SUPPLEMENTAL INFORMATION

Figure S1 (Related to Figure 1) (A) Summary diagram of the sorting scheme used in this study (B) qPCR graph showing relative mRNA levels of β-, α-, δ-, duct and acinar cell markers in each sorted fraction obtained from a 6-year-old donor (normalized to presort values). (C) FACS plots show distribution of juvenile islet cell populations using HPa2, HPi2 antibody combination. No specific sub population can be identified. (D) Cluster dendogram displaying Pearson correlation results between adult samples of this study and Morán et al., 2012.

Figure S2 (Related to Figure 2) (A) Volcano plot shows assigned raw *P*-values and fold change (juvenile versus adult) of genes detected in exocrine cells (dots) after DE-Seq analysis. Lines indicate differentially expressed genes with FDR 0.2 cutoff. (B) Box plots show secreted insulin levels of juvenile (n=3) and adult (n=10) islets exposed to 5.6 mM glucose. (C) Bar graphs represent area under the curve values of secreted insulin from juvenile (n=9) and adult (n=16) islets post 16.7 mmol/L glucose stimulation (* t-test $P<0.05$). Error bars indicate S.E.

Figure S3 (Related to Figure 3) Supplemental histone ChIP-Seq aggregate plots corresponding to additional multiple donor samples. Histone ChIP-Seq signals obtained from (A-C) 66 year-old adult donor, (D-F) 0.8, 1.4 and 5 year-old juvenile donors (*P* values obtained by Wilcoxon rank sum test comparing aggregated ChIP-Seq signal around the TSS of juvenile-upregulated and adult-upregulated genes).

Figure S4 (Related to Figure 5) (A) SIX2 or SIX3 antibody staining of EndoC-βH1 cells misexpressing these factors shows that the antibodies are specific and do not cross react. (B) Baseline expression of *PDX1*, *SIX2* and *SIX3* in EndoC-βH1 cells. Insulin content (C) and

secreted insulin levels (D) of EndoC-βH1 cells expressing *GFP, SIX2* or *SIX3.* Asterisks indicate statistically significant results reproduced in multiple experimental replicates ($P < 0.05$, two-way ANOVA followed by Fisher's Least Significant Difference test) (E) Relative expression levels of *PDX1, NKX6.1,* and *MAFA* in EndoC-βH1 cells misexpressing GFP, SIX2 or SIX3. (F) qPCR results of T-Antigen knock down experiments in EndoC-βH1 cells using three different siRNA duplexes (G) Immunohistology for KI67 shows reduced staining in EndoC-βH1 cells transfected with different siRNA duplexes to knockdown T-Antigen. Error bars indicate S.D.

Figure S5 (Related to Figure 5) (A) Heat map shows $\log_2$ fold changes in expression of age-dependent genes in human islet α- and β-cells, SIX3^{pos} EndoC-βH1^{TKD} or SIX2^{pos} EndoC-βH1^{TKD} cells. (B) GO Term enrichment analysis of genes differentially expressed in SIX3^{pos} EndoC-βH1^{TKD} cells. Top scoring biological process and molecular function terms are graphed with significance (*P*-value) on a negative $\log_{10}$ scale.

Table S1 (Related to Figure 1) List of RNA-Seq libraries sequenced in this study.

Table S2 (Related to Figure 2) List of age-dependent genes that are differentially expressed in α- and β-cells.

Table S3 (Related to Figure 2) List of age-dependent genes that are differentially expressed in duct and acinar cells.

Table S4 (Related to Figure 3) List of histone ChIP-Seq libraries sequenced in this study.

Table S5 (Related to Figure 5) List of genes differentially expressed in SIX3^{pos} EndoC-βH1^{TKD} cells compared to GFP EndoC-βH1^{TKD} control cells.

Supplemental Experimental Procedures

Flow Cytometry

Islet or acinar tissue was dispersed into single cells by enzymatic digestion using Accutase (Life Technologies) following manufacturer's protocol. Prior to antibody staining, cells were incubated with blocking solution containing FACS buffer (2% v/v fetal bovine serum in phosphate buffered saline) and goat IgG (Jackson Labs, 11.2 μg per million cells). LIVE/DEAD Fixable Aqua Dead Cell Dye (Life Technologies) was used as a viability marker. Cells were then stained with appropriate antibodies at 1:100 (v/v) final concentration. The following antibodies were used for FACS experiments: HPx1-Dylight 488 (Novus, NBP1-18951G), HPi2-Dylight 650 (Novus, NBP1-18946C), HPa2-Biotin (Novus, NBP1-18950B), CD26-PE (BioLegend, 302705), CD133/1 - Biotin (Miltenyi Biotec 130-090-664), CD133/2 - Biotin (Miltenyi Biotec 130-090-852), streptavidin-eFluor780 (eBioscience, 47-4317-82), streptavidin-APC (eBioscience, 17-4317-82), mouse anti-human Glucagon (Sigma, G 2654), guinea pig anti-human Insulin (DAKO, A0564), goat anti-guinea pig Dylight 550 (Abcam ab102385), donkey anti-mouse Alexa 488 (Molecular Probes, A-21202). All antibody incubation steps were performed on ice for 30 minutes. Intracellular hormone staining was performed as described (Hrvatin et al., 2014). Labeled cells were sorted on a special order 5-laser FACS Aria II (BD Biosciences) using a 100 μm nozzle, with appropriate compensation controls and doublet removal. Sorted cells were collected into low retention tubes containing 50 μL of FACS buffer. Cytometry data were analyzed and graphed using FlowJo software (TreeStar v.10.8).

Cloning of Lentivirus Expression Constructs And Virus Particle Generation

SIX2 (RC201328) and SIX3 (RC216286) full-length ORFs were obtained from Origene TrueORF collection (Origene Technologies) and cloned into pRRL lentivirus expression system (kindly provided by Dr. Gary Nolan, Stanford University, Stanford, California, USA) or pDUAL-

H2BmCherry vector using Gibson Assembly (NEB, 2611). After sequence verification, lentivirus particles were generated by co- transfection of 293T cells with pΔR8.2, pVSVG as well as the expression plasmid as described (Zufferey et al., 1997). Beginning at 24 hours post-transfection, the supernatants were collected approximately every 12 hours for three times and virus particles were precipitated using PEG-It solution following manufacturer's instructions (System Biosciences, LV810A-1).

Knockdown of T-Antigen in EndoC-βH1 Cells

RNA duplexes targeting T-Antigen were designed as described (Ihler et al., 2012). Cells were seeded onto coated 96-well plates at 50% density (Ravassard et al., 2011). On the day of transfection, siRNAs were delivered to the cells using Lipofectamine RNAiMAX Transfection reagent (Invitrogen, 13778075) at a final concentration of 90 nM of RNA duplex per well. siRNA delivery was repeated 3-4 days after the first transfection, and cells were analyzed at the end of 7 days.

GSIS Assays using EndoC-βH1 cells

Stable EndoC-βH1 lines were generated by transducing SIX2, SIX3 or control GFP lentivirus, with 100% efficiency as assessed by GFP fluorescence. Equal number of EndoC-βH1 cells expressing SIX2, SIX3 or GFP alone were seeded onto coated 96-well plates and cultured as described (Ravassard et al., 2011). The day before the GSIS assay, cells were incubated overnight in EndoC-βH1culture medium containing 1.8 mM glucose. The next day, stimulated insulin secretion was measured by static incubation of cells in appropriate buffer for one hour, and then collecting supernatants for subsequent ELISA analysis. For basal insulin measurement, cells were incubated in Krebs-Ringer bicarbonate buffer (KRBB) (115 mM NaCl, 5 mM KCl, 1.19

mM KH_2PO_4, 1 mM $CaCl_2$, 1 mM $MgCl_2$, 24 mM $NaHCO_3$, 10 mM HEPES, pH 7.4, and 0.2% BSA) for two hours before switching to KRBB containing 1.8 mM glucose. For high glucose the KRBB contained 16.7 mM glucose. To measure the insulin content, cells were lysed directly in plates using the acid:ethanol extraction protocol (0.18 M HCl in 96% ethanol [vol/vol], Mercodia Inc., Uppsala, Sweden). Insulin was quantified using a human insulin ELISA kit following manufacturer's protocol (10-1113-10, Mercodia Inc.). To control for cell number variability, insulin secretion and content measurements were adjusted to the DNA content, which was quantified using high-sensitive dsDNA Qubit assays (Q32851, Invitrogen).

RNA Isolation and RNA-Seq Assays

On average, $3x10^4$ cells were collected from each purified cell population. Sorted cells were pelleted by centrifugation at 650 *g*. After removal of the supernatant, cells were lysed in 100 μL of extraction buffer (PicoPure RNA isolation kit, Arctrus, Life Technologies). Total RNA was isolated following manufacturer's protocol, with the on-column Dnase I treatment option (RNase-free DNase I, Qiagen, 79254). RNA quality and quantity was measured using Bioanalyzer instrument (Agilent Technologies). RNA-Seq libraries were built using NuGen Ovation RNA-Seq v.2 (NuGen, 7102-08) and NEBNext library prep kit (NEB, E6240). Barcoded libraries were multiplexed and sequenced as paired-end 101bp reads on Illumina HiSeq2000 platform. To prepare RNA-Seq libraries of $SIX2^{pos}$ or $SIX3^{pos}$ EndoC-βH1^{TKD} cells, total RNA was isolated from approximately 1 x 10^6 cells using Trizol RNA extraction method (Life Technologies). The sequencing libraries were prepared using the KAPA mRNA-Seq kit (KAPA Biosystems, KK8400). Barcoded libraries were multiplexed and sequenced as single end 75 bp reads on the Illumina NextSeq sequencer.

Bioinformatic and Statistical Analysis of RNA-Seq Datasets

Sequencing reads were aligned to the Human Genome Assembly hg19 using STAR algorithm (Dobin et al., 2012) with default parameters (see Table S1 for mapping information). Transcript counts were obtained using a collapsed Ref-Seq reference transcriptome list available from the UCSC genome browser. After getting the transcript counts, differential gene expression calls were done using the DE-Seq R package (Anders and Huber, 2010). Genes that were considered as gender specific, based on the likelihood ratio test and had false discovery rate (FDR) adjusted p-values < 0.1 or on chromosome Y, were removed from the dataset. FDR was computed by p.adjust function of R using the Benjamini and Hochberg method. To identify genes of differential expression with age in endocrine cells, raw counts of genes with interquartile range > 2, and factors of interest were passed to DE-Seq to estimate the size factor of each sample for normalization and the dispersion of each gene. The factors of interest were age group and cell type (four levels consisted of alpha, beta, glucagon and insulin). Full and reduced negative binomial generalized linear models were fitted to evaluate the significance of age effect and calculate *P*-values. Column-wise, sample, quartiles of the normalized counts were used to subset genes into 3 groups: less than Q2, between Q2 and Q3, and greater than Q3. Within each of the three groups, genes with FDR < 0.2 in any of the groups were considered as differentially expressed in age. Genes with sum of mean adult and mean juvenile normalized counts less than 100 were filtered out. To cluster differentially expressed genes, normalized counts were shifted by +1, log base 2 transformed and mean centered. Hierarchical clustering of genes and Pearson correlation coefficients, from the cor() function of R, were done with the average linkage method of Cluster 3.0 software (de Hoon et al., 2004) and visualized with Java TreeView (Saldanha, 2004). The same procedure was applied in testing for age effect within exocrine, acinar and duct cells.

GO Term Analysis

GO Term analysis was performed using DAVID functional annotation tool version 6.7 as described (Huang et al., 2008) using the differentially expressed genes listed in Table S2, S3 or S5 as input lists (FDR 0.2).

Quantitative Reverse Transcription PCR (qRT-PCR)

cDNA was synthesized using total RNA extracted from sorted cells and the Superscript III first-strand synthesis system (Invitrogen, 18080051) according to the manufacturer's instructions. Quantitative PCR was performed using TaqMan assays and reagents from Applied Biosystems. The following TaqMan probes were used: *INS* (Hs00355773_m1), *GCG* (Hs00174967_m1), *SST* (Hs00356144_m1) *KRT19* (Hs00761767_s1), *CPA1* (Hs00156992_m1), *MAFA* (Hs01651425_s1), MKI67 (Hs01032443_m1), *PDX1* (Hs00236830_m1), *NKX6.1* (Hs00232355_m1), *SIX2* (Hs00232731_m1), *SIX3* (Hs00193667_m1). *ACTB* (Hs01060665_g1) was used as endogenous control. Relative mRNA abundance was calculated by comparative $\Delta\Delta C_T$ method using Applied Biosystems software.

ChIP-Seq Assays and Bioinformatic Analysis

Previously described methods (Rada-Iglesias et al., 2011; Wapinski et al., 2013) were adjusted as detailed below to perform ChIP-Seq assays from limited number of cells. On average, $2.5x10^5$ $HPi2^{pos}$ $HPx1^{neg}$ FACS purified cells were used for each ChIP-Seq experiment. Sorted cells were pelleted and washed with PBS, and were immediately crosslinked using freshly prepared 1% formaldehyde (v/v in PBS, Thermo #28908) for 10 minutes at room temperature. Crosslinked cells were then placed on ice and lysed in lysis buffer containing 10 mM Tris-HCl (pH 8.1), 1 mM EDTA (pH 8.0), 150 mM NaCl, 1 mM EGTA (pH 8.0), 0.1% sodium deoxycholate, 0.5% sodium laurylsarcosine, phenylmethanesulfonyl fluoride (Sigma) and protease inhibitor cocktails (Roche). Crosslinked chromatin was sheared to ~300 bp sized

fragments using a Misonix (S-4000) probe sonicator with 30% duty cycle. The lysate was centrifuged at 13,000 *g* to pellet the insoluble chromatin. The lysis buffer was neutralized with 1% Triton-X (Sigma). Sheared chromatin was split into three equal volumes and incubated overnight at 4ºC with 1 μg of the following antibodies: Rabbit H3K27ac (Abcam, ab4729), rabbit H3K4me3 (Abcam, ab8580), rabbit H3K27me3 (Diagenode, pAb-069-010). The next day chromatin-antibody complexes were captured using magnetic beads (Dynal, Invitrogen). Following washing steps, immunoprecipitated DNA was recovered by reversing crosslinks in elution buffer (50 mM Tris-Cl pH 8.0, 10 mM EDTA, 1% SDS) overnight at 65ºC. The DNA was treated with Rnase A (Roche) and purified using the phenol:chloroform:isoamylalcohol extraction method (Invitrogen). Sequencing libraries were prepared using NEBNext library prep kit (NEB, E6240). Barcoded ChIP-Seq and input libraries were multiplexed and sequenced as single-end 50 bp reads on Illumina HiSeq2000 platform. Sequencing reads were mapped to the human reference genome hg19 using Bowtie (Langmead et al., 2009) with default parameters. A custom script was used to calculate ChIP-Seq read counts 2 kb upstream and downstream of transcription start site of age-dependent genes. The TSS coordinates were downloaded from the UCSC genome browser using RefSeq reference transcriptome list.

Immunohistochemistry

Fresh human pancreata were fixed in 4% paraformaldehyde in PBS for overnight at 4ºC, and cryopreserved. 10 μm-thick frozen sections were cut and stained following standard cryo-staining protocols. Briefly, sections were washed in PBS, incubated with Streptavidin/Biotin blocking solution (Vector Labs, SP-2002) at room temperature, followed by incubation in permeabilization/blocking buffer (1% bovine serum albumin, 0.2% non-fat milk, 0.5% Triton-X in PBS) for one hour. Primary antibodies were mixed with permeabilization/blocking buffer at appropriate concentrations and slides were incubated at 4ºC overnight. The following antibodies

were used for immunohistochemistry experiments: Guinea pig anti-human Insulin, 1:1000 dilution (DAKO, A0564), guinea pig anti-human Glucagon, 1:1000 dilution (Takara, M182), rabbit anti-Ki67, 1:100 dilution (Novocastra/Leica, NCL-ki67p), rabbit anti-SIX2, 1:500 dilution (ProteinTech 11562-1-AP), rabbit anti-SIX3, 1:500 dilution (LifeSpan Biosciences, LS-B9336), goat anti-PDX1, 1:200 dilution (R&D systems, AF2419). Slides were washed with PBS, incubated with secondary antibodies at room temperature for two hours. After final wash with PBS, slides were preserved with mounting medium containing DAPI (Vector Labs, Vectashield H-1200). Images were obtained using Leica SP8 confocal microscope. Proliferating islet cells were quantified by counting $KI67^{pos}$ and $GLUCAGON^{pos}$ α-cells or $KI67^{pos}$ and $INSULIN^{pos}$ β-cells using pancreas sections from juvenile and adult donors. At least 50 islets per group were counted.

Supplemental References

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2012). STAR: ultrafast universal RNA-seq aligner. Bioinformatics bts635.

Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2008). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. *4*, 44–57.

Ihler, F., Vetter, E.V., Pan, J., Kammerer, R., Debey-Pascher, S., Schultze, J.L., Zimmermann, W., and Enders, G. (2012). Expression of a neuroendocrine gene signature in gastric tumor cells from CEA 424-SV40 large T antigen-transgenic mice depends on SV40 large T antigen. PloS One *7*, e29846.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25.

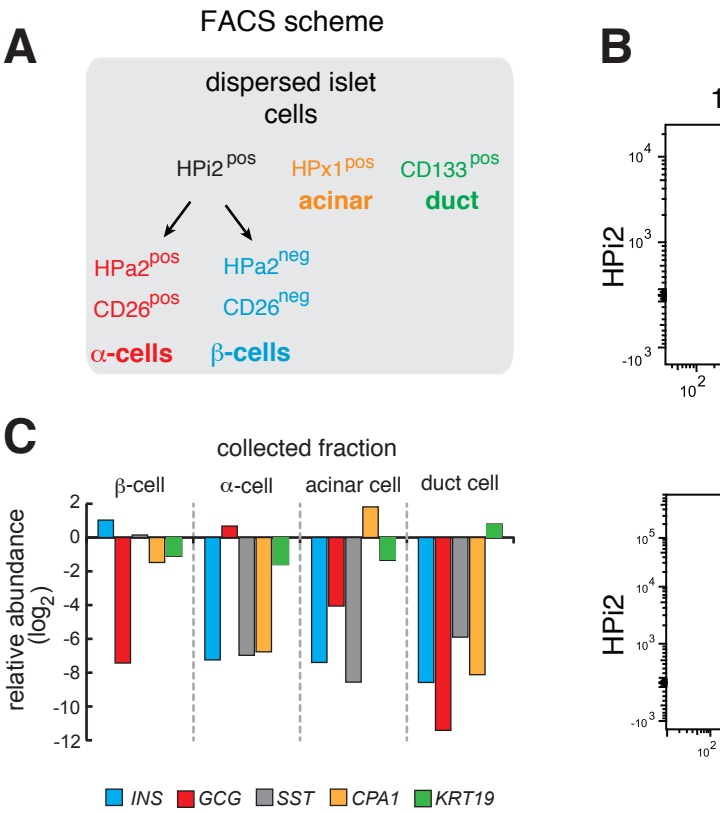
Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. Nature *470*, 279–283.

Saldanha, A.J. (2004). Java Treeview--extensible visualization of microarray data. Bioinforma. Oxf. Engl. *20*, 3246–3248.

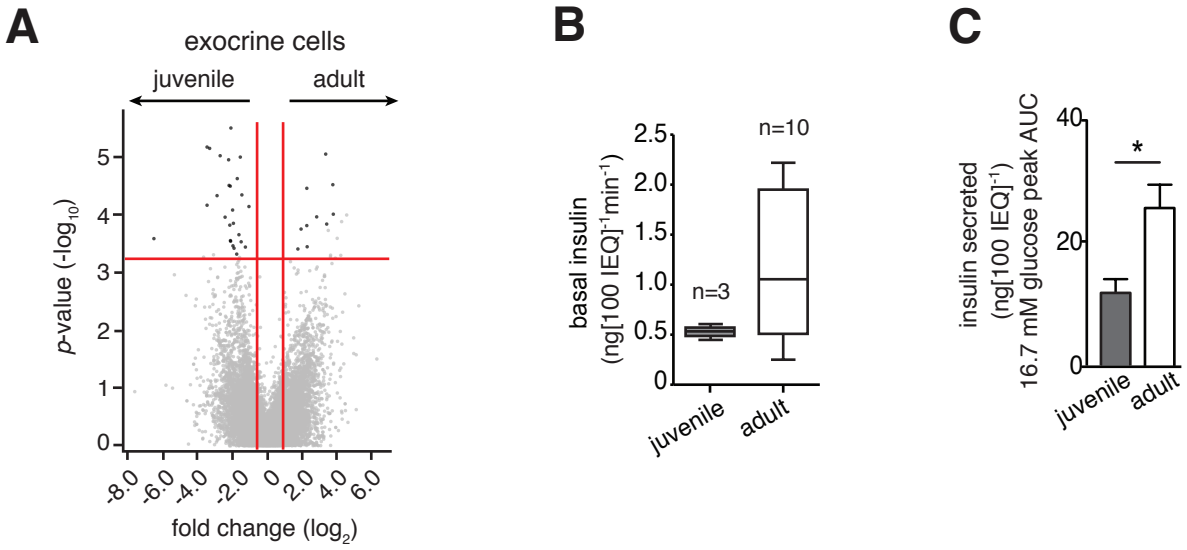
Wapinski, O.L., Vierbuchen, T., Qu, K., Lee, Q.Y., Chanda, S., Fuentes, D.R., Giresi, P.G., Ng, Y.H., Marro, S., Neff, N.F., et al. (2013). Hierarchical mechanisms for direct reprogramming of fibroblasts to neurons. Cell *155*, 621–635.

Zufferey, R., Nagy, D., Mandel, R.J., Naldini, L., and Trono, D. (1997). Multiply attenuated lentiviral vector achieves efficient gene delivery in vivo. Nat. Biotechnol. *15*, 871–875.

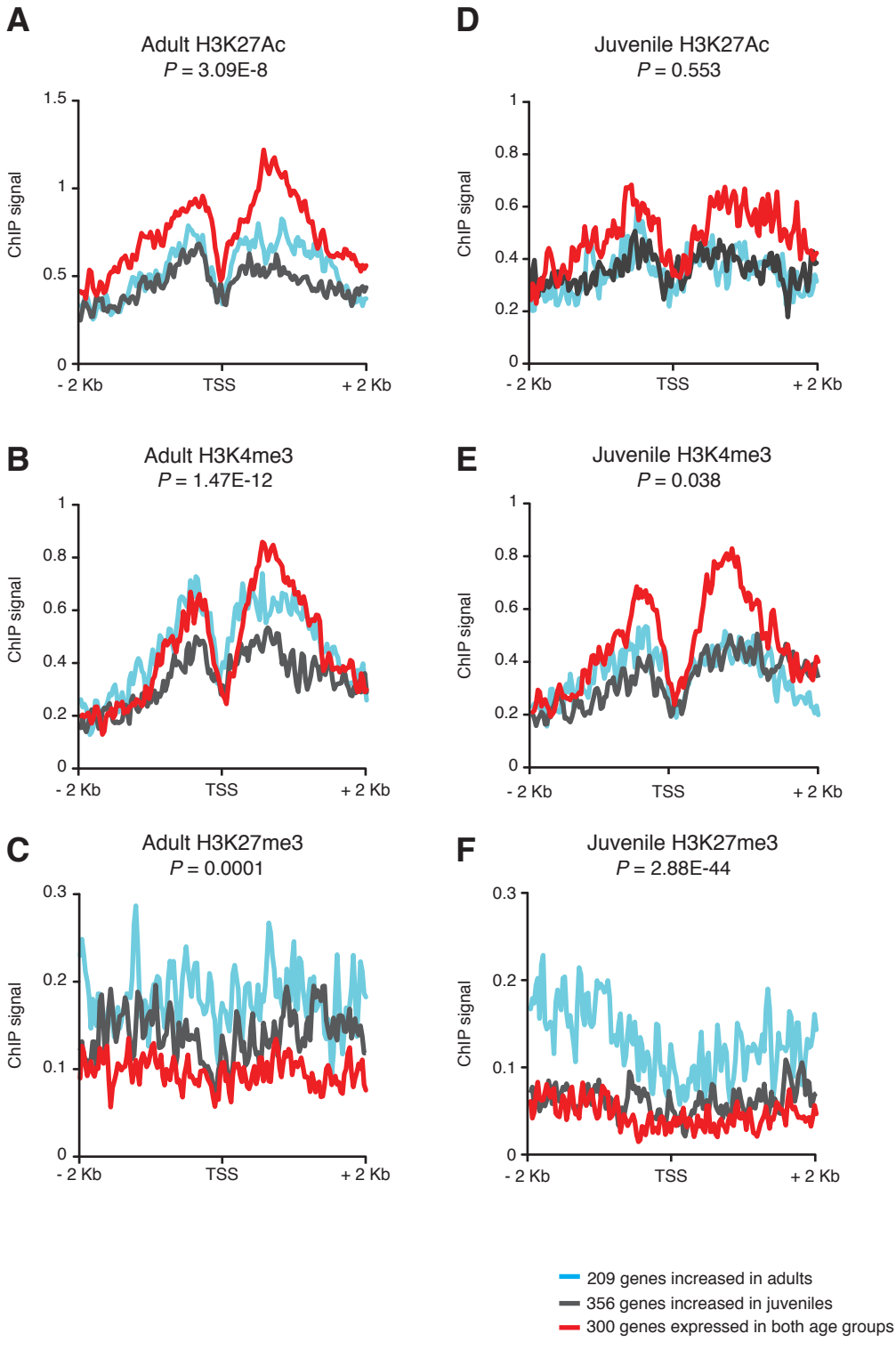
Arda et al, Figure S1



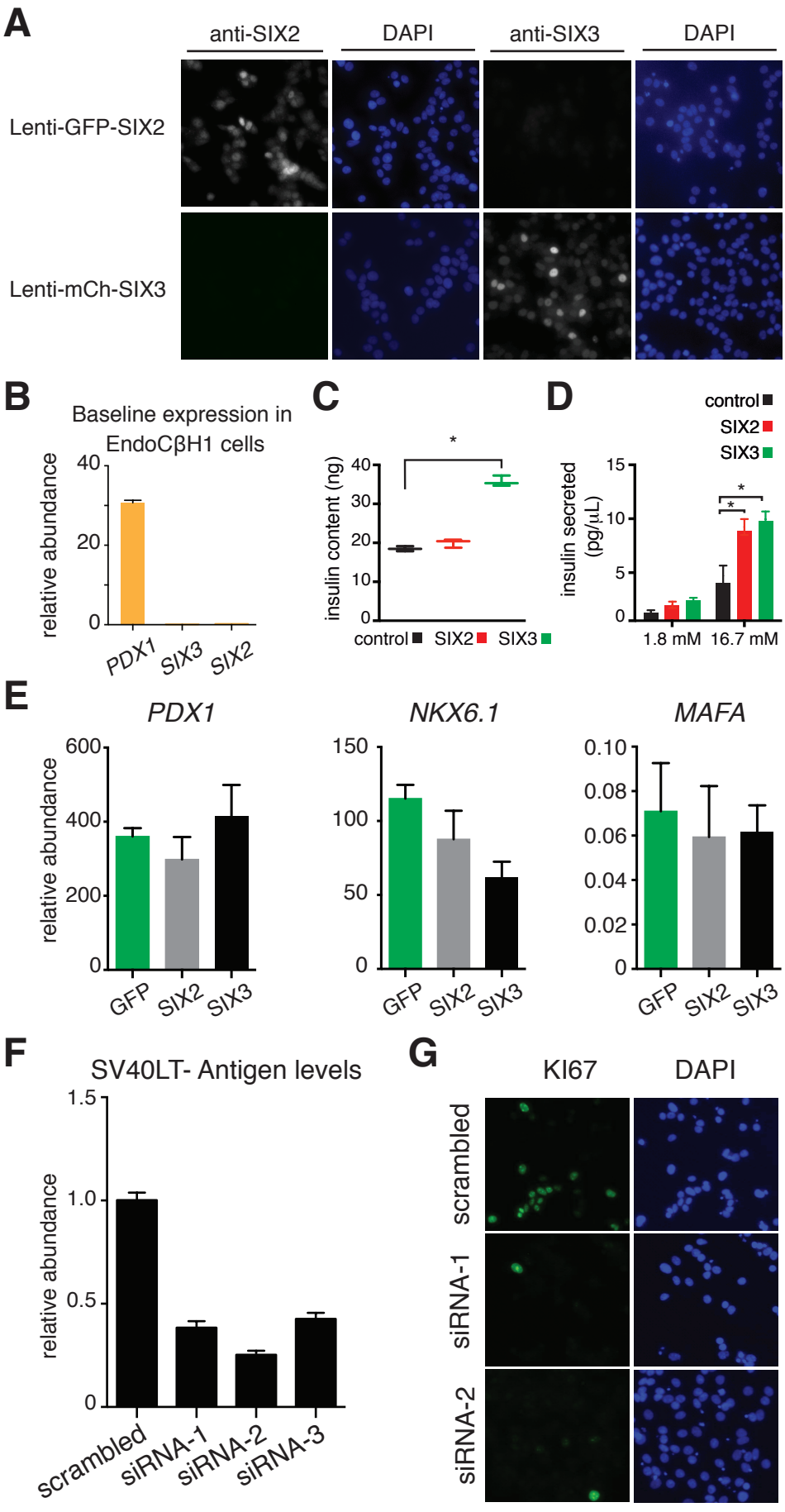
Arda et al, Figure S2



Arda et al, Figure S3



Arda et al, Figure S4



Arda et al, Revision 2, Figure S5

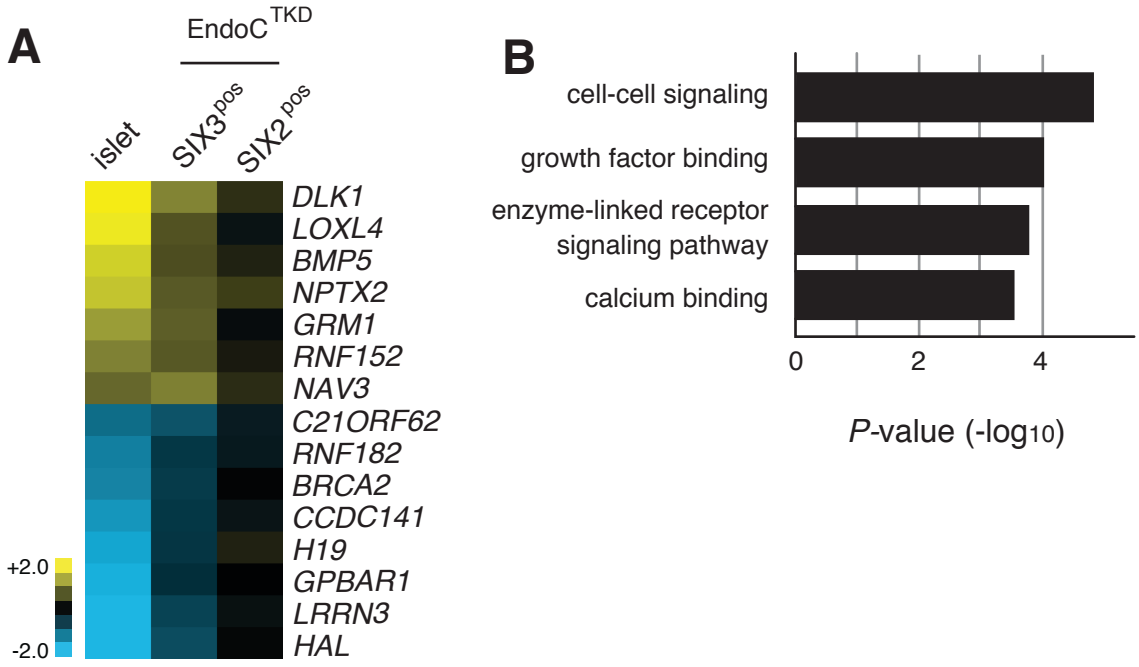


Table S1. List of RNA-Seq libraries sequenced in this study

RNA-Seq library	Donor Age	Sort	Enriched cell type	Number of uniquely mapped reads
RSq02	1.5	$HPx1^{pos}$ $HPi2^{neg}$	acinar	140,855,160
RSq03	33	$HPi2^{pos}$ $HPa2^{pos}$	alpha	135,448,352
RSq06	33	$HPx1^{pos}$ $HPi2^{neg}$	acinar	126,363,846
RSq09	40	$HPi2^{pos}$ $HPa2^{pos}$	alpha	177,446,818
RSq10	40	$HPi2^{pos}$ $HPa2^{neg}$	beta	156,206,766
RSq12	33	$HPi2^{pos}$ $HPa2^{neg}$	beta	187,959,114
RSq13	28	$HPi2^{pos}$ $HPa2^{pos}$	alpha	44,299,566
RSq14	28	$HPi2^{pos}$ $HPa2^{neg}$	beta	224,984,408
RSq15	33	$HPx1^{neg}$ $CD133^{pos}$	duct	112,648,006
RSq16	66	$HPi2^{pos}$ $HPa2^{pos}$ $CD26^{pos}$	alpha	213,237,528
RSq18	66	$HPi2^{pos}$ $HPa2^{neg}$ $CD26^{neg}$	beta	98,147,656
RSq24	5	$HPx1^{neg}$ $CD133^{pos}$	duct	167,134,810
RSq25	5	$HPi2^{pos}HPa2^{pos}$ $CD26^{pos}$	alpha	158,217,722
RSq26	5	$HPi2^{pos}$ $HPa2^{neg}$ $CD26^{neg}$	beta	168,028,674
RSq28	5	$HPx1^{pos}$ $HPi2^{neg}$	acinar	114,986,354
RSq29	48	$HPi2^{pos}$ $HPa2^{pos}$ $CD26^{pos}$	alpha	167,900,192
RSq31	48	$HPi2^{pos}$ $HPa2^{neg}$ $CD26^{neg}$	beta	143,647,714
RSq32	1.5	$HPx1^{neg}$ $CD133^{pos}$	duct	135,805,340
RSq33	54	$HPx1^{neg}$ $CD133^{pos}$	duct	135,085,684
RSq34	42	$HPx1^{pos}$ $HPi2^{neg}$	acinar	120,738,714
RSq35	5	$HPi2^{pos}$ $HPa2^{pos}$ $CD26^{pos}$	alpha	142,825,962
RSq37	5	$HPi2^{pos}$ $HPa2^{neg}$ $CD26^{neg}$	beta	131,284,904
RSq38	0.8	$HPi2^{pos}$ $HPa2^{pos}$ $CD26^{pos}$	alpha	65,029,350
RSq39	0.8	$HPi2^{pos}$ $HPa2^{neg}$ $CD26^{neg}$	beta	108,644,146
RSq40	1.4	$HPi2^{pos}$ $HPa2^{neg}$ $CD26^{neg}$	beta	123,746,376
RSq42	1.4	$HPi2^{pos}$ $HPa2^{pos}$ $CD26^{pos}$	alpha	81,215,588
RSq45	6	anti-INS (intracellular)	beta	31,476,878
RSq46	6	anti-GCG (intracellular)	alpha	32,010,992

Table S4. List of histone ChIP-Seq libraries sequenced in this study

ChIP-Seq library	Donor Age	Histone Mark	Number of uniquely mapped reads
EA25	48	H3K27ac	32,308,688
EA27	48	H3K4me3	15,720,693
EA28	48	H3K27me3	11,227,648
EA73	0.8	H3K27ac	28,772,012
EA45	0.5	H3K4me3	35,448,778
EA46	0.5	H3K27me3	30,120,960
EA13	66	H3K27ac	52,012,614
EA15	66	H3K4me3	25,296,036
EA16	66	H3K27me3	23,306,076
EA69	1.4	H3K27ac	27,065,899
EA51	5	H3K4me3	30,518,496
EA75	0.8	H3K27me3	30,154,281