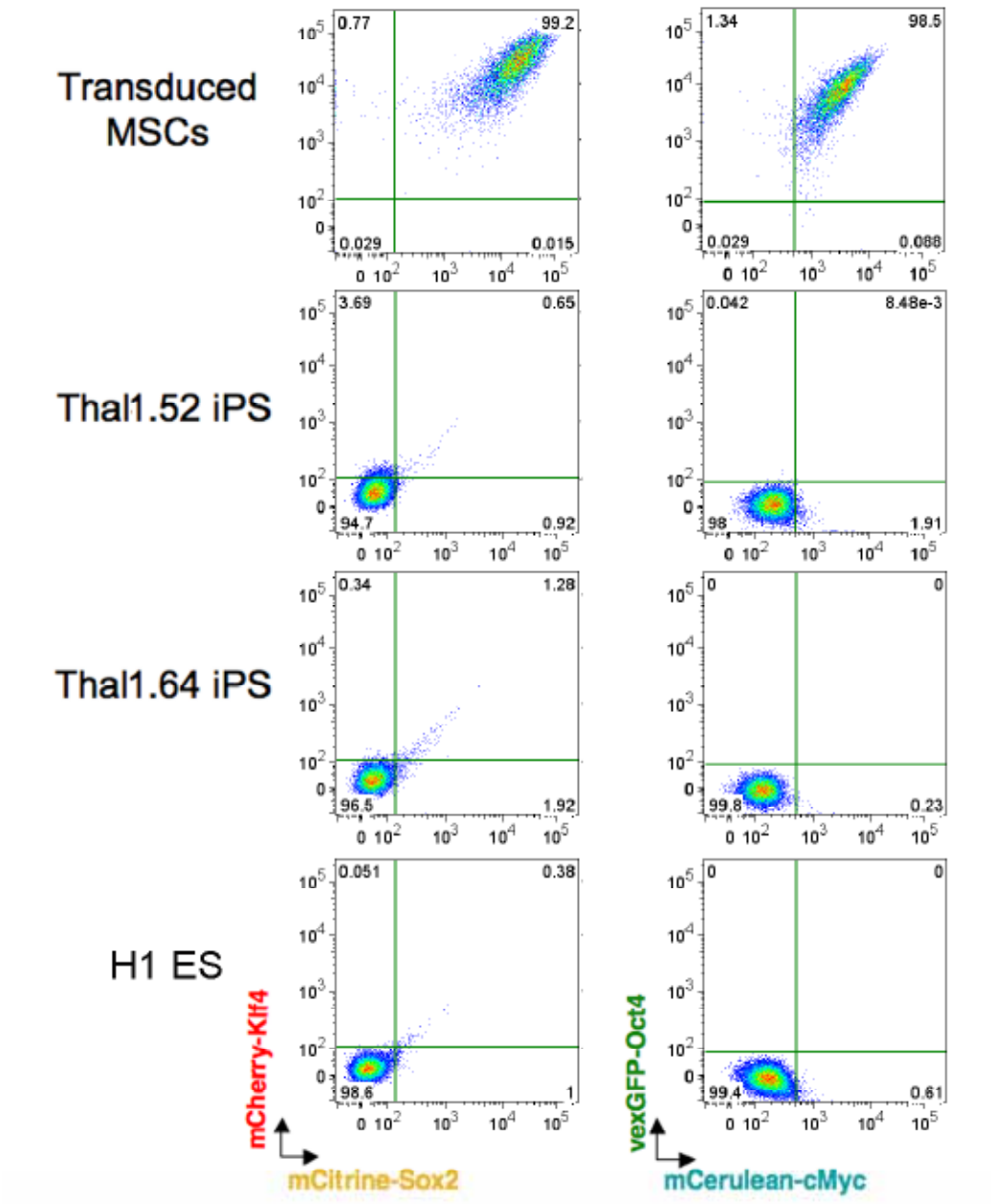
Supplementary Figure 5. Immunohistochemical analysis of teratomas. Immunohistochemical analysis of teratomas derived from iPS cell lines thal1.52, thal1.60, thal2.1, thal2.4, thal4.1, thal5.11 and thal5.10-Cre8, as indicated. **a-g**, cytokeratin (CK) 20-positive intestine-like epithelium (endoderm); **h-o**: vimentin-positive fibroblastic spindle cells (mesoderm); **p-v**: S100-positive peripheral nerve (**p-t**) or immature neural tissue (**u, v**) (ectoderm); **w-y**: higher magnification images of (p-r), respectively.

Nature Biotechnology: doi: 10.1038/nbt.1717

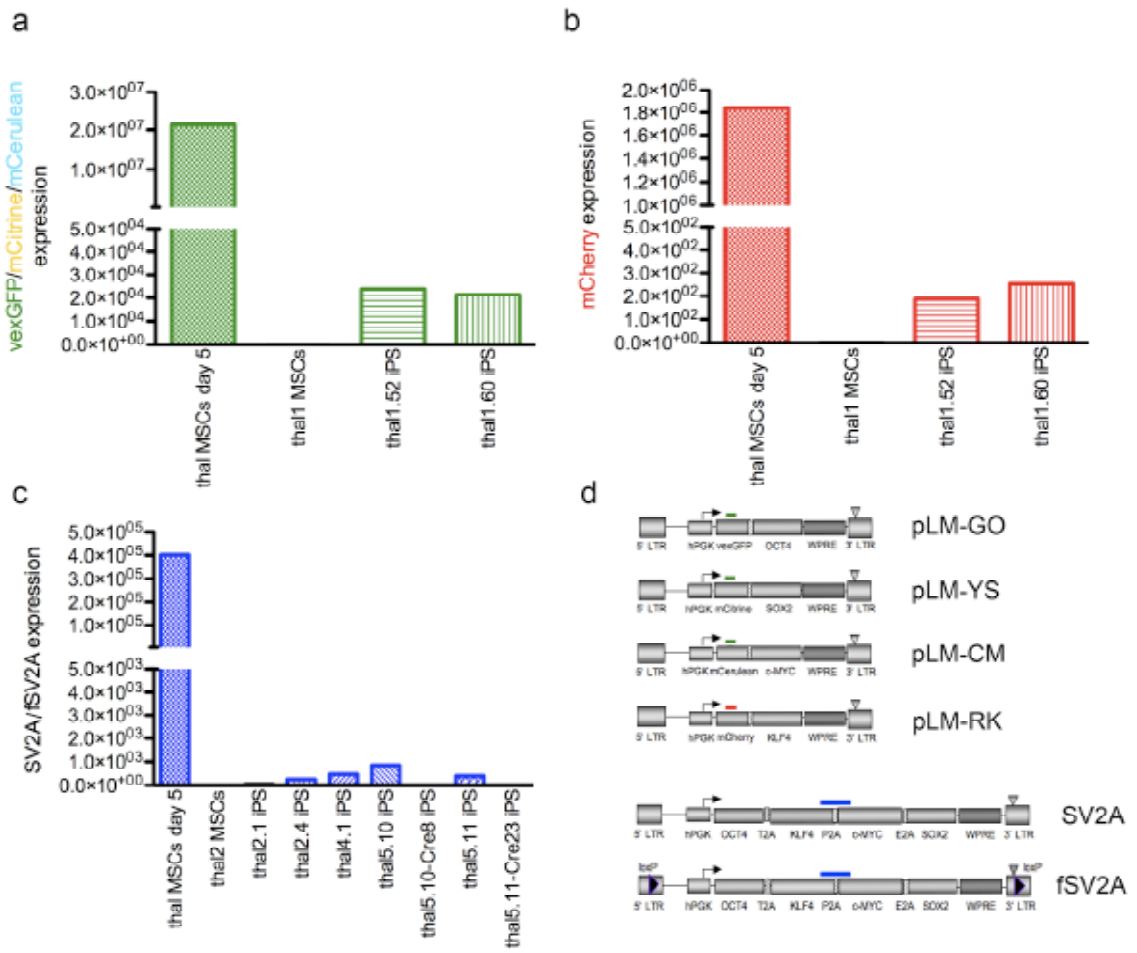


Supplementary Figure 6. Genotyping of thal-iPS cell lines. Representative genotyping results of thal-iPS cell lines. Lines thal1.52, thal2.1 and thal4.1 are compound heterozygous for the $ß^{+}$ mutations intervening sequence (IVS)I-110 (G→A), IVSI-1 (G→A), (IVS)I-5 (G→C), respectively, and the $ß^{0}$ mutation in codon 39 (C→T). Line thal5.10 is homozygous for the codon 39 mutation.

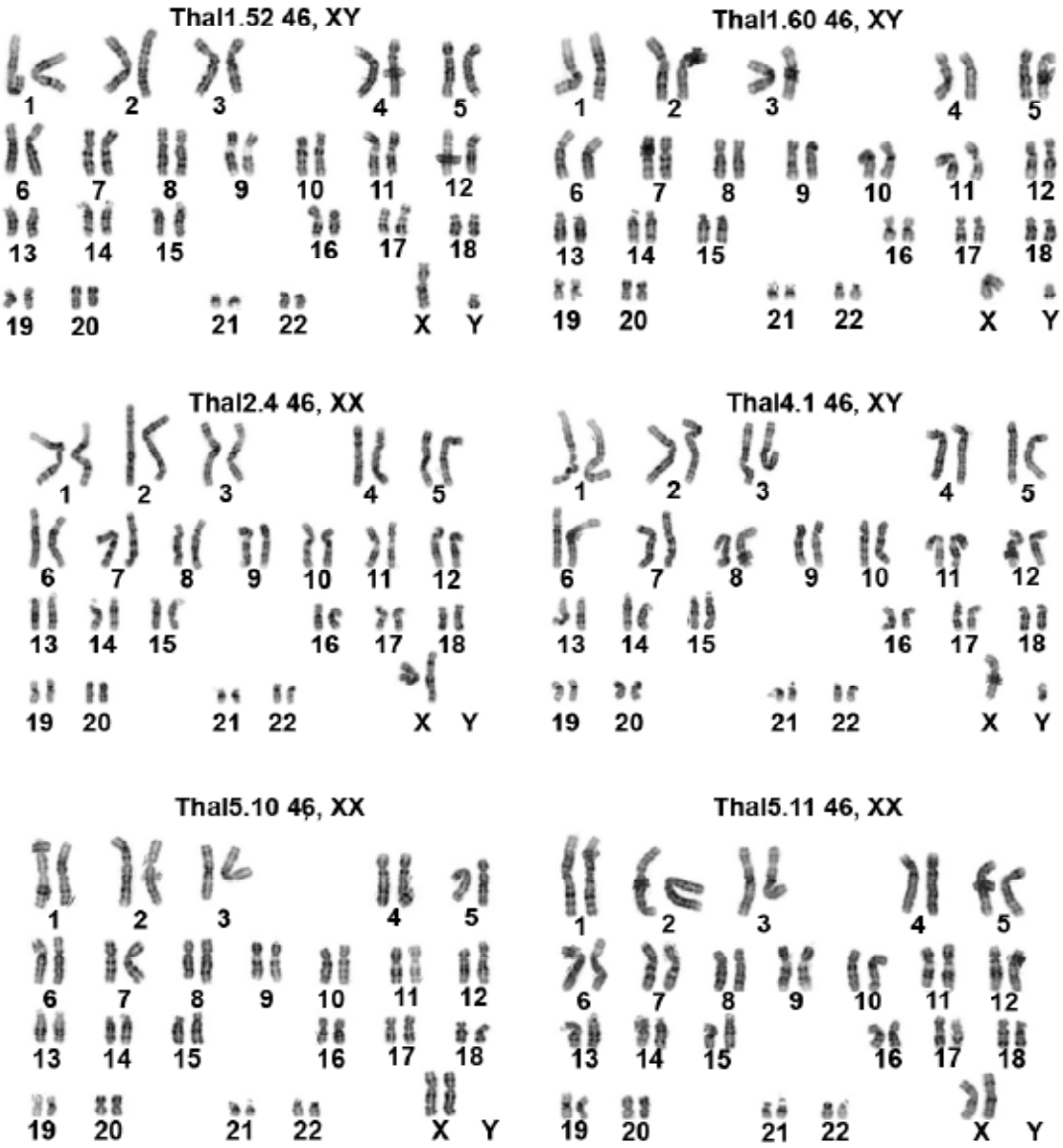
Nature Biotechnology: doi: 10.1038/nbt.1717



Supplementary Figure 7. Silencing of the reprogramming vectors. Silencing of the four vector-encoded factors in thal-iPS cell lines thal1.52 and thal1.64 derived from subject 1 with four bicistronic lentiviral vectors co-expressing OCT4, KLF4, c-MYC and SOX2 with a distinct fluorescent protein[15]. Upper panels: transgene expression in MSCs 48 hours after transduction with the four vectors. H1 ES: hES cell line H1.



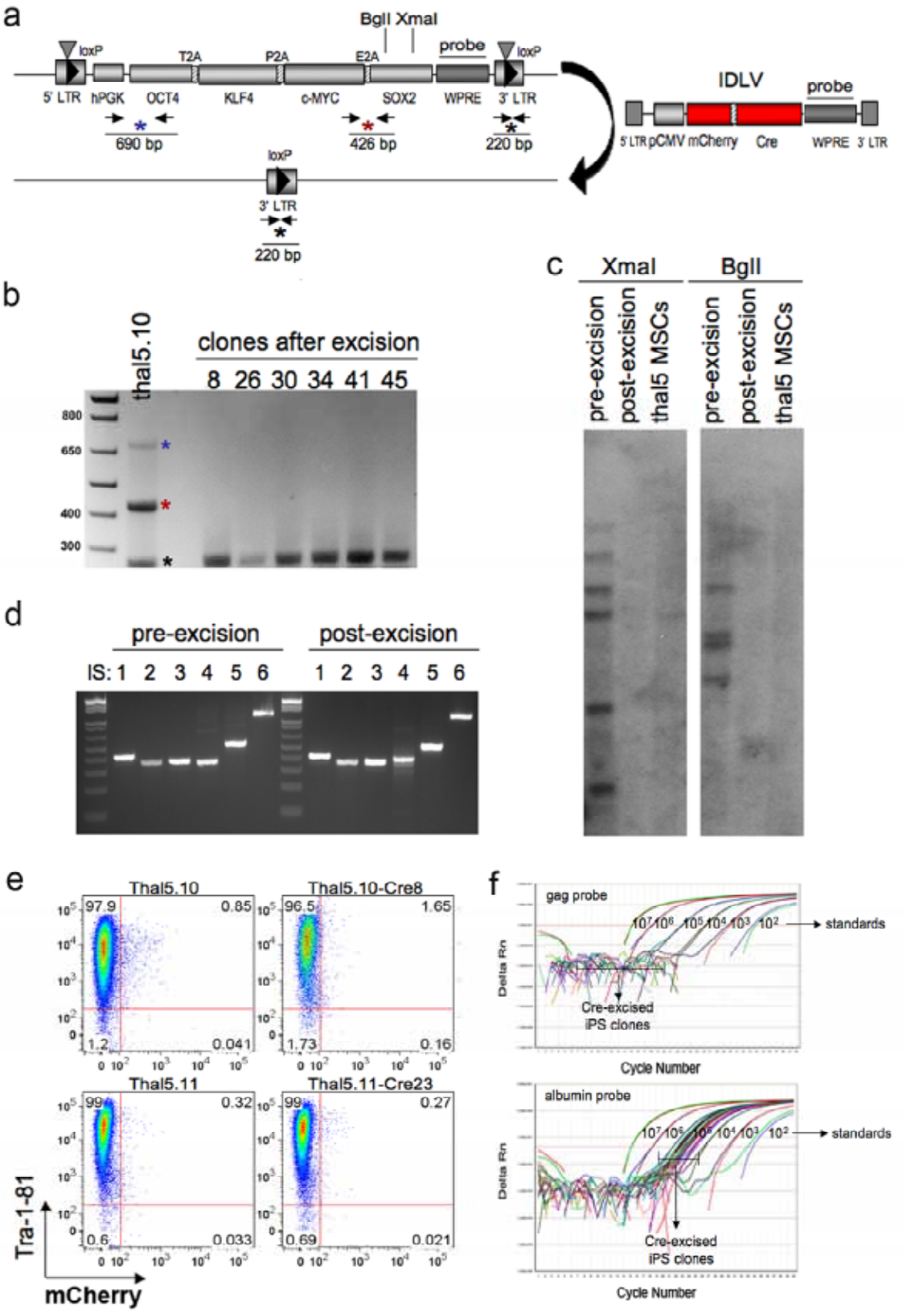
Supplementary Figure 8. Silencing of reprogramming vectors in thal-iPS cell lines by qRT-PCR. (**a, b**) Expression of the vector-encoded transcripts vexGFP-P2A-OCT4, mCitrine-P2A-SOX2, mCerulean-P2A-cMYC (a) and mCherry-P2A-KLF4 (b) in thal-iPS cell lines thal1.52 and thal1.60 derived from subject 1 with four bicistronic lentiviral vectors co-expressing OCT4, KLF4, c-MYC and SOX2 with a distinct fluorescent protein[15]. (**c**) Expression of the polycistronic OCT4-T2A-KLF4-P2A-cMYC-E2A-SOX2 transcript in thal-iPS cell lines thal2.1, thal2.4, thal4.1, thal5.10 and thal5.11. (**d**) Schematic representation of the vectors used for reprogramming with the positions of the probes used for detection of the vector-encoded transcript by qRT-PCR depicted. The primers and probe used in (a) detect all 3 GFP-derivative (vexGFP, mCitrine and mCerulean) transcripts. Thal MSCs day 5: MSCs 5 days after transduction with the 4 bicistronic vectors or the polycistronic SV2A vector. Thal1 MSCs, thal2 MSCs: untransduced MSCs from patients 1 and 2, respectively.



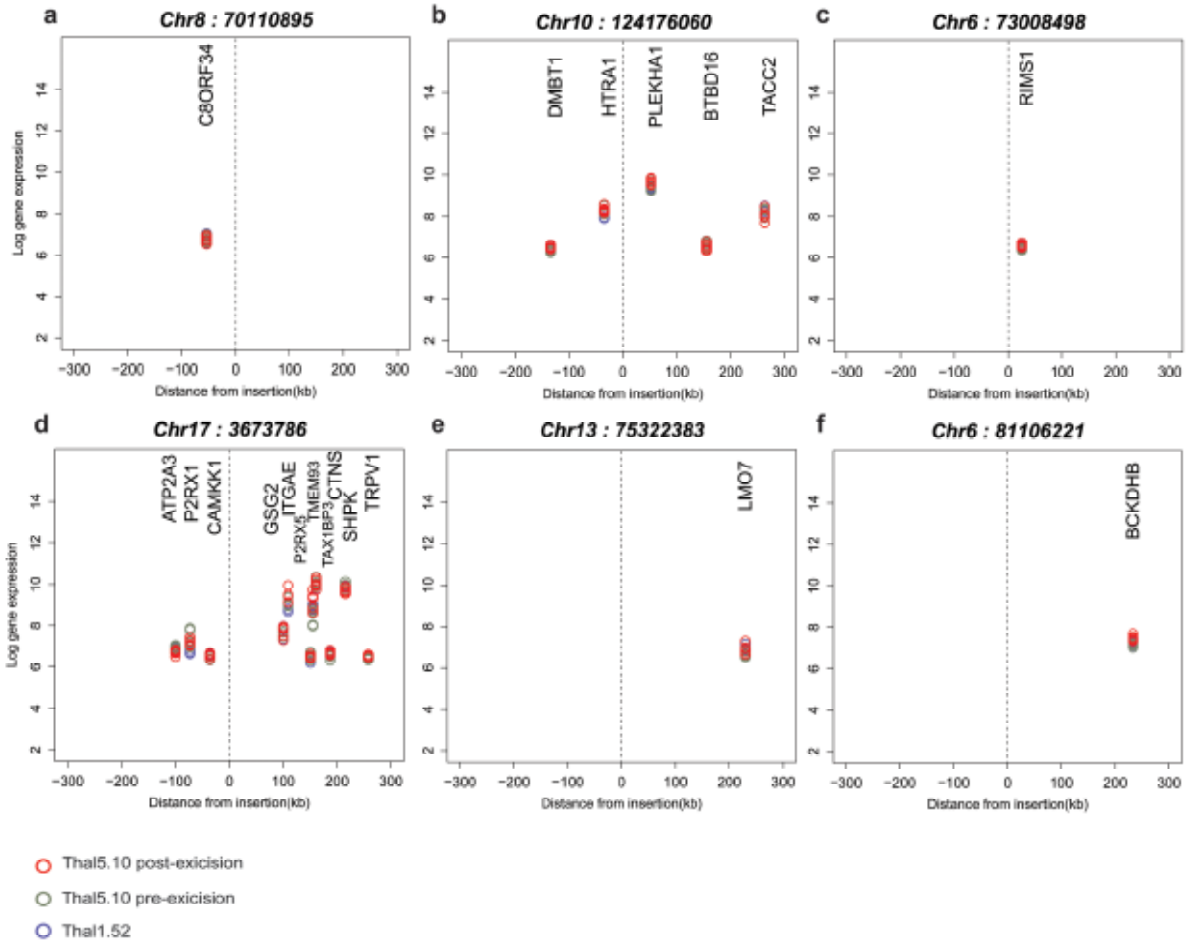
Supplementary Figure 9. Karyotype analysis of thal-iPS cell lines.

Nature Biotechnology: doi: 10.1038/nbt.1717

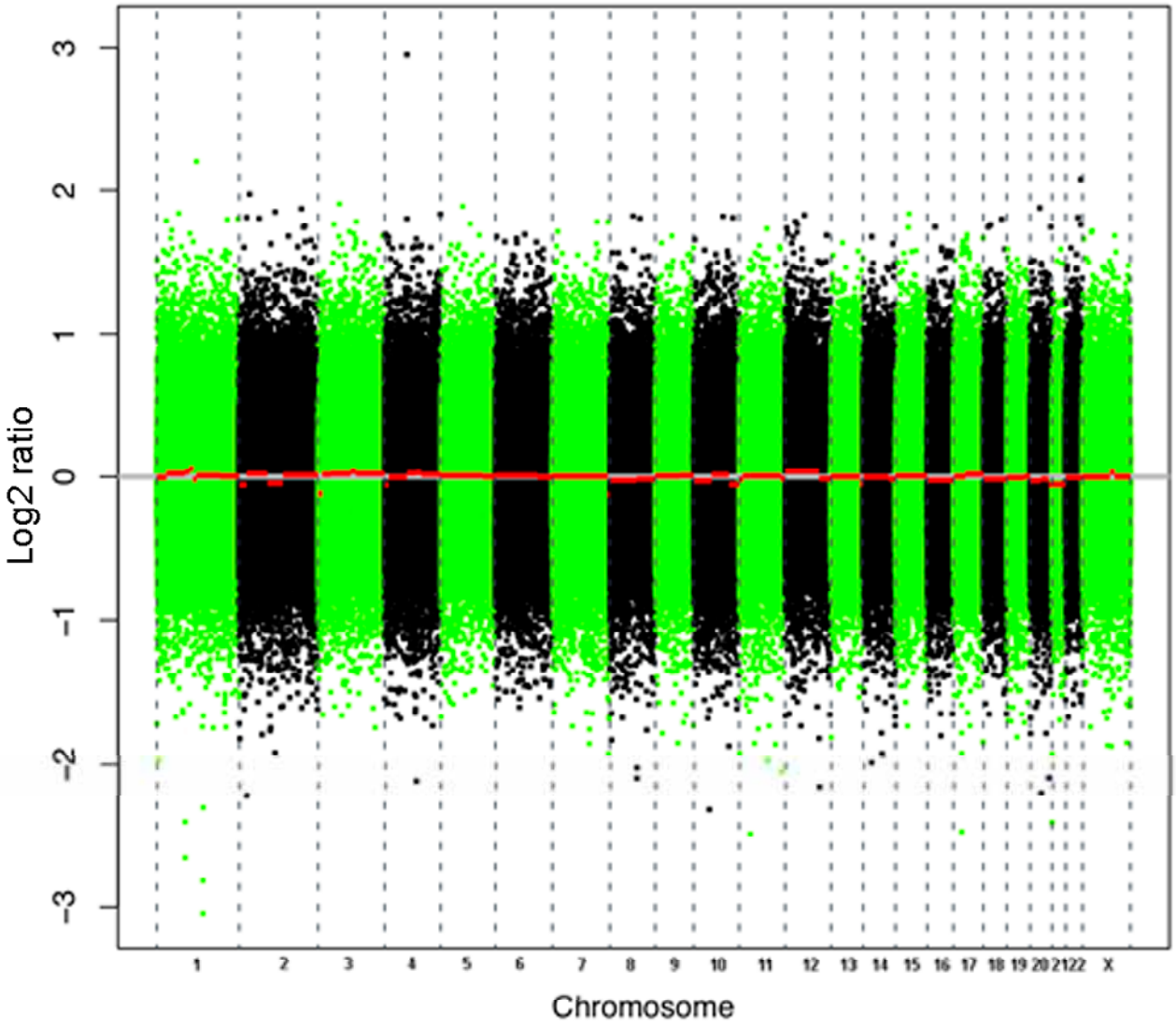
Supplementary Figure 10



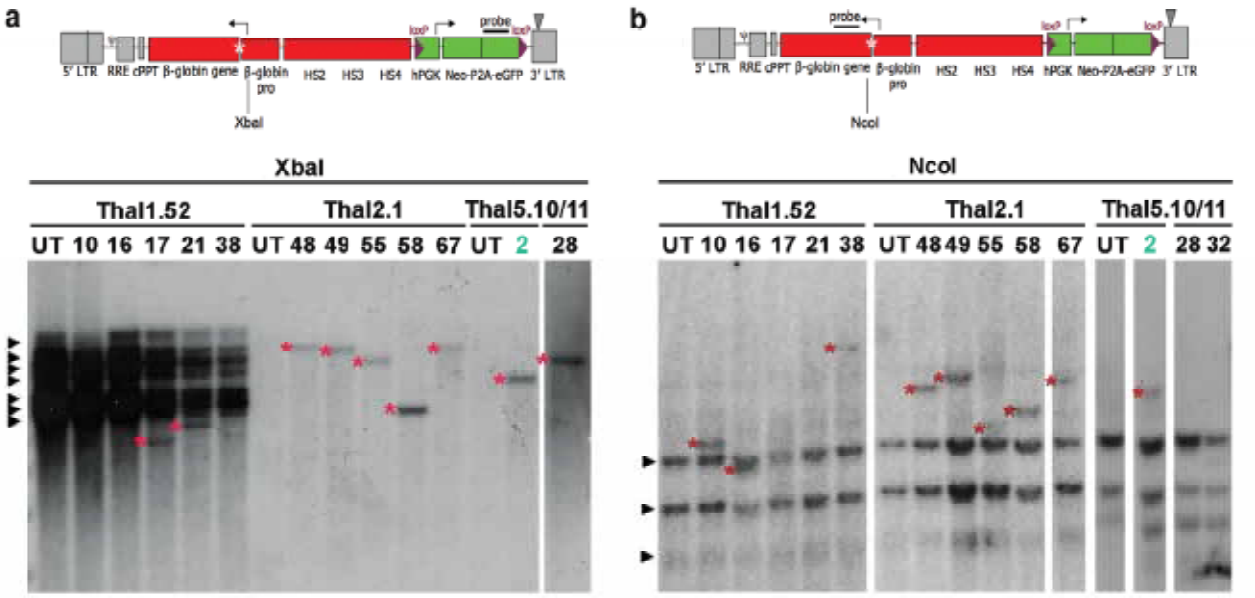
Supplementary Figure 10. Generation of transgene-free thal-iPS cells. (**a**) Upper panel: Schematic representation of the integrated fSV2A lentiviral vector co-expressing OCT4, KLF4, c-MYC and SOX2 on a single polycistronic transcript flanked by two loxP sites. Transduction of established thal-iPS cell lines thal5.10 and thal5.11 with an integrase-deficient lentiviral vector (IDLV) encoding bicistronically mCherry and Cre recombinase linked by a P2A peptide (right panel) results in excision of the vector leaving behind a viral long terminal repeat (LTR) with a deleted U3 region (lower panel). LTR: long terminal repeat; hPGK: human phosphoglycerate kinase promoter; WPRE: Woodchuck hepatitis virus posttranscriptional regulatory element. (**b**) Multiplex PCR for the integrated fSV2A provirus in 6 excised out of 47 screened thal-iPS cell clones (thal5.10-Cre8, -Cre26, -Cre30, -Cre34, -Cre41 and -Cre45) derived from line thal5.10, demonstrating excision of the vector. Asterisks show PCR products generated with primers shown in (a). (**c**) Southern Blot analysis of lines thal5.10 (pre-excision) and thal5.10-Cre8 (post-excision) after digestion with *XmaI* (left panel) or *BglI* (right panel) using the WPRE as probe (depicted in a). (**d**) PCR with a forward LTR-specific primer and reverse primers unique for each of the 6 integration sites (IS) in lines thal5.10 (pre-excision) and thal5.10-Cre8 (post-excision). The genomic coordinates of the 6 integration sites are shown in Supplementary Table 3. (**e**) Analysis of Tra-1-81 and mCherry expression in vector-excised lines thal5-10-Cre8 and thal5.11-Cre23, as well as in the parental lines thal5.10 and thal5.11, showing no residual mCherry-Cre expression. (**f**) Delta Rn vs cycle plots of multiplex qPCR with primers and probes specific for the gag region (upper panel) and the human albumin gene (lower panel) in the excised clones derived from thal5.10 shown in (b) (thal5.10-Cre8, -Cre26, -Cre30, -Cre34, -Cre41 and -Cre45). Standards corresponding to 10^2 up to 10^7 templates of both the gag and albumin amplicons run simultaneously are also shown, demonstrating the sensitivity of detection and excluding the possibility of residual integrated proviruses. All samples were run in duplicate.



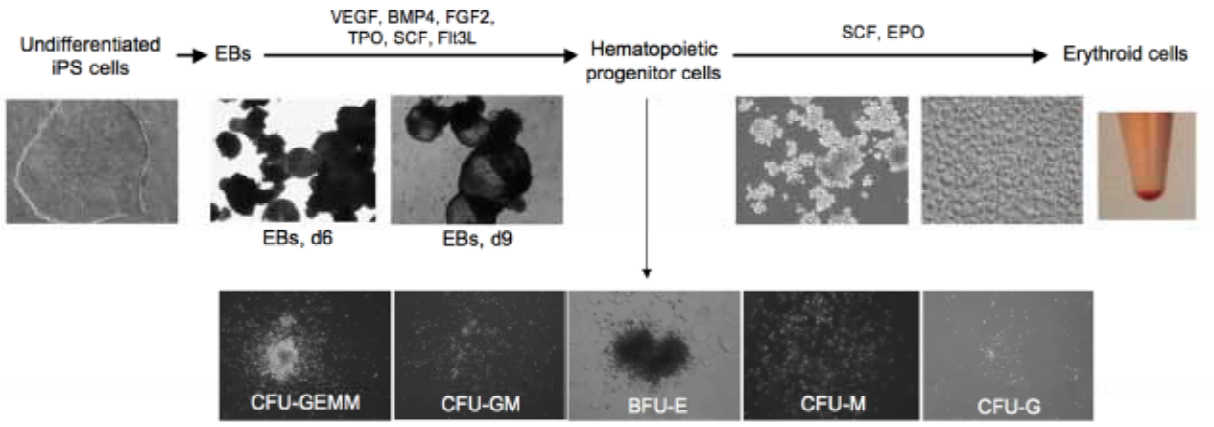
Supplementary Figure 11. Expression of genes within 300 kb of the 6 fSV2A vector integration sites in thal-5.10 line before and after vector excision. Expression of genes within 300 kb on either side of the fSV2A vector insertions (a-f: insertion sites 1-6, respectively) was assessed by microarray analysis in the thal-iPS cell line thal5.10 before (grey circles) and after (line thal5.10-Cre8, red circles) excision of all 6 copies of the fSV2A vector, relative to the expression in another thal-iPS cell line (thal1.52, blue circles). No significantly differentially expressed genes ($p < 0.05$) were observed in any of the vector insertions. Biological triplicates for all 3 samples were analyzed.



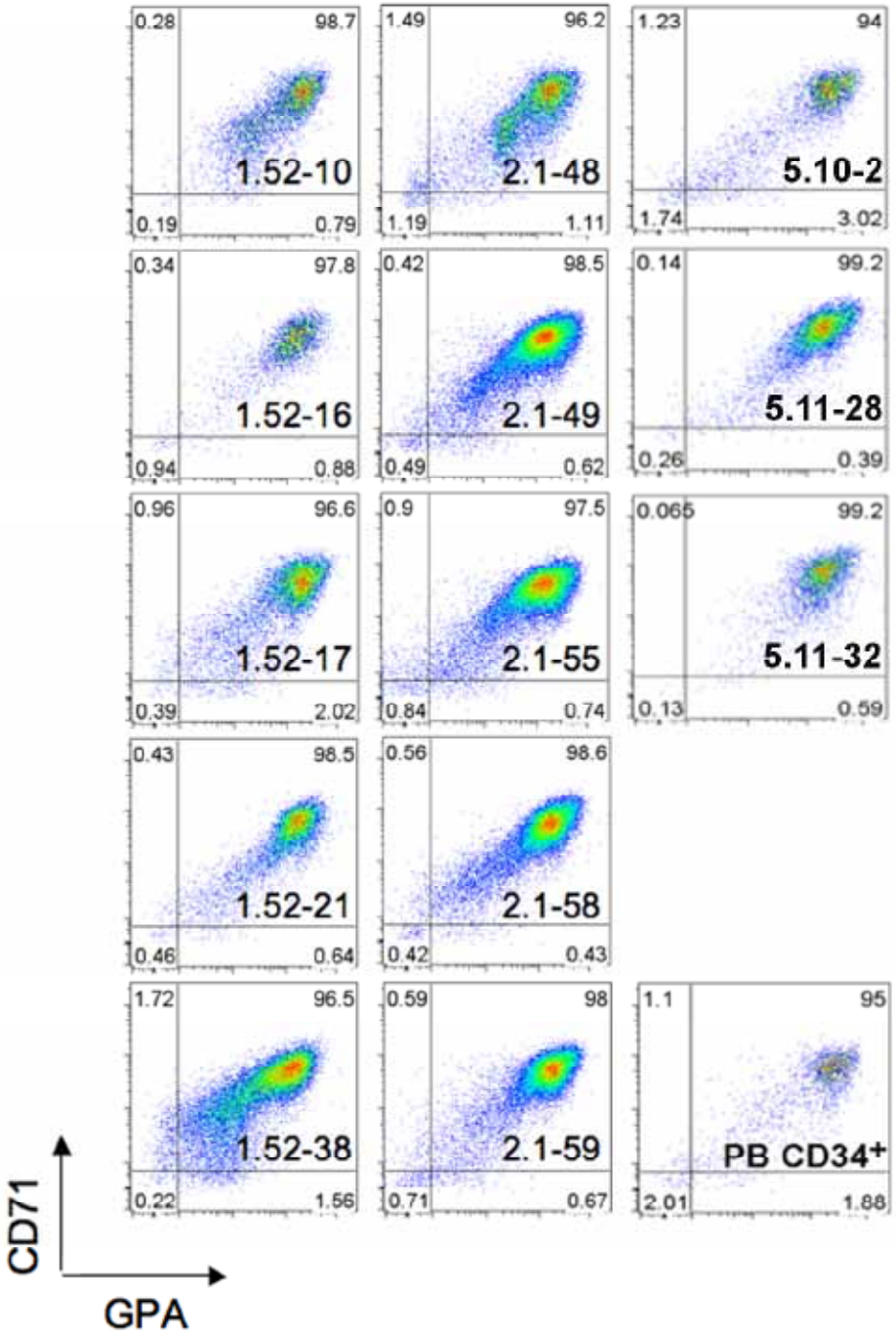
Supplementary Figure 12. Comparative Genomic Hybridization (CGH) array analysis of iPS cell line thal5.10-Cre8 and the parental thal5 MSCs. Genome wide copy number profiles are shown for all the chromosomes (X axis). The Y-axis represents the raw probe intensity from the CGH array. The red line represents the segment means from Circular Binary Segmentation. None of the segments are > 0.3 or < −0.3, indicating that there are no significant copy number aberrations.



Supplementary Figure 13. Southern Blot analysis of single copy iPS cell clones. Southern blot analysis to ascertain single integrations of the TNS9.3/fNG vector and clonality. Genomic DNA was digested with *Xba*I (a) or *Nco*I (b). The probe used in (a) is eGFP (shown in the upper panel). The probe in (b) spans exons 1 and 2 of the ß-globin gene (shown in the upper panel). The parental thal-iPS cell lines thal1.52, thal2.1, thal5.10 and thal5.11 and the clone number are depicted above the lanes. UT: untransduced. Arrowheads in (a) indicate bands corresponding to the reprogramming vectors pLM-GO, pLM-YS and pLM-CM (diagrammatically depicted in Supplementary Fig. 6d) present in the thal1.52 line and the clones derived from it. Arrowheads in (b) indicate endogenous bands (corresponding to the endogenous ß-globin locus). Asterisks depict unique vector integration bands.

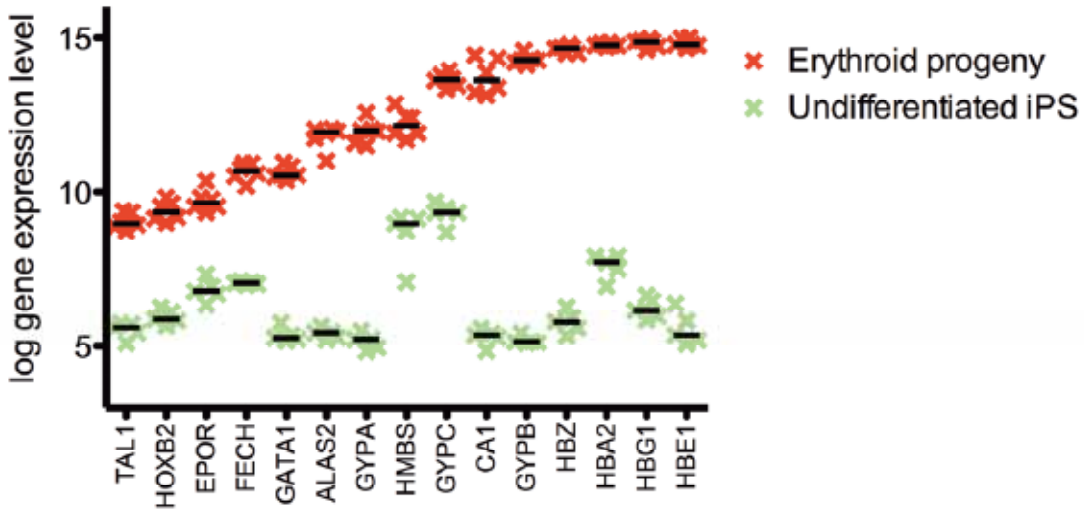


Supplementary Figure 14. Hematopoietic/erythroid differentiation of thal-iPS cells. Upper panels from left to right: undifferentiated iPS; embryoid bodies (EBs) at day 6 and 9 of culture, as indicated; erythroid cells in suspension culture at day 21, with macroscopic signs of hemoglobinization. Lower panels: Methylcellulose colonies derived from iPS cell line thal5.10 at day 9 of EB differentiation. CFU-GEMM: colony-forming unit-granulocyte, erythrocyte, monocyte, megakaryocyte; CFU-GM: colony-forming unit-granulocyte, monocyte; BFU-E: burst-forming unit-erythrocyte; CFU-M: colony-forming unit-monocyte; CFU-G: colony-forming unit-granulocyte.

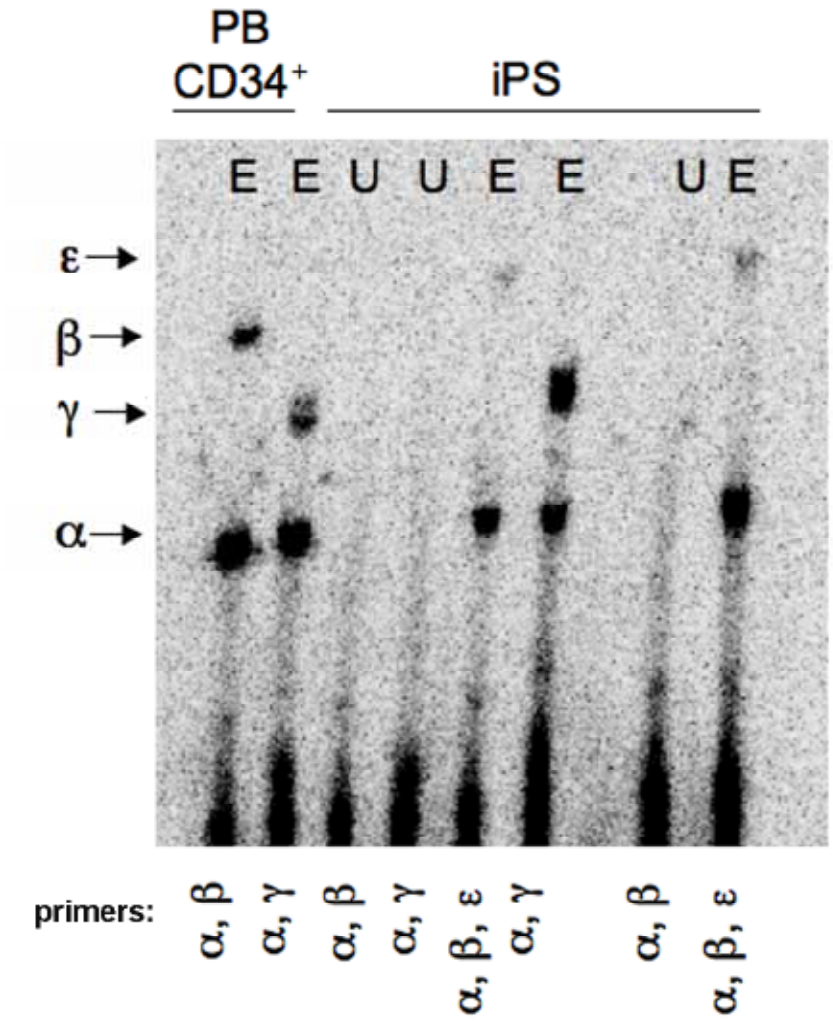


Supplementary Figure 15. Expression of erythroid lineage markers in the erythroid progeny of single globin vector copy clones. Expression of cell surface markers CD71 and GPA in the erythroid progeny of the 13 thal-iPS cell clones harboring a single globin vector copy, as well as in erythroid cells derived in vitro from peripheral blood (PB) $CD34^+$ cells of a healthy individual.

Nature Biotechnology: doi: 10.1038/nbt.1717

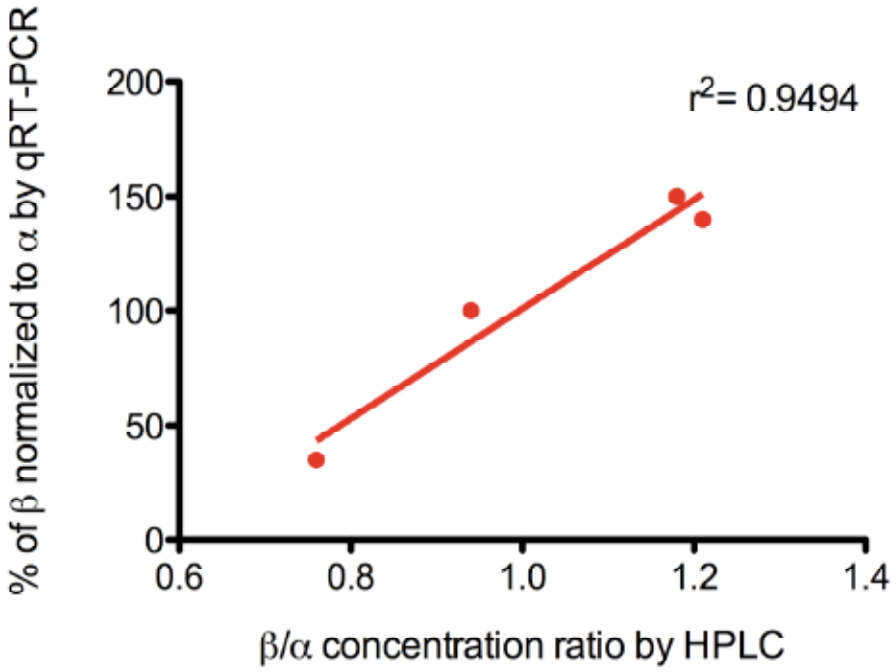


Supplementary Figure 16. Induction of erythroid-specific genes in erythroid cell derivatives of thal-iPS cells. Expression levels of erythroid-specific genes (listed below the X axis) in single vector copy thal-iPS cell clones, before (green) and after (red) erythroid differentiation. Horizontal lines denote median expression.

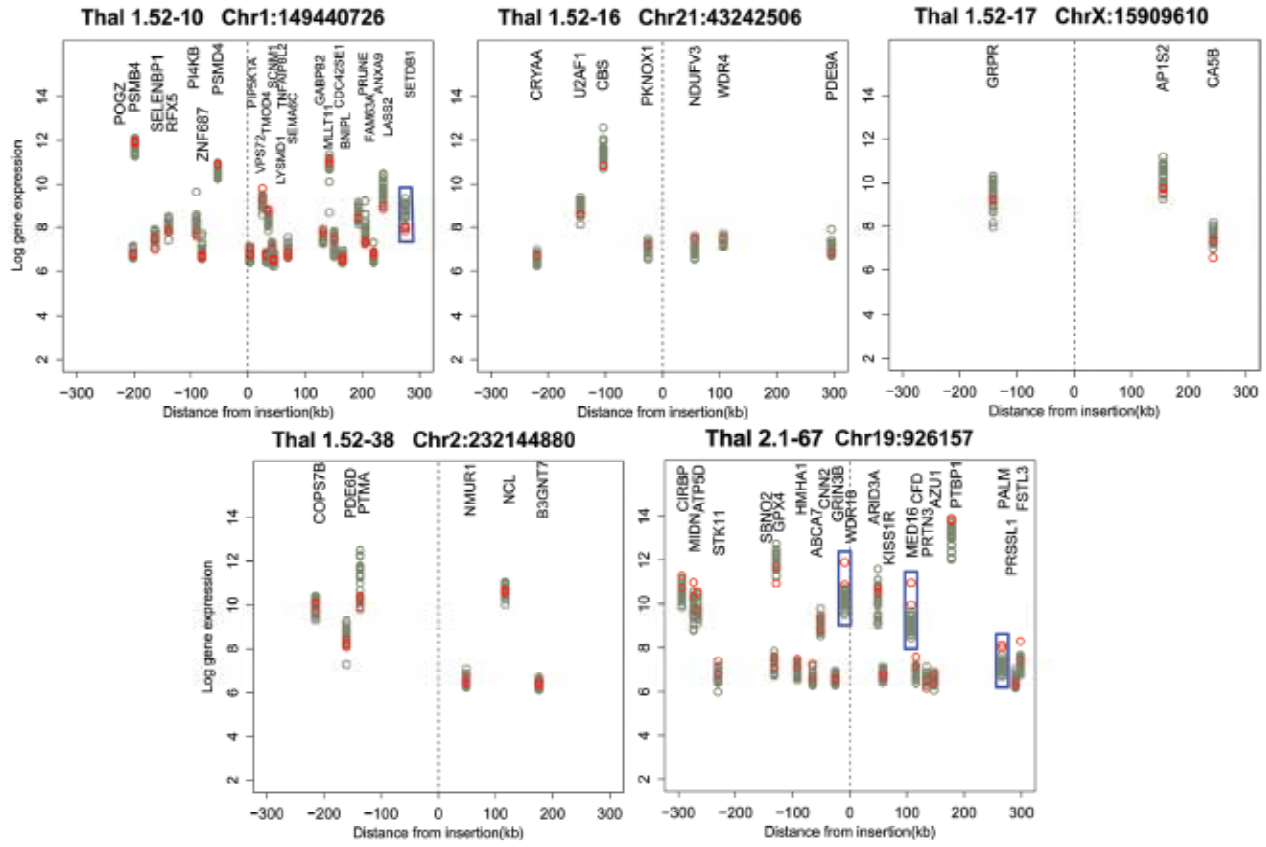


Supplementary Figure 17. Globin expression in the erythroid progeny of iPS cells. Analysis of α-, ß-, γ- and ε- globin expression by primer extension (combinations of primers are indicated below the blot) in erythroid cells (E) derived in vitro from peripheral blood (PB) $CD34^+$ cells of a normal donor and from the iPS cell line FDCT, generated from fibroblasts of a healthy individual, as well as in undifferentiated FDCT iPS cells (U).

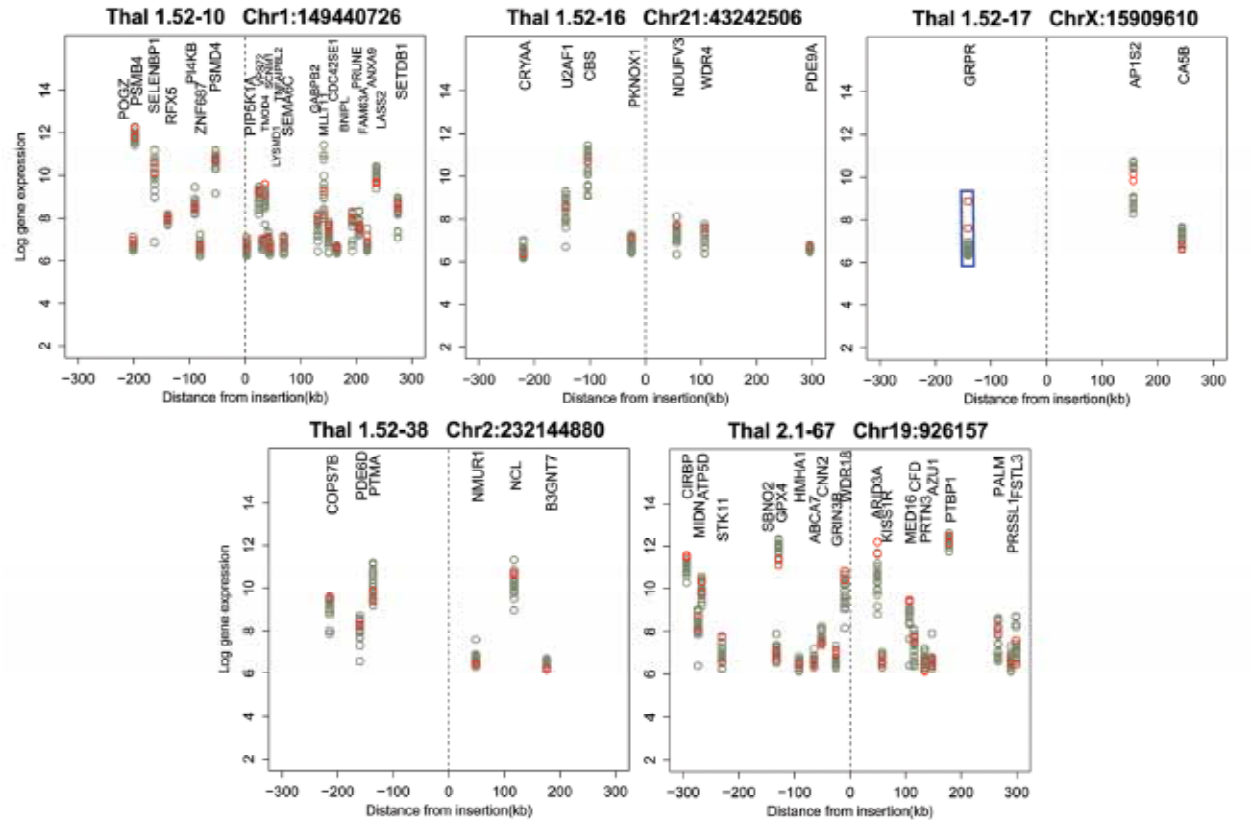
Nature Biotechnology: doi: 10.1038/nbt.1717



Supplementary Figure 18. Correlation of vector-encoded β-globin expression measured at the mRNA and protein level. X axis: ratio of concentration of β- to α- globin measured by HPLC (Supplementary Table 7). Y axis: % of ß-globin expression per gene copy, relative to the average endogenous ß-globin expressed in the in vitro differentiated erythroid progeny of peripheral blood $CD34^+$ cells from 4 healthy individuals, and normalized to the endogenous α-globin expression, measured by qRT-PCR.



Supplementary Figure 19. Expression of endogenous genes in the vicinity of the vector integration site in undifferentiated single copy thal-iPS cell clones. Expression of all genes present within 300 kb on either side of the single vector insertion in undifferentiated thal-iPS cell clones 1.52-10, 1.52-16, 1.52-17, 1.52-38 and 2.1-67. Red open circles denote gene expression measured in the given clone harboring each integration, as indicated. Grey open circles denote gene expression in the other clones, as well as in untransduced lines thal1.52 and thal5.11-Cre23. Differentially expressed genes are marked with blue boxes and include SETDB1 in clone thal1.52-10 and WDR18, MED16 and PALM in clone thal2.1-67. No other genes – either in the vector neighborhood or in a genome-wide survey - were found to be significantly differentially expressed ($p < 0.05$).



Supplementary Figure 20. Expression of endogenous genes in the vicinity of the vector integration site in the erythroid progeny of single copy thal-iPS cell clones. Expression of all genes present within 300 kb on either side of the single vector insertion in the erythroid progeny of thal-iPS cell clones 1.52-10, 1.52-16, 1.52-17, 1.52-38 and 2.1-67. Red open circles denote gene expression measured in the given clone harboring each integration, as indicated. Grey open circles denote gene expression in the other clones, as well as in untransduced lines thal1.52 and thal5.11-Cre23. We find that gene GRPR is differentially expressed in thal1.52-17 clone. We did not find any genes – either in the vector neighborhood or in a genome-wide survey - to be significantly differentially expressed ($p < 0.05$) in the other clones.

Supplementary Table 1

ABCB10, ABL2, ACADM, ALG14, AMPD1, APH1A, ARID1A, ARNT, BCL10, BCL9, BTBD19, BTG2, C1ORF183, C1ORF55, C1ORF63, C1ORF64, CD46, CD52, CDC73, CDK11A, CDK11B, CDK18, CDKN2C, CHD5, CKS1B, CLIC4, CNR2, COL11A1, CR2, CSF1, CSF3R, CYR61, DHRS3, DNAJC6, DPT, E2F2, ELK4, EPHA2, EPS15, ETV3, FABP3, FCGR2B, FCRL4, FCRL5, FGR, FH, FMOD, GADD45A, GFI1, GPR161, GSTM1, HIVEP3, ID3, IL24, INADL, IRF2BP2, JAK1, JUN, LAPTM5, LCK, LGR6, LPHN2, MACF1, MAP3K6, MCL1, MDM4, MDS2, MEF2D, MLLT11, MPL, MTHFR, MUC1, MUTYH, MYCL1, NAV1, NBL1, NOTCH2, NRAS, NTRK1, NUP133, OBSCN, PADI2, PAPPA2, PAX7, PBX1, PDE4DIP, PIGR, PLA2G2A, PLCH2, PPAP2B, PPP1R8, PRCC, PRDM16, PRDM2, PRDX1, PRG4, PRRX1, PSMA5, PTCH2, PTPN14, PTPRU, RABGAP1L, RABGGTB, RALGPS2, RAP1A, RAP1GAP, RASAL2, RBBP5, RBM15, RBM34, RCC2, REG4, RERE, RFX5, RGL1, RGS1, RHBG, RHOC, RLF, RNASEL, RORC, RPL22, RPS6KA1, RUNX3, S100A14, S100A4, SDC3, SDCCAG8, SDHB, SDHC, SFPQ, SHC1, SKI, SLAMF6, SLC45A3, SMYD3, STIL, TACSTD2, TAL1, TGFB2, TGFBR3, THBS3, THRAP3, TIE1, TNFRSF14, TNFRSF1B, TNFRSF8, TNN, TNR, TP73, TPM3, TPR, TRIM33, TRIM62, UCHL5, USP24, USP48, ZBTB48, ZDHHC18, ABI1, ACSL5, ADAM12, ADARB2, ADD3, ALOX5, AVPI1, BMI1, BMPR1A, C10ORF137, CCDC6, CDK1, CUBN, DIP2C, DMBT1, DNAJC9, DUX4, ERCC6, FAS, FBXW4, FGF8, FGFR2, GATA3, GDF10, HHEX, HNRNPA3P1, IL2RA, KLF6, LDB1, MAP3K8, MAPK8, MLLT10, MXI1, MYST4, NCOA4, NFKB2, PAX2, PLCE1, PRF1, PSAP, PTEN, PTPRE, RET, SGMS1, SUFU, TCF7L2, TET1, TLX1, VCL, WDR11, ZMIZ1, ACCS, ADAMTS15, ARHGAP20, ARHGEF12, ATHL1, ATM, BAD, BIRC3, BRMS1, C11ORF30, CALCA, CARS, CBL, CCND1, CD248, CD44, CD6, CD82, CDKN1C, CLP1, CTNND1, CTTN, DDB2, DDX10, DDX6, DNAJC24, ETS1, EXT2, FANCF, FGF3, FGF4, FLI1, FNBP4, FOSL1, GSTP1, GTF2H1, GUCY1A2, HRAS, IGF2, INTS4, LMO1, LMO2, LTBP3, MALAT1, MAML2, MAP4K2, MDK, ME3, MEN1, MLL, MRE11A, MRVI1, MYEOV, MYOD1, NRXN2, NUMA1, NUP98, PAFAH1B2, PCSK7, PICALM, PLEKHB1, POU2AF1, PPP1CA, PPP1R14B, PRKCDBP, PTH, RASGRP2, RASSF7, RELA, RHOG, RRAS2, SCN3B, SCYL1, SDHAF2, SDHD, SIK3, SLC22A18, SLC43A1, SORL1, SPI1, ST5, SWAP70, TCIRG1, TECTA, THRSP, TMEM86A, TSG101, WT1, ZBTB16, AGAP2, ALDH2, APAF1, ARHGAP9, ATF1, ATF7IP, BCL7A, BHLHE41, BTG1, C3AR1, CAMKK2, CCND2, CCNT1, CDK2, CDK2AP1, CDK4, CDKN1B, CELA1, CHFR, CORO1C, DDIT3, DTX3, ELK3, EMP1, EPS8, ERBB3, ERC1, ETV6, FGF6, FMNL3, GLI1, HMGA2, HNF1A, HOXC11, HOXC13, HSP90B1, IGF1, IGFBP6, INHBE, KCNH3, KDM2B, KDM5A, KIAA0528, KITLG, KRAS, KRT73, LMO3, LRMP, LTBR, MARS, MDM2, NACA, NCOR2, P2RX7, PPFIBP1, PPP1CC, PPP1R1A, PPTC7, PRKAB1, PTHLH, PTPN11, PTPN6, PXN, RAD52, RAP1B, RARG, RECQL, RHOF, SARNP, SBNO1, SCARB1, SILV, SLC38A1, SLC38A2, SOX5, SSPN, STK38L, TNFRSF1A, WNT1, WNT10B, ZNF384, BRCA2, CDX2, ERCC5, FGF14, FGF9, FLT1, FLT3, FOXO1, GAS6, GPC5, HMGB1, INTS6, KLF5, LCP1, LHFP, LMO7, PDS5B, RAP2A, RASA3, RB1, UBL3, ZMYM2, AKT1, ARF6, ARID4A, BATF, BCL11B, BCL2L2, C14ORF49, CCNB1IP1, CDKN3, CHGA, CRIP2, DICER1, DLST, EVL, FOS, FOXG1, FSCB, GMFB, GOLGA5, GPHN, GPR68, GTF2A1, HSP90AA1, IFI27, JDP2, KIAA1409, KTN1, MAX, MMP14, MNAT1, MTA1, NDRG2, NFKBIA, NIN, PAPOLA, PAX9, PGF, PLEK2, PNMA1, PRKD1, PTGR2, RAD51L1, RNASE10, SEL1L, SERPINA1, SIX1, SIX4, SYNE2, TCL1A, TCL6, TGFB3, TRAF3, TRIP11, TSHR, XRCC3, ZFYVE26, ABHD2, ADAMTSL3, AKAP13, ANP32A, BCL2A1, BLM, BUB1B, C15ORF2, C15ORF21, C15ORF55, CASC5, CHD2, CHRNB4, CORO2B, CRTC3, CSPG4, CYP19A1, CYP1A1, FES, FGF7, GABPB1, GATM, IDH2, IGF1R, IL16, LACTB, MAP2K5, MKRN3, NTRK3, PML, RASGRF1, RASGRP1, RHCG, SEMA4B, SMAD3, TCF12, THBS1, UBL7, USP3, ABCA3, ABCC1, ADAMTS18, AMFR, BCAR1, C16ORF80, CBFA2T3, CBFB, CCNF, CDH1, CDH11, CDH13, CHD9, CIITA, CORO1A, COTL1, CREBBP, CYLD, DOC2A, E2F4, ERCC4, FANCA, FOXC2, FUS, GALNS, GSPT1, HERPUD1, HPR, IGFALS, IL21R, IL4R, ITGAX, JMJD5, KIAA0182, MAF, MMP2, MVP, MYH11, NFAT5, NFATC3, NME3, NQO1, PALB2, PLK1, RBBP6, RBL2, SCNN1B, SOCS1, SOX8, STX4, TFAP4, TNFRSF17, TSC2, UQCRC2, VAC14, WFDC1, ZDHHC7, ZFHX3, ZNF423, ZNF668, AATF, ABR, ACAP1, ALOX12, ALOX15, ASPSCR1, BCAS3, BRCA1, BRIP1, BZRAP1, CANT1, CASC3, CBX4, CCL7, CCR7, CD79B, CDC27, CDK5R1, CHD3, CLTC, COL1A1, CRK, CUEDC1, CYTSB, DDX5, DGKE, DNAH9, DPH1, ELAC2, ERBB2, ETV4, FGF11, FLCN, FLII, GAS7, GIP, GRB2, HEXIM1, HIC1, HLF, HOXB8, ITGB4, KCTD2, KLHL10, LASP1, LGALS3BP, MAFG, MAP2K4, MLLT6, MSI2, MYH1, NBR1, NF1, NGFR, NLE1, NME1, PECAM1, PER1, PFN1, PHB, PPM1E,

PRKAR1A, RAB5C, RABEP1, RARA, RECQL5, RNF157, RNF213, SEPT9, SLC16A3, SLC26A11, SMG6, SSTR2, STARD3, STAT3, STAT5A, STAT5B, SUZ12, TAF15, TBX2, TBX21, THRA, TIMP2, TK1, TMEM104, TNFAIP1, TNFRSF13B, TOP2A, TP53, TRAF4, TRIM37, USP6, WNT3, ATP8B1, BCL2, CDH20, DCC, DLGAP1, DYM, GRP, KDSR, KIAA0427, KIAA1632, MALT1, MC2R, MPPE1, NFATC1, SERPINB3, SERPINB4, SERPINB5, SMAD2, SMAD4, SS18, TNFRSF11A, YES1, ZNF521, ACTL9, AKT2, APC2, AXL, BAX, BBC3, BCL3, BRD4, CBLC, CCNE1, CD79A, CDC34, CDKN2D, CEACAM3, CEACAM5, CEACAM7, CEBPA, CIC, CIRBP, CRTC1, ELL, EPOR, EPS15L1, ERCC2, FOSB, FSTL3, GRIN2D, HKR1, ICAM1, ICAM5, IER2, IFI30, JAK3, JUNB, JUND, KEAP1, KLK2, KLK3, KLK6, LOC729991-MEF2B, LYL1, MARK4, MATK, MLLT1, NAPA, NFIC, NFIX, NLRP2, NLRP8, NR2F6, PDE4A, PIK3R2, PLEKHG2, POLD1, POU2F2, PPP1R13L, PRSSL1, PSG2, PTBP1, PTPRH, RAB8A, RAD23A, RELB, RFX2, RRAS, S1PR4, SCAMP4, SH3GL1, SLC27A1, SMARCA4, SPPL2B, SSBP4, STK11, SYDE1, TCF3, TFPT, TGFB1, TPM4, VAV1, ZFP36, ZNF146, ZNF331, ZNF442, ZNF569, ACSL3, AFF3, ALK, ANKRD44, ARHGAP25, ARHGEF4, ARID5A, ASAP2, ATIC, BCL11A, BIN1, BIRC6, BOK, BUB1, C2ORF65, CAPG, CHN1, CNNM4, COL4A3, CREB1, CXCR7, DIRC1, DIRC3, DUSP2, E2F6, EIF2AK2, EML4, ERBB4, ERCC3, FEV, FN1, FOSL2, GALNT5, GEN1, GLI2, HDAC4, HDLBP, HNRNPA3, HOXD11, HOXD13, ID2, IDH1, IGFBP2, IGFBP5, INHA, KDM3A, LHCGR, LRP2, LRRFIP1, MAP2, MEIS1, MERTK, MSH2, MSH6, MXD1, MYCN, NCOA1, NEURL3, OTOF, PAX3, PAX8, PLCL1, PMS1, PRKRA, PTMA, QPCT, RAB1A, RALB, RAMP1, RANBP2, RAPH1, REG3A, REL, RHOB, SAG, SLC9A2, SP110, TGFA, THADA, TTL, USP39, VAMP8, WNT6, XDH, YWHAQ, ZC3H8, ZEB2, ZFP36L2, ARFRP1, ASXL1, BCAS4, BCL2L1, CDC25B, CEBPB, CSE1L, CSNK2A1, CSRP2BP, E2F1, EEF1A2, ERGIC3, GHRH, GNAS, HCK, MAFB, MAPRE1, MMP9, MYBL2, NCOA3, NCOA6, PCNA, PHF20, PKIG, PLAGL2, PLCB1, PMEPA1, PPP1R16B, PTPRT, RBL1, RBM39, RGS19, SERINC3, SNAI1, SNX5, SRC, SRSF6, SS18L1, STK4, SULF2, TH1L, TOP1, WISP2, YWHAB, ZFP64, ZNF217, CRYAA, DOPEY2, ERG, ETS2, IFNAR1, IFNGR2, OLIG2, PDXK, PKNOX1, PSMG1, PTTG1IP, RSPH1, RUNX1, SLC19A1, TFF1, TFF2, TFF3, TIAM1, TMPRSS2, TMPRSS3, BCR, BID, CDC42EP1, CECR5, CHEK2, CLTCL1, CRKL, CYP2D6, CYP2D7P1, DNAL4, EP300, EWSR1, GGA1, GSTT1, IL2RB, MCM5, MKL1, MMP11, MN1, MTMR3, MYH9, NF2, PATZ1, PDGFB, PICK1, PIM3, PKDREJ, SEPT5, SMARCB1, SOX10, ST13, THOC5, TIMP3, TMPRSS6, TOB2, ACPP, ALS2CL, APEH, ATR, BCL6, CBLB, CCNL1, CCR1, CDC25A, CHL1, CMTM7, CNBP, CNTN4, CNTN6, COL7A1, CTNNB1, DIRC2, DNASE1L3, ECT2, EIF1B, EIF4A2, EIF4G1, EPHA3, ETV5, FANCD2, FGF12, FHIT, FILIP1L, FLNB, FOXL2, FOXP1, GATA2, GHRL, GLB1, GMPS, GNAI2, GPX1, HCLS1, HES1, HSPBAP1, IFRD2, IL17RB, IRAK2, ITGA9, LPP, MANF, MCM2, MECOM, MITF, MLF1, MLH1, MME, MST1, MST1R, NCK1, NCKIPSD, P2RY14, PIK3CA, PLSCR4, PPARG, PSMD2, PTPRG, RAF1, RAP2B, RARB, RASSF1, RHOA, RPN1, RRP9, SATB1, SEMA3F, SEMA5B, SETD2, SKIL, SLC9A9, SRGAP3, ST6GAL1, STAG1, TCTA, TDGF1, TFG, TFRC, TGFBR2, THPO, TM4SF1, TOP2B, TTLL3, TUSC2, UBA7, USP4, VEPH1, VHL, WDR82, WNT5A, XIRP1, XPC, ZDHHC3, ZNF35, ADAM29, AFF1, ANAPC10, ARAP2, AREG, CAMK2D, CCNA2, CHIC2, CXCL1, CXCL10, CXCL2, CXCL3, EGF, EIF4E, FBXW7, FGF2, FGF5, FGFR3, FIP1L1, FRYL, GAB1, GAK, HPSE, IGFBP7, IL2, JAKMIP1, KDR, KIT, KLF3, LEF1, LRBA, MAD2L1, MAPK10, NFKB1, PDGFRA, PHOX2B, RAP1GDS1, RCHY1, RG9MTD2, RHOH, SH3BP2, SLC25A4, SYNPO2, TEC, TET2, USP38, UTP3, WHSC1, ACSL6, AFF4, APC, ARHGAP26, CCNB1, CCNG1, CCNH, CD74, CDC25C, CDH10, CDK7, CPLX2, CSF1R, CTNNA1, CTNND2, DAB2, DBN1, EBF1, ENC1, ERCC8, FAT2, FBXL7, FGF1, FGFR4, GHR, HAPLN1, HMGCR, IL3, IL5, IL6ST, IL9, IRF1, ITK, LIFR, LNPEP, MAPK9, MARCH3, MCC, MEF2C, MRPS30, MSH3, N4BP3, NIPBL, NKX2-5, NPM1, NR2F1, NSD1, PCDHB7, PDGFRB, PIK3R1, PITX1, PRLR, RANBP17, RASA1, RASGRF2, SKP1, SKP2, SLC30A5, SLC6A3, TERT, TGFBI, TLX3, TRIM23, UBE2D2, AHI1, AKAP12, BACH2, BAK1, BRD2, CAGE1, CCNC, CCND3, CD109, CDKN1A, COL19A1, CRISP1, CTGF, CYB5R4, DDR1, DEK, DST, E2F3, EEF1A1, ESR1, EYA4, FANCE, FBXO5, FGFR1OP, FOXO3, FYN, GOPC, HFE, HIST1H1B, HIST1H1C, HIST1H4I, HIVEP1, HMGA1, HSP90AB1, IGF2R, IRF4, KCNK5, KCNQ5, KPNA5, LTA, LTB, MAS1, MLLT4, MYB, NFKBIE, NKAIN2, NOTCH4, PBX2, PHIP, PIM1, PKHD1, POLH, POU5F1, PRDM1, PRPF4B, PTCRA, PTK7, ROS1, RREB1, RTN4IP1, RUNX2, SERPINB1, SH3BGRL2, SLC29A1, SLC44A4, SOX4, SRSF3, STL, SYNE1, TAP1, TAP2, TFEB, THBS2, TNF, TNFAIP3, TRIM27, VEGFA, WISP3, ZDHHC14, ZNF318, ABCB1, ABCB4, ABCB8, ACTB, ADCYAP1R1, AKAP9, ASL, BRAF, C7ORF27, CADPS2, CALCR, CAMK2B, CARD11, CDK5, CDK6, CLDN4, COL28A1, CREB3L2, CUX1, DTX2, EGFR, ELMO1, ELN, EPHA1, EPHB6, EPO, ETV1, EXOC4, EZH2, GIMAP7, GLI3, GNA12, GPNMB, GRB10, HGF,

HIP1, HNRNPA2B1, HOXA11, HOXA13, HOXA3, HOXA7, HOXA9, IGFBP1, IGFBP3, IKZF1, ING3, INTS1, IQCE, JAZF1, KIAA1147, KIAA1549, LFNG, MAFK, MET, MLL3, MNX1, NOD1, NRCAM, OSBPL3, PDAP1, PDGFA, PLEKHA8, PMS2, POU6F2, PTN, RALA, SBDS, SH2B2, SHH, SMO, SRI, SRPK2, TPK1, TRIM24, TRRAP, VGF, WNT2, AP3M2, ASAP1, BAALC, CHCHD7, COX6C, CSMD3, DLC1, EIF3E, EXT1, FGFR1, GLI4, GML, HOOK3, IL7, LY6E, LYN, MFHAS1, MOS, MYBL1, MYC, MYST3, NAT1, NBN, NCOA2, NKX3-1, NOV, PCM1, PDGFRL, PLAG1, PLEKHA2, PSCA, PTP4A3, PVT1, RB1CC1, RECQL4, RNF139, RSPO2, RUNX1T1, SNAI2, TCEA1, TMEM74, TRHR, WHSC1L1, WRN, ABCA1, ABL1, ABO, ALAD, ANP32B, ANXA1, AUH, BAG1, BRD3, BSPRY, C9ORF103, CD274, CDKN2A, CDKN2B, CEP110, CKS2, CRB2, DAB2IP, EHMT1, FANCC, FANCG, FNBP1, GADD45G, GAS1, GGTA1, GNAQ, GSN, JAK2, KDM4C, KLF4, MAMDC4, MLLT3, NFIB, NOTCH1, NR4A3, NUP214, OMD, PAX5, PRRX2, PSIP1, PTCH1, PTPRD, RAD23B, RGS3, SET, SNAPC3, SPTAN1, SUSD3, SYK, TAL2, TEK, TGFBR1, TRAF1, TSC1, UHRF2, VAV2, XPA, AIFM1, AR, ARAF, ARHGAP4, ARMCX1, ARMCX2, ARMCX3, BCOR, BGN, BRWD3, BTK, CDK16, CFP, CRLF2, DACH2, DKC1, EGFL6, EIF1AX, ELF4, ELK1, ERAS, FAM123B, FGF13, FOXO4, GATA1, GPC3, GRPR, HDAC6, IL2RG, IL3RA, KDM5C, KDM6A, L1CAM, MAGEE1, MCF2, MID1, MSN, MTCP1, NAA10, NONO, P2RY8, PHF6, PIM2, PRPS1, SEPT6, SOX3, SSX1, SSX2, SSX4, STARD8, TBX22, TFE3, TIMP1, WAS,

Supplementary Table 1. List of cancer-related genes. The database is periodically updated, combining updates of the Sanger Center Cancer Gene Census and various other sources that can be found at http://microb230.med.upenn.edu/protocols/cancergenes.html. The current database comprises 1468 genes (most recent update 7/2010).

Patient	1						2						3	4	5		
Genre	Male						Female						Female	Male	Female		
Beta-thalassemia genotype	ß^0/ß$^+$						ß^0/ß$^+$						ß^0/ß$^+$	ß^0/ß$^+$	ß^0/ß^0		
Beta-thalassemia mutations	ß39/IVS I-110						ß39/IVS I-1						ß39/IVS I-6	ß39/IVS I-5	ß39/ß39		
Reprogramming vector	4 vectors[15]						SV2A						NA	fSV2A	fSV2A		
Starting somatic cell type	BM MSC						BM MSC						skin fibroblast	skin fibroblast	BM MSC		
Total putative thal-iPS cell lines generated	8						7						0	1	4		
Thal-iPS cell lines characterized	1.52					1.60	2.1					2.4		4.1	5.10	5.11	
hES cell morphology	YES					YES	YES					YES		YES	YES	YES	
Reprogramming vector copy number (qPCR)	41					97	10					11		10	6**	6**	
Reprogramming vector excision	NA					NA	NA					NA		ND	YES (6/47)	YES (7/42)	
Karyotype	normal					normal	normal					normal		normal	normal	normal	
Pluripotency markers	YES					YES	YES					YES		YES	YES	YES	
Expression of pluripotency genes	YES					YES	YES					YES		YES	YES	YES	
Silencing of reprogramming factors	YES					YES	YES					YES		YES	YES	YES	
Oct4 promoter methylation	YES					ND	YES					ND		ND	YES	YES	
Teratoma formation	YES					YES	YES					YES		YES	YES	YES	
Globin vector transduction	+						+								+	+	
Number of clones screened*	16						38								17	36	
Number of single vector copy clones (qPCR)	10						16								8	17	
Clonality confirmed by Southern Blot	5/5						5/5								1/1	2/4	
Single copy integrant mapped to genome	1.52-10	1.52-16	1.52-17	1.52-21	1.52-38		2.1-48	2.1-49	2.1-55	2.1-58	2.1-67				5.10-2	5.11-28	5.11-32
Within 50 kb of 5' end of any gene	+	+	-	-	+		+	+	+	-	+				-	-	+
Within 300 kb of oncogene	+	+	+	+	+		+	+	+	-	+				-	-	-
Within 300 kb of miRNA	-	-	-	-	+		-	-	-	-	-				-	-	-
Within transcription unit	+	-	-	+	-		-	+	+	+	-				-	+	+
Within UCRs	-	-	-	-	-		-	-	-	-	-						
Fulfills "safe harbor" criteria	-	-	-	-	-		-	-	-	-	-				+	-	-
eGFP expression	+	+	+	-	+		-	-	+	-	+				+	+	ND
Beta-globin expression	37%	19%	30%	32%	52%		9%	0%	26%	159%	44%				85%	156%	41%
Microarray gene expression analysis	YES	YES	YES	ND	YES		ND	ND	ND	ND	YES				YES	ND	ND

ND: not done

* see Supplementary Table 5 for details

**Lines 5.10 and 5.11 have the same 6 fSV2A integrations

Supplementary Table 2. Flow chart of thal-iPS cell lines generated and globin clones analyzed in this study. (Lines thal5.10 and thal5.11 were established by picking independent colonies, expanded and characterized separately and were found - after mapping of the fSV2A vector integration sites - to harbor the same 6 reprogramming vector integrations.)

IS	Chromosome	Position	Orientation	Inside gene	Inside exon
1	8	70110895	-	-	-
2	10	124176060	-	+	-
3	6	73008498	-	+	-
4	17	3673786	-	+	-
5	13	75322383	-	+	-
6	6	81106221	-	+	-

Supplementary Table 3. Integration sites of the fSV2A vector in iPS cell line thal5.10.

Criteria	Human Genome	Globin vector integration sites in thal-iPS cells (n=5840)	Globin vector integration sites in thal-iPS cell clones (n=36)
> 50kb of 5' end of any gene	62.85	45	44.44
> 300kb of oncogene 5' or 3' end	75.49	58.58	47.22
> 300kb of miRNA 5' or 3' end	91.27	83.8	86.11
Outside a gene transcription unit	62.38	34.91	38.89
Outside UCRs	99.9959	99.98	100
All five criteria combined	37.94	17.29	8.33

Supplementary Table 4. Bioinformatics analysis of the proportion of human genome and lentiviral vector integration sites within regions meeting our five "safe harbor" criteria, independently and in combination. Numbers denote percentages. Integration site datasets analyzed include 5840 integration sites mapped in polyclonal thal-iPS cells 5 days following transduction with TNS9.3/fNG, as well as in a total of 36 integrations retrieved in single and multiple copy thal-iPS cell clones after the selection process. The 36 sites are shown individually in Table 1 (single-copy clones) and in Supplementary Table 6 (multiple copy clones). Total bases in human genome draft hg18: 3080419480, total genes used: 20884, total cancer-related genes: 1468, total miRNAs: 705.

Clone	MOI	%eGFP$^+$ 48h	VCN
1.52-10	0.3	35%	1
1.52-14	0.3	35%	1
1.52-15	0.3	35%	2
1.52-16	0.3	35%	1
1.52-17	0.3	35%	1
1.52-18	0.3	35%	2
1.52-19	0.3	35%	1
1.52-20	0.3	35%	2
1.52-21	0.3	35%	1
1.52-22	0.3	35%	1
1.52-23	0.3	35%	2
1.52-24	0.3	35%	3
1.52-25	0.3	35%	1
1.52-36	0.3	35%	1
1.52-37	0.3	35%	5
1.52-38	0.3	35%	1

Single copy clones	10
Clones screened	16

Clone	MOI	%eGFP$^+$ 48h	VCN
2.1-35	0.3	26%	2
2.1-36	0.3	26%	5
2.1-37	0.3	26%	1
2.1-38	0.3	26%	2
2.1-39	0.3	26%	2
2.1-40	0.3	26%	4
2.1-41	0.3	26%	0
2.1-42	0.3	26%	2
2.1-43	0.3	26%	2
2.1-44	0.3	26%	2
2.1-45	0.3	26%	1
2.1-46	0.3	26%	1
2.1-47	0.3	26%	2
2.1-48	0.3	26%	1
2.1-49	0.3	26%	1
2.1-50	0.3	26%	2
2.1-51	0.3	26%	2
2.1-52	0.3	26%	3
2.1-53	0.3	26%	1
2.1-54	0.3	26%	2
2.1-55	0.3	26%	1
2.1-56	0.3	26%	1
2.1-57	0.3	26%	1
2.1-58	0.3	26%	1
2.1-59	0.3	26%	2
2.1-60	0.1	13%	1
2.1-61	0.1	13%	1
2.1-62	0.1	13%	2
2.1-63	0.1	13%	0
2.1-64	0.1	13%	1
2.1-65	0.1	13%	1
2.1-66	0.1	13%	0
2.1-67	0.1	13%	1
2.1-68	0.1	13%	2
2.1-69	0.1	13%	0
2.1-70	0.1	13%	0
2.1-71	0.1	13%	0
2.1-72	0.1	13%	1

Single copy clones	16
Clones screened	38

Clone	MOI	%eGFP$^+$ 48h	VCN
5.10-1	0.5	40%	2
5.10-2	0.5	40%	1
5.10-3	0.5	40%	1
5.10-4	0.5	40%	2
5.10-5	0.5	40%	4
5.10-6	0.5	40%	1
5.10-20	0.5	40%	1
5.10-21	0.5	40%	1
5.10-22	0.5	40%	2
5.10-23	0.5	40%	2
5.10-24	0.5	40%	3
5.10-25	0.5	40%	1
5.10-30	0.5	40%	3
5.10-31	0.5	40%	1.5
5.10-34	0.5	40%	1
5.10-41	0.1	5%	1
5.10-42	0.1	5%	0

Single copy clones	8
Clones screened	17

Clone	MOI	%eGFP$^+$ 48h	VCN
5.11-1	0.3	32%	2
5.11-2	0.3	32%	5
5.11-3	0.3	32%	1
5.11-4	0.3	32%	2
5.11-5	0.3	32%	2
5.11-6	0.3	32%	4
5.11-7	0.3	32%	11
5.11-8	0.3	32%	1
5.11-9	0.3	32%	2
5.11-10	0.3	32%	2
5.11-11	0.3	32%	1
5.11-12	0.3	32%	1
5.11-14	0.3	32%	1
5.11-15	0.3	32%	2
5.11-16	0.3	32%	1
5.11-17	0.3	32%	2
5.11-18	0.3	32%	2
5.11-19	0.3	32%	1
5.11-20	0.3	32%	2
5.11-21	0.3	32%	3
5.11-22	0.1	8%	1
5.11-24	0.1	8%	1
5.11-25	0.1	8%	2
5.11-26	0.1	8%	0
5.11-27	0.1	8%	1
5.11-28	0.1	8%	1
5.11-29	0.1	8%	0
5.11-30	0.1	8%	1
5.11-32	0.1	8%	1
5.11-37	0.1	8%	1
5.11-38	0.1	8%	2
5.11-39	0.1	8%	0
5.11-40	0.1	8%	1
5.11-41	0.1	8%	1
5.11-42	0.1	8%	1
5.11-43	0.1	8%	2

Single copy clones	17
Clones screened	36

Total single copy clones	51
Total clones screened	107

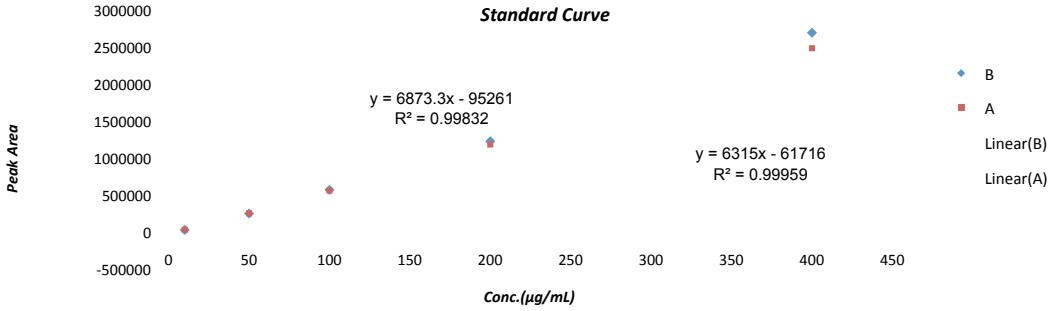
Supplementary Table 5. Screening of vector copy number (VCN) in clones from thal-iPS cell lines thal1.52, thal2.1, thal5.10 and thal5.11 transduced with the TNS9.3/fNG vector by qPCR. All clones with a VCN value close to 1 were re-analyzed in at least 2 independent qPCR reactions, in duplicate or triplicate wells. The average VCN value for each clone, rounded to the nearest whole number, is shown. 15 of these clones found to harbor a single vector copy were further analyzed by Southern Blotting (Fig. 2a, b and Supplementary Fig. 13) and selected for further studies. MOI: multiplicity of infection, %eGFP$^+$ 48h: percentage of eGFP$^+$ cells 48h after transduction, VCN: vector copy number.

Chromosome	Position	Orientation	Within 50kb of 5' end of any gene	Within 300kb of cancer gene 5' or 3' end	Within 300kb of miRNA 5' or 3' end	Inside a gene transcription unit	Within UCR
1	210,993,568	+	+	-	-	+	-
3	50,035,283	+	-	+	+	+	-
3	72,936,602	-	+	-	-	+	-
4	1,671,756	+	+	+	+	+	-
5	14,247,861	+	-	-	-	+	-
7	30,357,713	+	+	+	+	+	-
7	151,522,467	-	-	+	-	+	-
8	96,308,121	+	+	-	-	-	-
10	3,046,320	+	-	-	-	-	-
10	93,733,556	+	-	-	-	+	-
11	45,951,822	-	-	-	-	+	-
11	55,611,666	-	+	-	-	-	-
12	88,418,633	-	+	-	-	+	-
12	109,469,464	-	+	+	-	+	-
15	37,718,440	-	-	+	-	+	-
16	13,149,701	+	-	-	-	+	-
16	56,100,096	-	+	-	-	-	-
17	67,328,980	+	-	-	-	-	-
17	71,300,479	+	+	+	-	+	-
19	2,559,078	-	-	+	-	+	-
19	15,182,452	-	+	+	+	-	-
22	35,619,057	-	+	+	-	-	-
X	8,484,509	-	-	-	-	+	-

Supplementary Table 6. TNS9.3/fNG vector integration sites in thal-iPS cell clones. Additional TNS9.3/fNG integration sites retrieved in multiple copy thal-iPS cell clones. Highlighted in green are sites that meet all 5 "safe harbor" criteria.

Standard Set:

Standard Number	Conc. (µg/mL)	RT (min) 23.2 Peak Area β	RT (min) 27.1 Peak Area α
STD5	10	24689	27343
STD4	50	246630	249298
STD3	100	565504	556850
STD2	200	1222491	1178450
STD1	400	2688113	2478877



Sample Set:

Sample name	RT (min) 23.4 Peak Area β	RT (min) 27.4 Peak Area α	Meas. Conc. (µg/mL) β	Meas. Conc. (µg/mL) α	Dilution Factor	Calc. Conc. (µg/mL) β	Calc. Conc. (µg/mL) α	β/α (conc ratio)	%β (qRT-PCR)
H1 hES	Not integratable	32104	ND	14.86	0.50	ND	7.43	ND	0
Thal1.52 UT	Not integratable	70933	ND	21.01	0.5	ND	10.5	ND	0
Thal1.52-16	3979	58206	14.44	18.99	0.50	7.22	9.50	0.76	0.37
Thal5.11-28	6285	17317	14.77	12.52	0.50	7.39	6.26	1.18	1.56
Thal5.10-2	11264	18874	15.50	12.76	0.50	7.75	6.38	1.21	1.42
Thal2.1-58	12783	43587	15.72	16.68	0.50	7.86	8.34	0.94	1.01

** All cell samples were treated with 0.1 % SDS solution. Injection Vol: 100 µL. (Standard injection Vol: 50 µL)

Supplementary Table 7. Quantification of vector-encoded β-globin expression by HPLC. α- and β-globin standards of known concentration were used to generate calibration curves to permit conversion of peak areas measured in the chromatograms (Fig. 3d) to protein concentration amounts for each globin chain against the reference standards. Last column to the right: β-globin mRNA, normalized to α-globin mRNA, is given as a percentage of one endogenous normal allele, calculated by qRT-PCR, in the same cell sample.

Oligonucleotide	Sequence (5' to 3')
NcoI-Klf4 F primer	CCGGTTccATGGCTGTCAGCGACGCGCTGCTCCC
Klf4-P2A R primer	agggccgggattctcctccacgtcacctgcttgtttgagtagtgagaagtttgttgctccagatccAAA ATGCCTCTTCATGTGTAAGGC
P2A-cMyc F primer	GGATCTGGAgcaacaaacttctcactactcaaacaagcaggtgacgtggaggagaatcccg gccctATGCCCCTCAACGTTAGCTTCACCAA
cMyc-EcoRI R primer	CCGGTTgaattcTTACGCACAAGAGTTCCGTAGCTGTTC
AgeI-cMyc F primer	CCGGTTaccggtATGCCCCTCAACGTTAGCTTCACCAA
cMyc-E2A R primer	gggaccggggttactttcaacatcgccagcgagtttcaacaaagcgtagttagtacattgtccagatcc CGCACAAGAGTTCCGTAGCTGTTC
E2A-Sox2 F primer	GGATCTGGAcaatgtactaactacgctttgttgaaactcgctggcgatgttgaaagtaaccccgg tcccATGTACAACATGATGGAGACGGAGCTG
Sox2-SalI-EcoRI R primer	ccggttgaattcgtcgacTCACATGTGTGAGAGGGGCAGTGTGC
NsiI-Oct4-GSG-T2A-NcoI 5' linker	TtcaaacGGATCTGGAgagggcagaggaagtcttctaacatgcggtgacgtggaggagaat cccggccC
NsiI-Oct4-GSG-T2A-NcoI 3' linker	CATGGggccgggattctcctccacgtcaccgcatgttagaagacttcctctgccctcTCCAGAT CCgtttgaATGCA
loxP-NheI 5' linker	ctagcATAACTTCGTATAATGTATGCTATACGAAGTTATg
loxP-NheI 3' linker	ctagcATAACTTCGTATAGCATACATTATACGAAGTTATg
AgeI-Neo F primer	CCGGTTaccggtATGGGATCGGCCATTGAACAAGATGGATTGCACGCA GG
Neo-P2A R primer	AGGGCCGGGATTCTCCTCCACGTCACCTGCTTGTTTGAGTAGTGAGA AGTTTGTTGCTCCAGATCCgaagaactcgtcaagaaggcgat
P2A-eGFP F primer	GGATCTGGAGCAACAAACTTCTCACTACTCAAACAAGCAGGTGACGT GGAGGAGAATCCCGGCCCTatggtgagcaagggcgaggagctg
eGFP-BglII R primer	CCGGTTagatctTTACTTGTACAGCTCGTCCATGCCG
loxP-MluI 5' linker	cgcgtATAACTTCGTATAATGTATGCTATACGAAGTTATt
loxP-MluI 3' linker	cgcgaATAACTTCGTATAGCATACATTATACGAAGTTATa
loxP-SalI 5' linker	tcgagATAACTTCGTATAATGTATGCTATACGAAGTTATg
loxP-SalI 3' linker	tcgacATAACTTCGTATAGCATACATTATACGAAGTTATc
XmaI-mCherry F primer	CCGGTTcccgggATGGTGAGCAAGGGCGAGGAG
mCherry-P2A R primer	agggccgggattctcctccacgtcacctgcttgtttgagtagtgagaagtttgttgctccagatccCTT GTACAGCTCGTCCATGCCG
P2A-Cre F primer	GGATCTGGAGCAACAAACTTCTCACTACTCAAACAAGCAGGTGACGT GGAGGAGAATCCCGGCCCTatgtccaatttactgaccgtacac
Cre- SalI R primer	ccggttgtcgacCTAATCGCCATCTTCCAGCAGGCGC

Supplementary Table 8. List of oligonucleotides used for vector construction.

Supplementary Table 9

Oligonucleotide	Application	Sequence (5' to 3')
ß-thal-1 F primer	ß-thalassemia genotyping	GAAGAGCCAAGGACAGGT
ß-thal-1 R primer	ß-thalassemia genotyping	CCTTCCTATGACATGAACTTAACCAT
Actin F primer	qRT-PCR (SYBR Green)	TGAAGTGTGACGTGGACATC
Actin R primer	qRT-PCR (SYBR Green)	GGAGGAGCAATGATCTTGAT
OCT4-endo F primer	qRT-PCR (SYBR Green)	CCTCACTTCACTGCACTGTA
OCT4-endo R primer	qRT-PCR (SYBR Green)	CAGGTTTTCTTTCCCTAGCT
SOX2-endo F primer	qRT-PCR (SYBR Green)	CCCAGCAGACTTCACATGT
SOX2-endo R primer	qRT-PCR (SYBR Green)	CCTCCCATTTCCCTCGTTTT
KLF4-endo F primer	qRT-PCR (SYBR Green)	GATGAACTGACCAGGCACTA
KLF4-endo R primer	qRT-PCR (SYBR Green)	GTGGGTCATATCCACTGTCT
cMYC-endo F primer	qRT-PCR (SYBR Green)	TGCCTCAAATTGGACTTTGG
cMYC-endo R primer	qRT-PCR (SYBR Green)	GATTGAAATTCTGTGTAACTGC
hTERT F primer	qRT-PCR (SYBR Green)	TGTGCACCAACATCTACAAG
hTERT R primer	qRT-PCR (SYBR Green)	GCGTTCTTGGCTTTCAGGAT
GDF3 F primer	qRT-PCR (SYBR Green)	AAATGTTTGTGTTGCGGTCA
GDF3 R primer	qRT-PCR (SYBR Green)	TCTGGCACAGGTGTCTTCAG
NANOG F primer	qRT-PCR (SYBR Green)	TGAACCTCAGCTACAAACAG
NANOG R primer	qRT-PCR (SYBR Green)	TGGTGGTAGGAAGAGTAAG
UTF1 F primer	qRT-PCR (SYBR Green)	CCGTCGCTGAACACCGCCCTGCTG
UTF1 R primer	qRT-PCR (SYBR Green)	CGCGCTGCCCAGAATGAAGCCCAC
REX1 F primer	qRT-PCR (SYBR Green)	TCGCTGAGCTGAAACAAATG
REX1 R primer	qRT-PCR (SYBR Green)	CCCTTCTTGAAGGTTTACAC
DPPA4 F primer	qRT-PCR (SYBR Green)	GGAGCCGCCTGCCCTGGAAAATTC
DPPA4 R primer	qRT-PCR (SYBR Green)	TTTTTCCTGATATTCTATTCCCAT
DNMT3B F primer	qRT-PCR (SYBR Green)	TGCTGCTCACAGGGCCCGATACTTC
DNMT3B R primer	qRT-PCR (SYBR Green)	TCCTTTCGAGCTCAGTGCACCACAAAAC
OCT4-pro F primer	Bisulphite sequencing	TAGTTGGGATGTGTAGAGTTTGAGA
OCT4-pro R primer	Bisulphite sequencing	TAAACCAAAACAATCCTTCTACTCC
CCG F primer	qRT-PCR for vector silencing	TCTTCAAGTCCGCCATGC
CCG R primer	qRT-PCR for vector silencing	AAGTCGATGCCCTTCAGC
CCG probe	qRT-PCR for vector silencing	FAM-TCGGCGCGGGTCTTGTAGTT-TAMRA
mCherry F primer	qRT-PCR for vector silencing	AAGCTGAAGGTGACCAAGG
mCherry R primer	qRT-PCR for vector silencing	TTGGAGCCGTACATGAACTG
mCherry probe	qRT-PCR for vector silencing	FAM-CCCTTCGCCTGGGACATCCTG-TAMRA
OKMS F1 primer	qRT-PCR for vector silencing	ACCTCGCCTTACACATGAAG
OKMS R1 primer	qRT-PCR for vector silencing	AGTAGAAATACGGCTGCACC
OKMS probe	qRT-PCR for vector silencing	FAM-CCCCTCAACGTTAGCTTCACCAACA-TAMRA
fSV2A-1 F primer	PCR for vector excision	CAGTCGGCTCCCTCGTTG
fSV2A-1 R primer	PCR for vector excision	GCTTCCTCCACCCACTTCTG
fSV2A-2 F primer	PCR for vector excision	AAGAGGACTTGTTGCGGAAA

LTR F primer	PCR for vector excision	GGCTAATTCACTCCCAACGA
LTR R primer	PCR for vector excision	CAACAGACGGGCACACACTA
Gag F primer	qPCR for vector excision	GGAGCTAGAACGATTCGCAGTT
Gag R primer	qPCR for vector excision	GTTGTAGCTGTCCCAGTATTTGTC
Gag probe	qPCR for vector excision	ACAGCCTTCTGATGTTTCTAACAGGCCAGG
GV1 F primer	qPCR for VCN	GCCTCCCAAAGTGCTATCTG
GV1 R primer	qPCR for VCN	CAGTCAAAGTCGAATGCAGC
GV1 probe	qPCR for VCN	FAM/TCGAGACTAGTGGTACAGATAGGCGG/TAMRA
hAlb F primer	qPCR for VCN	TGAAACATACGTTCCCAAAGAGTTT
hAlb R primer	qPCR for VCN	CTCTCCTTCTCAGAAAGTGTGCATAT
hAlb probe	qPCR for VCN	VIC-TGCTGAAACATTCACCTTCCATGCAGA-TAMRA
iPCR F primer	Inverse PCR for TNS9.3/fNG	CGTTAACGATCTTAGCCACTTTT
iPCR R primer	Inverse PCR for TNS9.3/fNG	CTTGTACAGCTCGTCCATGC
iPCR -fSV2A F primer	Inverse PCR for fSV2A	GGAACCCACTGCTTAAGCCTCAATAAAGCTTG
iPCR -fSV2A R primer	Inverse PCR for fSV2A	CAAAGGGAGATCCGACTCGTCTGAGGG
LAM-LTR1-bio F primer	LAM-PCR	GAGCTCTCTGGCTAACTAGG
LAM-LTR2-bio F primer	LAM-PCR	AGCTTGCCTTGAGTGCTTCA
LAM-LTR3-bio F primer	LAM-PCR	AGTAGTGTGTGCCCGTCTGT
LAM-linker-univ	LAM-PCR	GACCCGGGAGATCTGAATTCAGTGGCACAGCAGTTAGG
LAM-linker-Tsp509I	LAM-PCR	AATTCCTAACTGCTGTGCCACTGAATTCAGATC
LAM-LC-1 R primer	LAM-PCR	GACCCGGGAGATCTGAATTC
LAM-LC-2 R primer	LAM-PCR	AGTGGCACAGCAGTTAGG
GV F primer	Integration site-specific PCR	GCCCGGGTTATAATTACCTCA
5.10 IS1 R primer	Integration site-specific PCR	GCTCGTCAAAGGGAAGTTTCT
5.10 IS2 R primer	Integration site-specific PCR	TTCACATTGAAAGCCAGACG
5.10 IS3 R primer	Integration site-specific PCR	AGGCTGTGCTAATTCCATCC
5.10 IS4 R primer	Integration site-specific PCR	CCTCAGCATCCCAAAGTGTT
5.10 IS5 R primer	Integration site-specific PCR	TGCCTGTAATCCCAGCTACC
5.10 IS6 R primer	Integration site-specific PCR	GCTTTCCGTCCAGTGAAAAA
PE-alpha primer	Primer Extension	TTGTCGGCAGGAGACAGCACCAT
PE-beta primer	Primer Extension	AGACTTCTCCTCAGGAGTCAG
Alpha-globin F primer	qRT-PCR	CTGGCGAGTATGGTGCG
Alpha-globin R primer	qRT-PCR	GAAGTGCGGGAAGTAGGTC
Alpha-globin probe	qRT-PCR	VIC-ATGTTCCTGTCCTTCCCCACCAC-TAMRA
Beta-globin F primer	qRT-PCR	TGCACGTGGATCCTGAGAACT
Beta-globin R primer	qRT-PCR	AATTCTTTGCCAAAGTGATGGG
Beta-globin probe	qRT-PCR	FAM-CAGCACGTTGCCCAGGAGCCTG-TAMRA

Supplementary Table 9. List of oligonucleotides used for molecular analyses of iPS cell characterization, globin vector copy number, integration and expression.