Supplementary Materials for

Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants

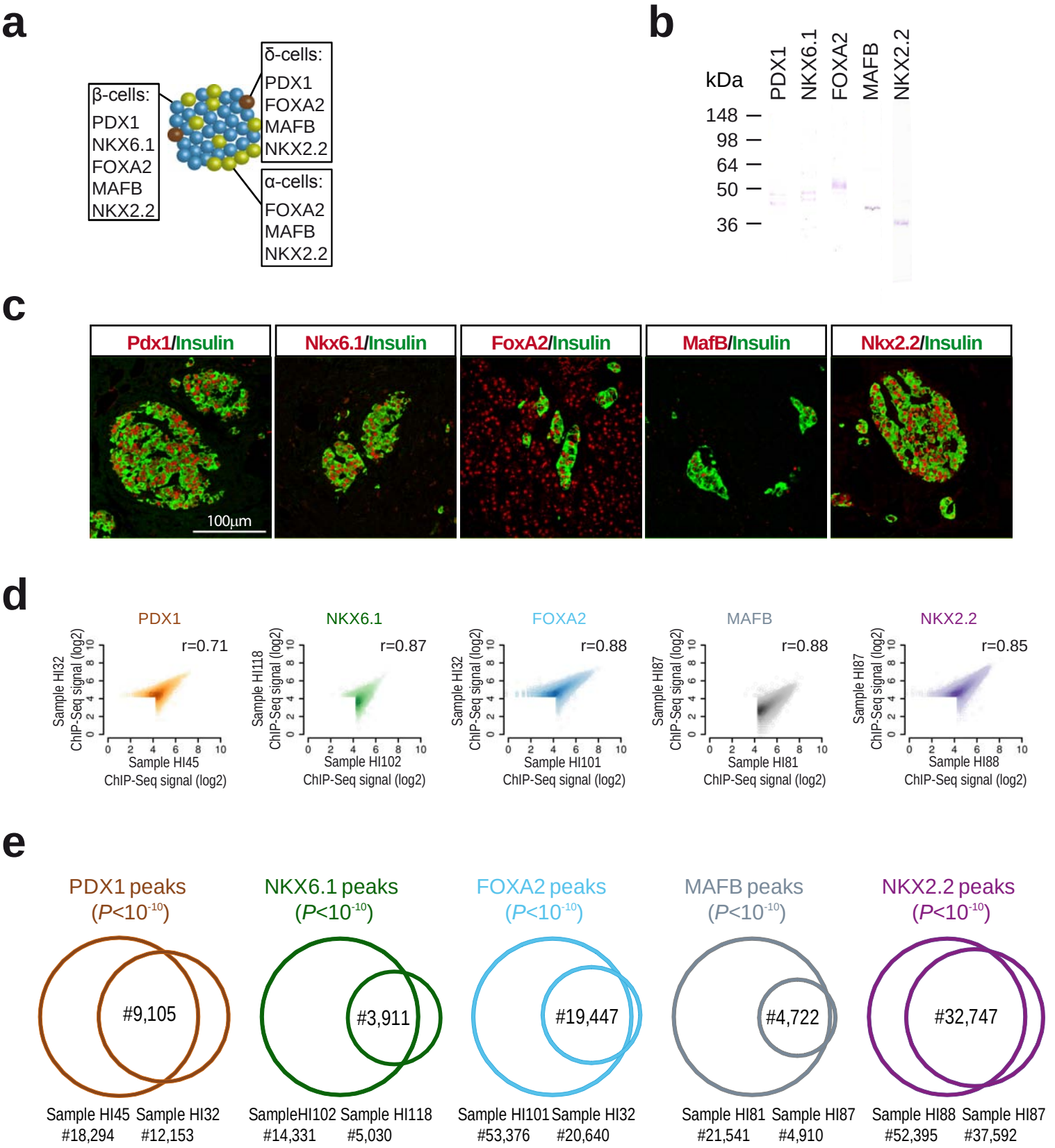
Lorenzo Pasquali[#], Kyle J. Gaulton[#], Santiago A. Rodríguez-Seguí[#], Loris Mularoni, Irene Miguel-Escalada, İldem Akerman, Juan J. Tena, Ignasi Morán, Carlos Gómez-Marín, Martijn van de Bunt, Joan Ponsa-Cobas, Natalia Castro, Takao Nammo, Inês Cebola, Javier García-Hurtado, Miguel Angel Maestro, François Pattou, Lorenzo Piemonti, Thierry Berney, Anna L. Gloyn, Philippe Ravassard, José Luis Gómez Skarmeta, Ferenc Müller, Mark I. McCarthy, Jorge Ferrer,*

[#] Authors contributed equally to this work

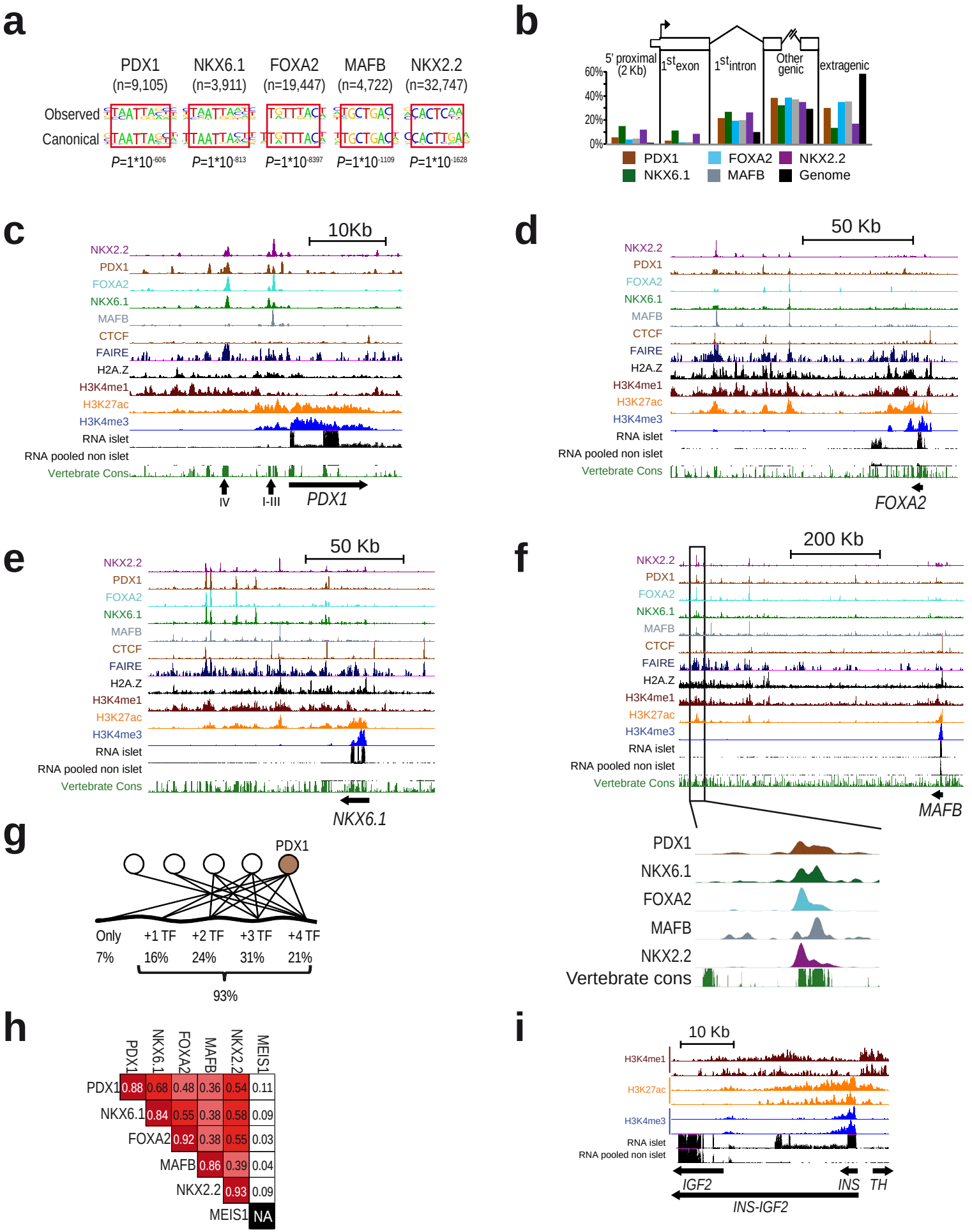
*Correspondence to: Jorge Ferrer. e-mail: j.ferrer@imperial.ac.uk

---

This PDF file includes:

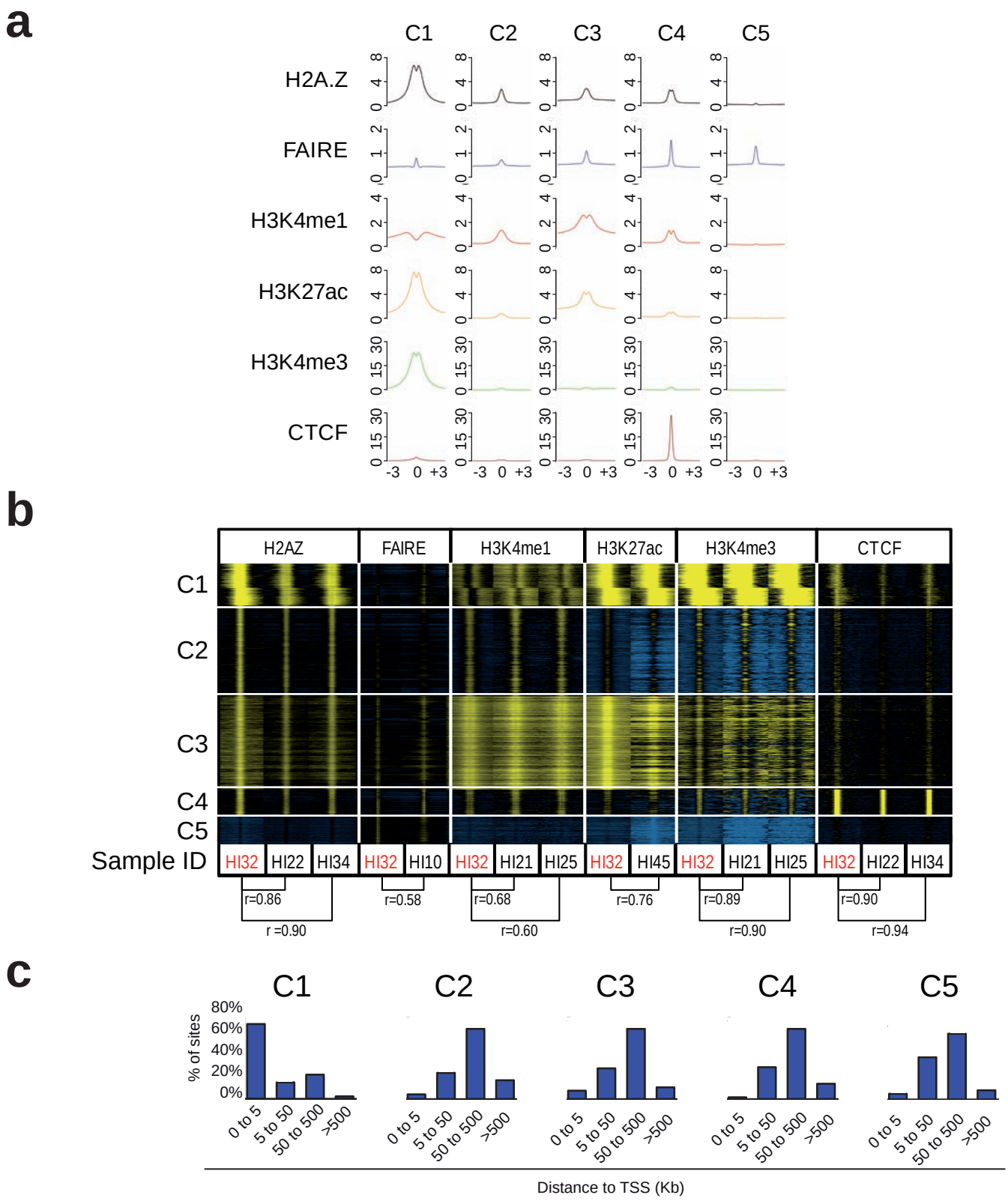


Supplementary Figure 1. Transcription factor ChIP-Seq signals are specific and consistent in biological replicates. (a) Schematic of transcription factor expression in three major pancreatic islet-cell types. **(b)** Assessment of antibody specificity by western blot using human islet nuclear extracts revealed expected migration patterns for all 5 transcription factors. As a control, western blots using the same antibodies but nuclear extracts obtained from Hela cells did not show specific bands (not shown). Molecular-mass markers (in kDa) are shown on the left. **(c)** Antibody assessment by immunofluorescence confocal microscopy of formaldehyde-fixed human pancreatic sections, showing expected islet-cell type nuclear localization of the tested transcription factors. FOXA2 was also found to localize to the pancreatic acinar tissue, as expected from its known expression profile. **(d)** Normalized ChIP-Seq signal correlation between two biological human islet replicates computed genome-wide over 1Kb bins. r values represent Pearson correlation coefficients. **(e)** Transcription factor binding peaks identified at a stringent threshold ($P < 10^{-10}$) in both biological replicates were retained for subsequent analysis.

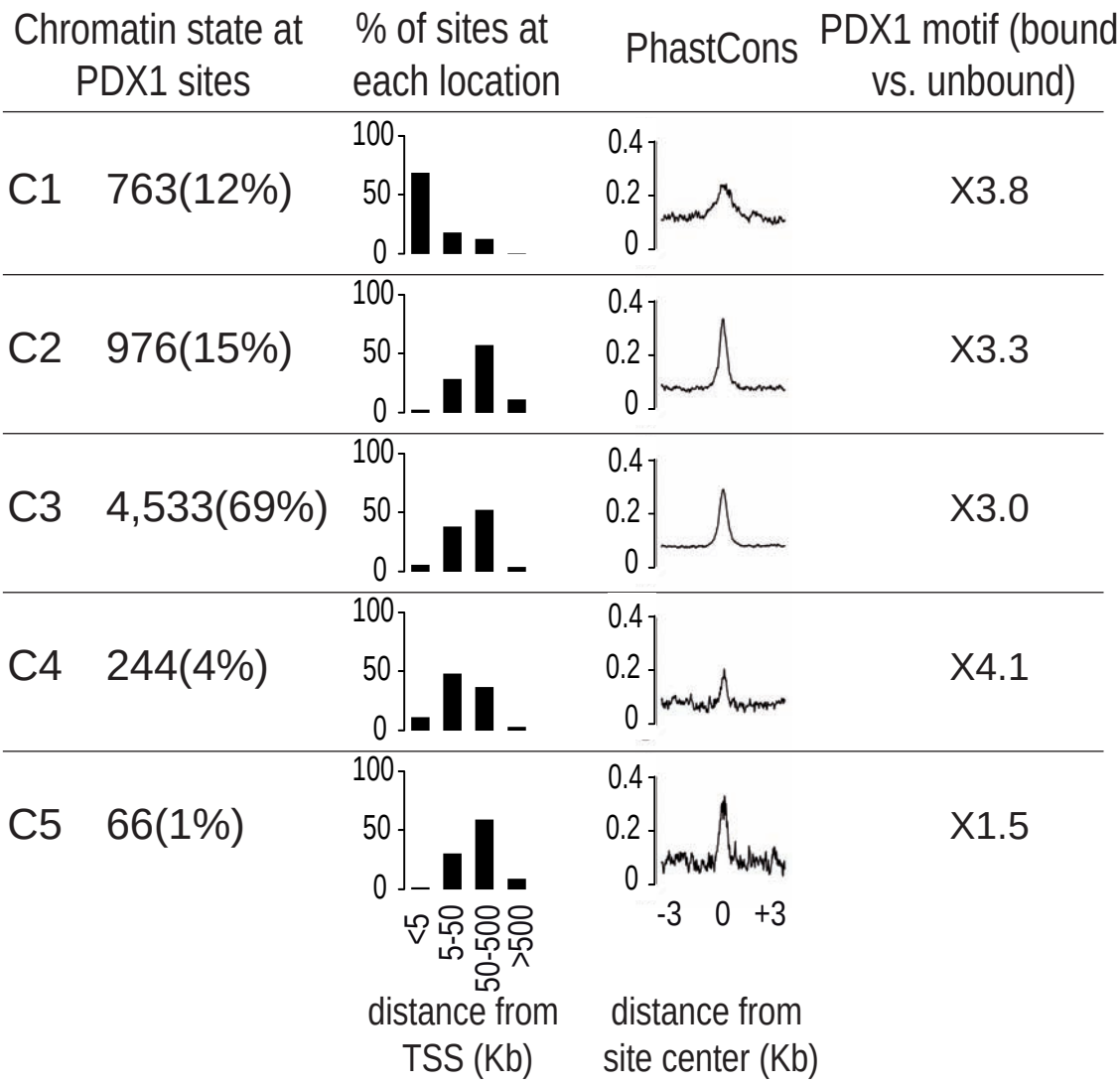


Supplementary Figure 2. A transcription factor network in human islet-cells. (a) Number of high-confidence transcription factor peaks and logo of the enriched sequence motifs compared with the *in vitro* reported motif of each

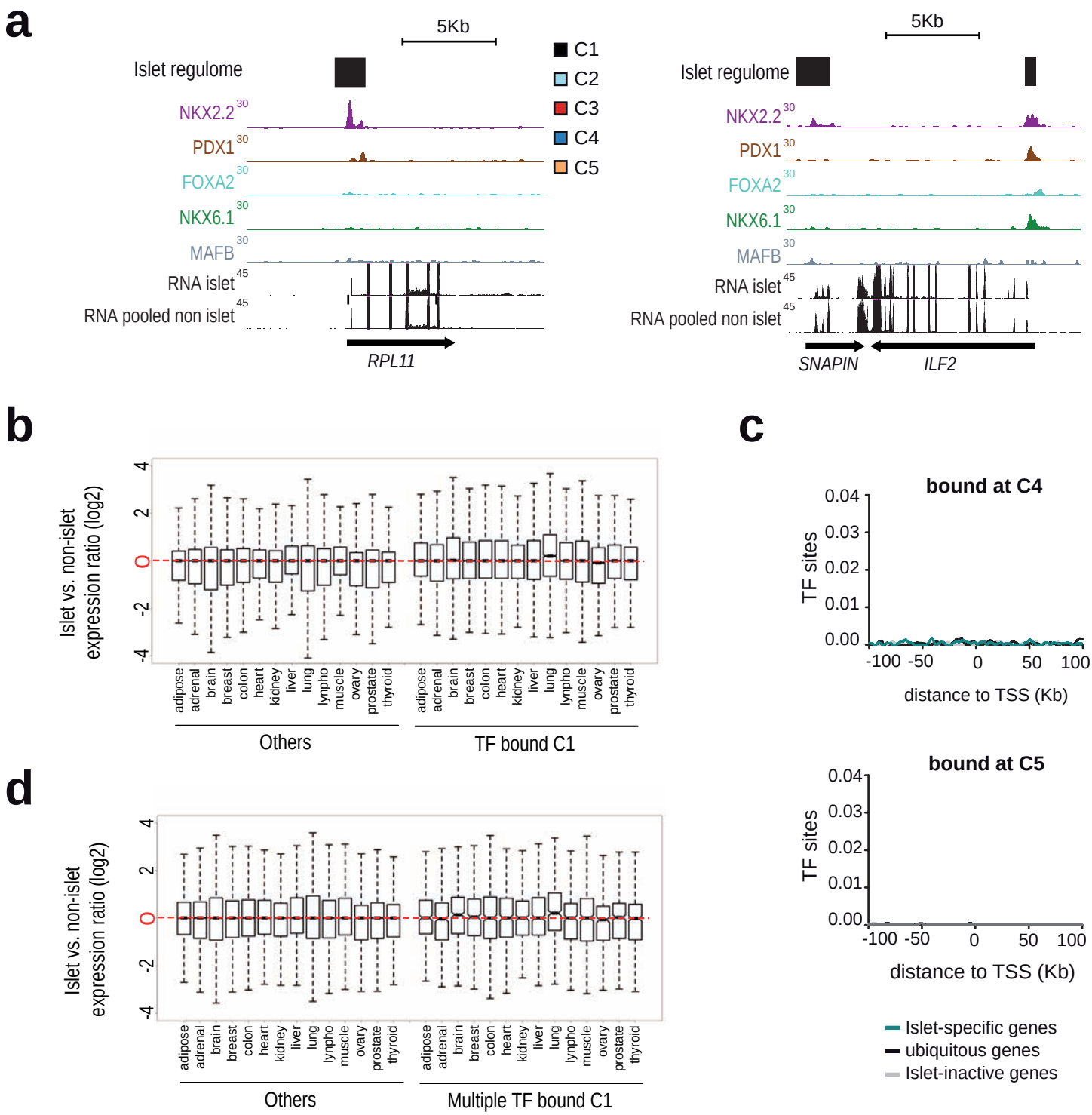
transcription factor, along with p-values for enrichment using HOMER[1]. As expected, most enriched transcription factor sequence motifs underlying bound sites included those previously reported *in vitro* for the same transcription factors. High-confidence transcription factor peaks were defined as those called at $P < 1x10^{-10}$ in replicate islet samples. **(b)** Transcription factor binding site distribution relative to gene annotations. Transcription factor binding sites were enriched near the 5' end of annotated genes, although in absolute terms most binding sites were more distant. The relative distribution of the different compartments in the entire genome is shown as black bars. **(c-f)** Examples of transcription factor binding patterns at loci harboring the *PDX1*, *FOXA2*, *NKX6-1*, and *MAFB* genes, showing frequent co-occupancy and clusters of binding sites of the five transcription factors at multiple sites near their own genes and near each other's genes. Arrows indicate transcription factor binding to known enhancers (Area I-IV) in *PDX1*. **(g)** Over 90% of PDX1 high-confidence binding sites show co-occupancy by at least one other factor in at least one replicate sample, illustrating that islet transcription factors very frequently bind to shared locations. Co-occupancy is defined by overlap of peaks by at least 1 bp. Comparable findings were encountered when we assessed co-occupancy relative to the binding sites of the four other transcription factors (not shown). **(h)** Transcription factor co-occupancy was also computed genome-wide by correlating binding signals between all transcription factor pairs, or between replicates for the same transcription factor. This analysis is restricted to sites bound by NKX6.1 and/or PDX1, because both are predominantly present in β-cells and therefore ensure that the correlation analysis is restricted to binding events occurring in the same cell type. Comparisons with ChIP-Seq data for MEIS1 in an umbilical cord blood cell line is shown as a reference. Numbers represent the Pearson's r correlation coefficient. NA: unavailable data. **(i)** Enrichment of histone modification marks (H3K4me3, H3K27ac and H3K4me1) at the *INS* (insulin) locus highlight the high purity and appropriate differentiated state of the human pancreatic islets used throughout this study.



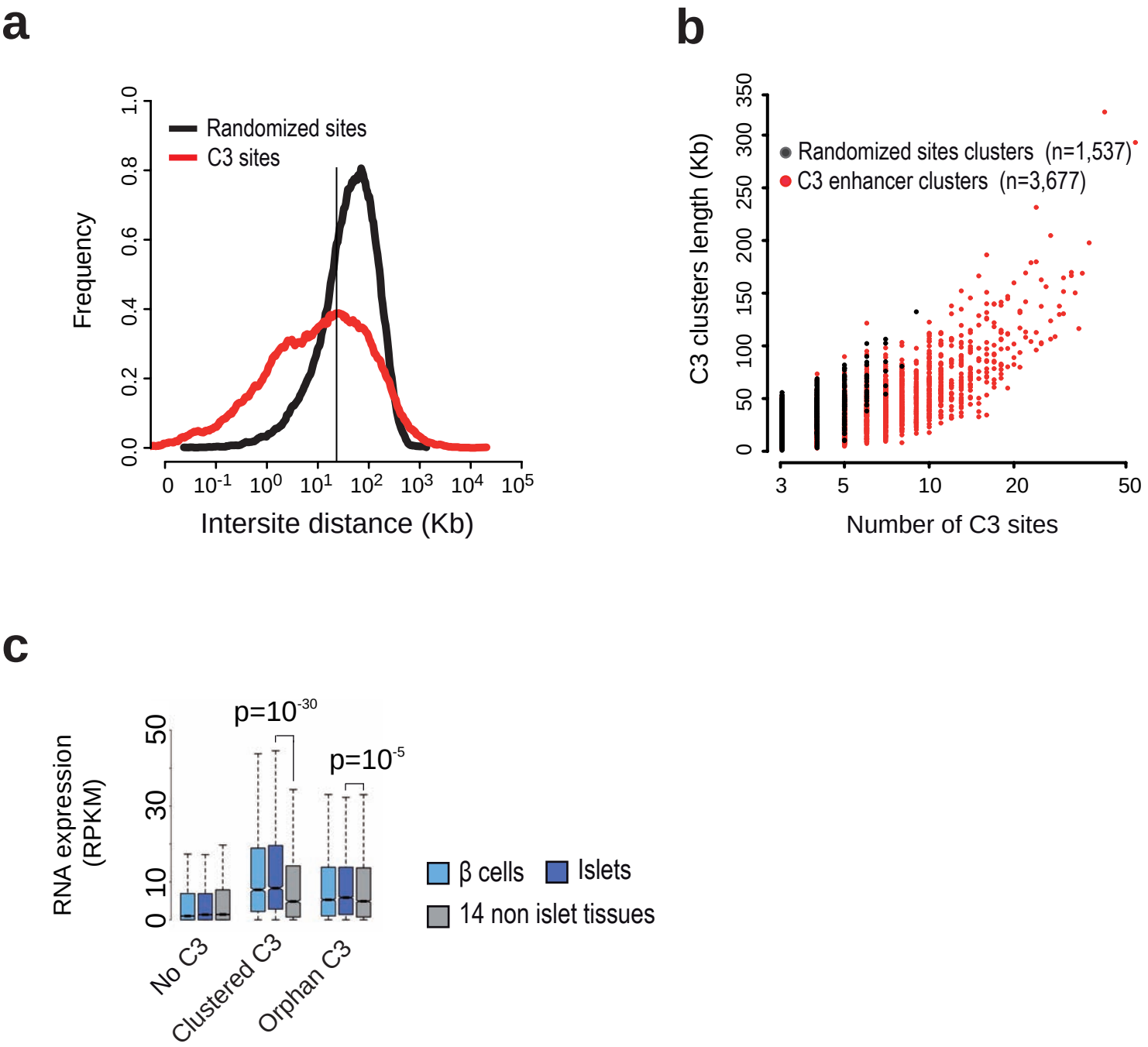
Supplementary Figure 3. Sample consistency and genomic location of accessible chromatin subclasses. (a) Read density for FAIRE, H2A.Z, different active histone modification marks and CTCF in 6 Kb windows centered on the 5 different classes of accessible chromatin defined by *K*-median clustering as shown in Fig. 2a. C1 sites showed strong H3K4me3 enrichment, C2 sites showed monomodal H3K4me1 enrichment without H3K4me3 or H3K27ac enrichment, C3 sites showed bimodal H3K4me1 enrichment without H3K4me3 and strong H3K27ac enrichment, C4 sites showed strong CTCF occupancy, and C5 sites lacked enriched active histone modifications or CTCF binding. This analysis was performed with data from sample HI32, for which all marks were available. **(b)** FAIRE, H2A.Z, histone modification signals, and CTCF signals were comparable in replicate human islet samples. Read densities in 6 Kb windows were centered on the 5 different classes of open chromatin defined by *K*-median clustering of sample HI32 as shown in Fig. 2a, and the read density in the same genomic sites is shown for replicate samples. Pearson's correlation coefficients for read densities between replicate samples in these sites are shown beneath. **(c)** Distribution relative to gene TSS of accessible chromatin classes shows that as expected C1 (promoter-like) accessible chromatin sites are preferentially located at the 5'end of annotated genes, unlike other accessible chromatin sites.



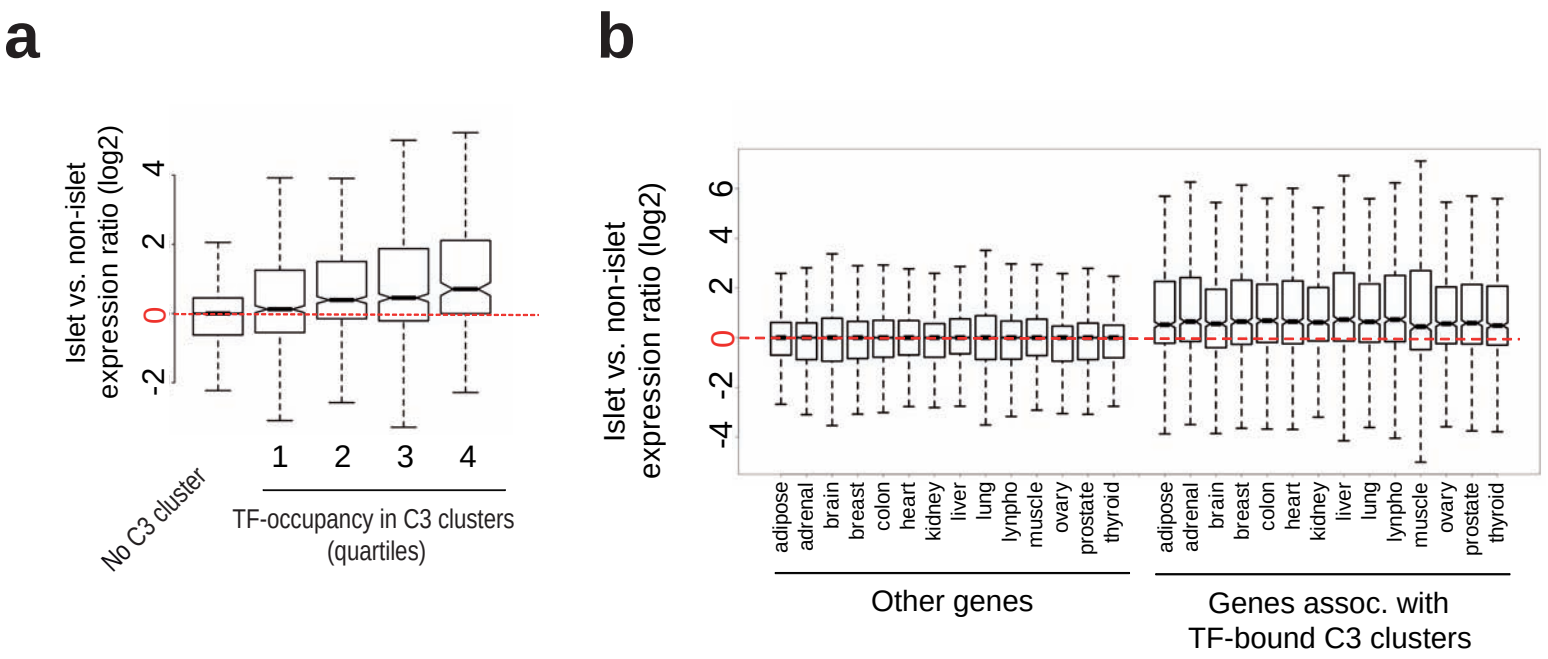
Supplementary Figure 4. Genomic features and chromatin state at PDX1 binding sites. A large fraction of PDX1 binding sites are associated with active enhancer chromatin, although PDX1 binds to all major classes of accessible chromatin, and in all cases shows a similarly high evolutionary sequence conservation or DNA-binding recognition motif enrichment. From left to right: Number and percentage of PDX1-bound sites that overlap with each accessible chromatin class, percentage of PDX1-bound sites that fall into each distance interval from the transcriptional start site (TSS), average PhastConst sequence conservation scores at -3 to +3 Kb relative to the binding site center, and fold enrichment of the PDX1 consensus recognition motif in PDX1-bound sites vs. non-bound sites of the same accessible chromatin class.

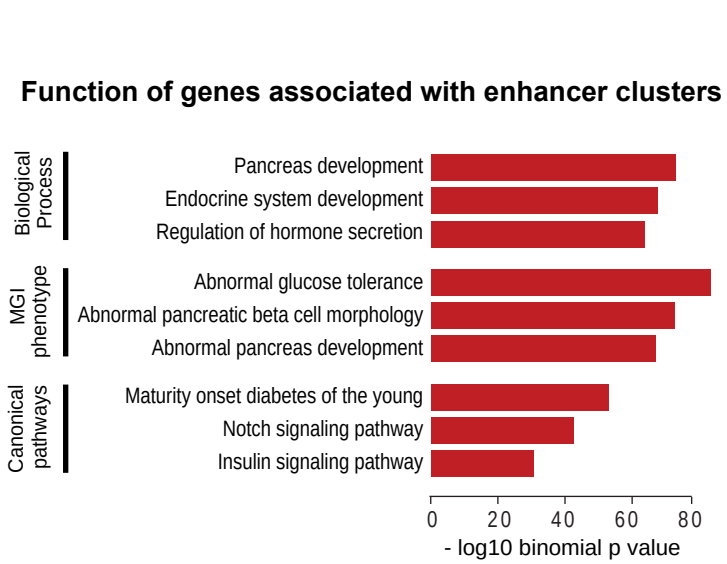


Supplementary Figure 5. Transcription factor binding at non-C3 accessible chromatin sites does not occur predominantly in islet-specific transcribed genes. (a) Examples of islet-specific transcription factors binding to the 5' flanking regions of two ubiquitously expressed genes. Expression in non-islet tissues is depicted as an average signal for all 14 tissues for simplicity, yet shows comparably high levels in all individual tissues. **(b)** Genes that are bound by one or more islet-specific transcription factor at their promoter (C1 chromatin), but lack active enhancer (C3) chromatin at <25Kb from TSS, tend to be transcriptionally active at a similar level in human islets and non-islet tissues. The boxes show Log2 ratios of quantile-normalized expression levels in islets vs. indicated non-islet tissues, and depict the interquartile range (IQR). Whiskers extend to either the maximum value or to 1.5 times the IQR, and notches indicate 95% confidence intervals of the median. **(c)** Density of C4 (CTCF-bound) and C5 (no active histone modifications) accessible chromatin sites bound by two or more transcription factors in the vicinity of the TSSs of 1,000 most islet-specific genes, ubiquitously active genes, or islet inactive genes. The results are depicted at the same scale as Fig 2c, and show that, unlike C1 and C3, transcription factor binding to C4 and C5 is not enriched in any of the three gene subsets. **(d)** Genes that are bound by three or more islet-specific transcription factor at promoter (C1) accessible chromatin, but lack active enhancers (C3) chromatin <25Kb from TSS, tend to be transcriptionally active at similar level in human islets and non islet tissues. The results are presented as described above for panel b. The statistical analysis of this data was performed with pooled non-islet data, and is shown in Fig 2g.



Supplementary Figure 6. Pancreatic islet enhancer clusters. (a) Distribution of distances between adjacent C3 sites, compared to adjacent randomized C3 sites. We calculated the distances between all adjacent C3 sites, and between adjacent randomized C3 sites taken from 1,000 iterations that used only the mappable genome of each individual chromosome separately, excluding sex chromosomes. The graph depicts the distribution of intersite distances for chromosome 1 as an example, with a vertical black line marking the 25th percentile of intersite distances for the randomized C3 sites. This threshold was used to create clusters, which were formed by three or more C3 sites that were maximally separated by that distance. **(b)** Size distribution of the 3,677 clusters of islet C3 sites, formed by 3-54 C3 sites each. The size range of the C3 clusters is 2 to 322 Kb (median 23 Kb). The clusters formed by randomized C3 sites are shown in black for comparison. **(c)** Gene expression in β-cells, islets, and 14 non-pancreatic tissues for genes that do not contain a C3 site, those that have a clustered C3 site, or an orphan C3 within < 25Kb from the TSS. The boxes show RNA expression levels in islets or the collection of non-islet tissues, expressed as the interquartile range (IQR). Whiskers extend to either the maximum value or to 1.5 times the IQR, and notches indicate 95% confidence intervals of the median. All P values resulted from a Wilcoxon Rank Sum test.





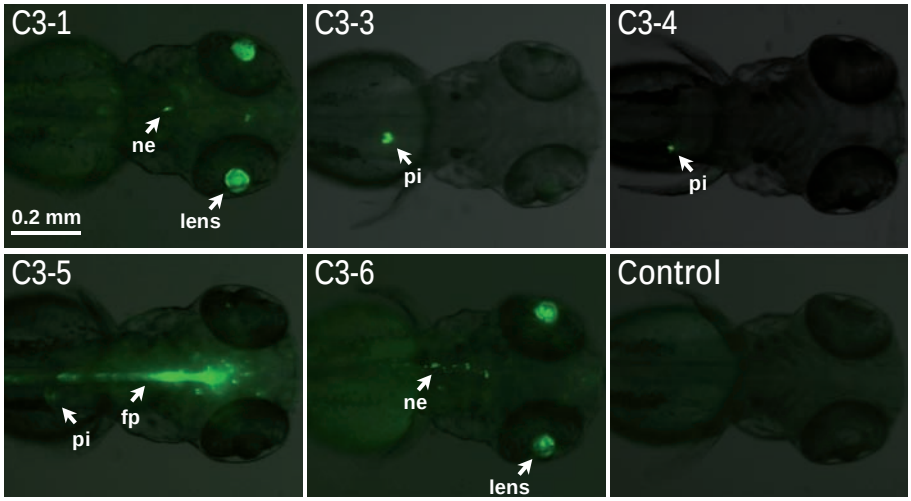
d

Examples of known β-cell genes associated with enhancer clusters:

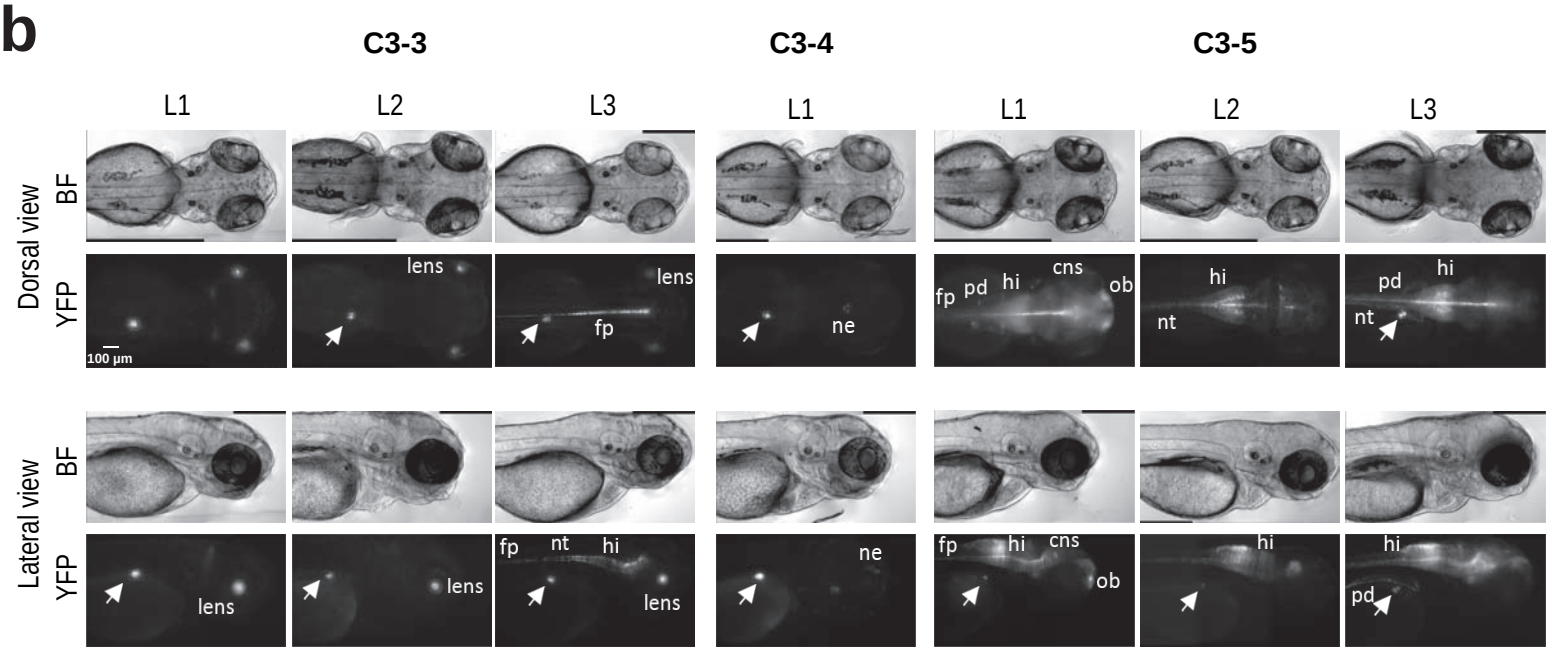
ABCC8	*GJD2*	*NKX2.2*	*SCG2*
CDKAL1	*GLIS3*	*NKX6.1*	*SCG5*
CHGA	*GLP1R*	*PAX6*	*SLC2A2*
CPE	*IAPP*	*PCSK1*	*SLC30A8*
DACH1	*INSM1*	*PDE3B*	*SOX4*
FFAR1	*ISL1*	*PDX1*	*ST18*
FOXA2	*KCNJ11*	*PTPRN*	*SULT4A1*
FOXO1	*MNX1*	*PTPRN2*	*SYT7*
G6PC2	*MYT1*	*RFX3*	*TCF7L2*
GAD2	*NCAM1*	*RFX6*	*UCN3*
GIPR	*NEUROD1*	*RGS4*	*WNT4*

Supplementary Figure 7. Pancreatic islet transcription factor bound enhancer clusters. (a) Enriched transcription in human islets is associated with clusters of enhancers that exhibit high transcription factor occupancy. We devised a transcription factor occupancy score that measures the extent to which the enhancers that form part of clusters are bound by multiple transcription factors (see methods), and used it to divide clusters in quartiles. Clusters were linked to a gene if they were located within < 25 Kb from the gene's TSS. The box plots show the expression ratio in islets vs. non-islet tissues for all genes in each category, expressed as the IQR. Whiskers extend to either the maximum value or to 1.5 times the IQR, and notches indicate 95% confidence intervals of the median. **(b)** Genes that were linked to clusters of enhancers that exhibit high transcription factor occupancy (clusters showing the upper 50% average transcription factor occupancy scores) show islet-enriched expression. The boxes show Log2 ratios of quantile-normalized expression levels in islets vs. indicated non-islet tissues, and depict the IQR. Whiskers extend to either the maximum value or to 1.5 times the IQR, and notches indicate 95% confidence intervals of the median. **(c)** Genes associated with clusters of enhancers that are highly bound by islet-specific transcription factors are enriched in typical β-cell functions. The analysis was performed by linking clusters to nearby genes with Genomic Regions Enrichment of Annotations Tool (GREAT) using default parameters[2], and the results illustrate the most significantly enriched functional annotation categories. **(d)** Most genes that have established roles in β-cell function and identity are associated with islet transcription factor bound enhancer clusters. A complete list with references is shown in Supplementary Table 2.

a



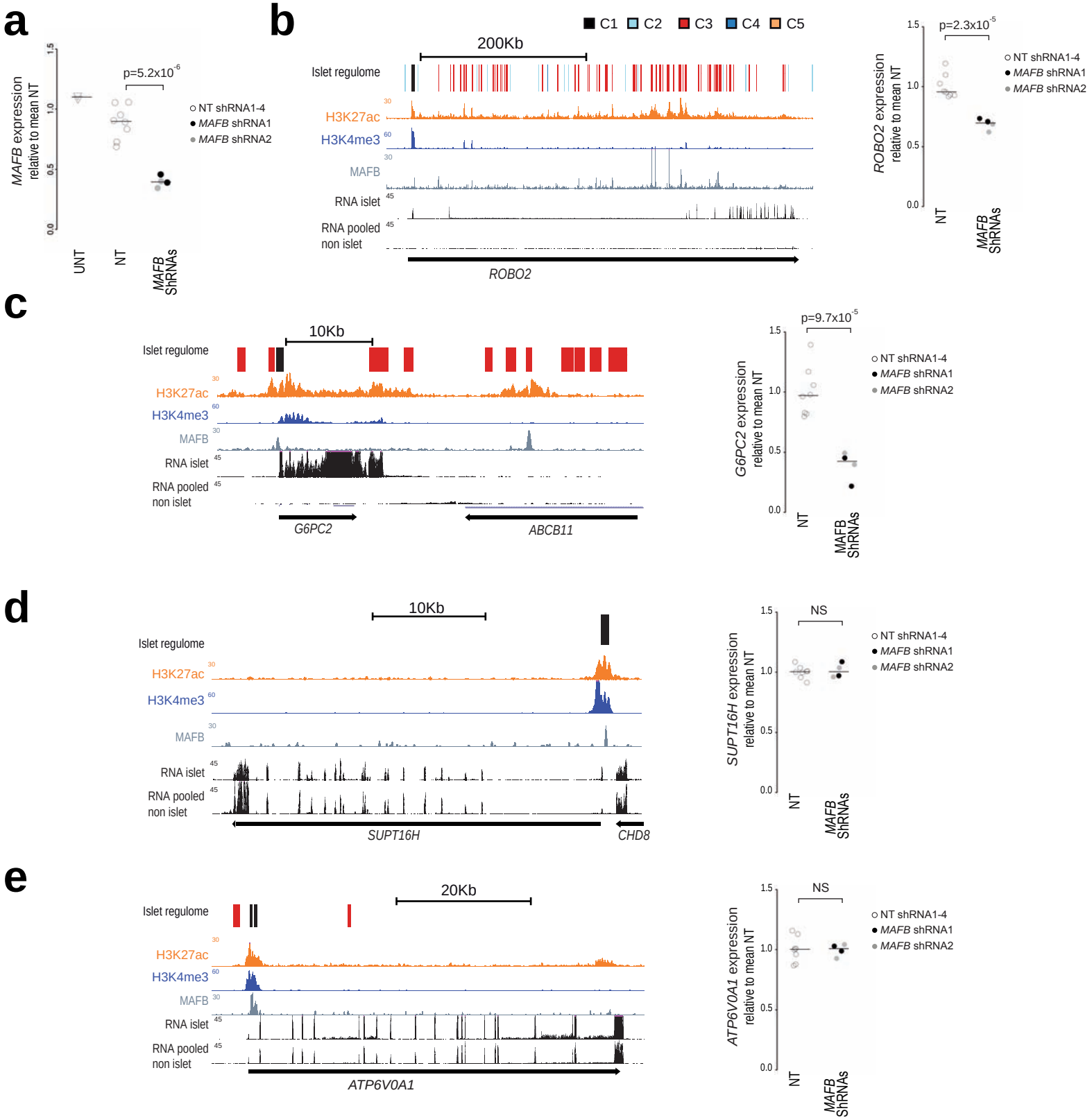
b



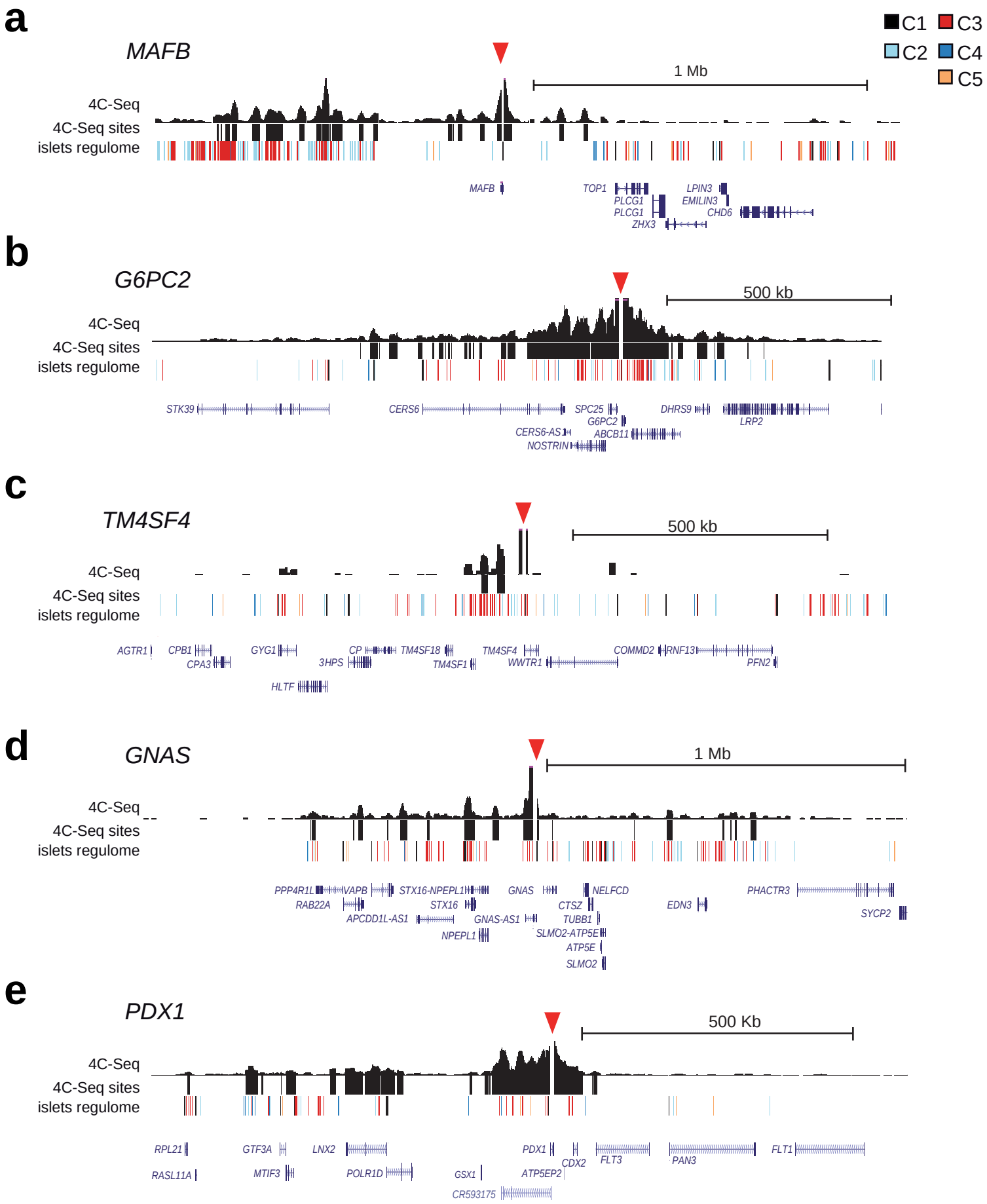
c

Name	*Nearby gene upstream (Kb from TSS)*	*Nearby gene downstream (Kb from TSS)*	*hg18 coordinates*	*Pancreatic Islet enhancer activity*	*Non pancreatic enhancer activity*
C3-1	*AGPAT9* (663)	*NKX6-1* (299)	chr4:85339334-85339883	-	Broad neuronal pattern
C3-3	*ISL1* (1,108)	*PELO* (296)	chr5:51822729-51823938	+	-
C3-4	*PROX1* (14)	*SMYD2* (277)	chr1:212242123-212243697	+	Muscle
C3-5	*LOC100128568* (88)	*RFX2* (43)	chr19:6017422-6018365	+	Hindbrain, pronephric duct, floor plate
C3-6	*TCF7L2* (27)	*HABP2* (573)	chr10:114726967-114727826	-	Broad neuronal pattern

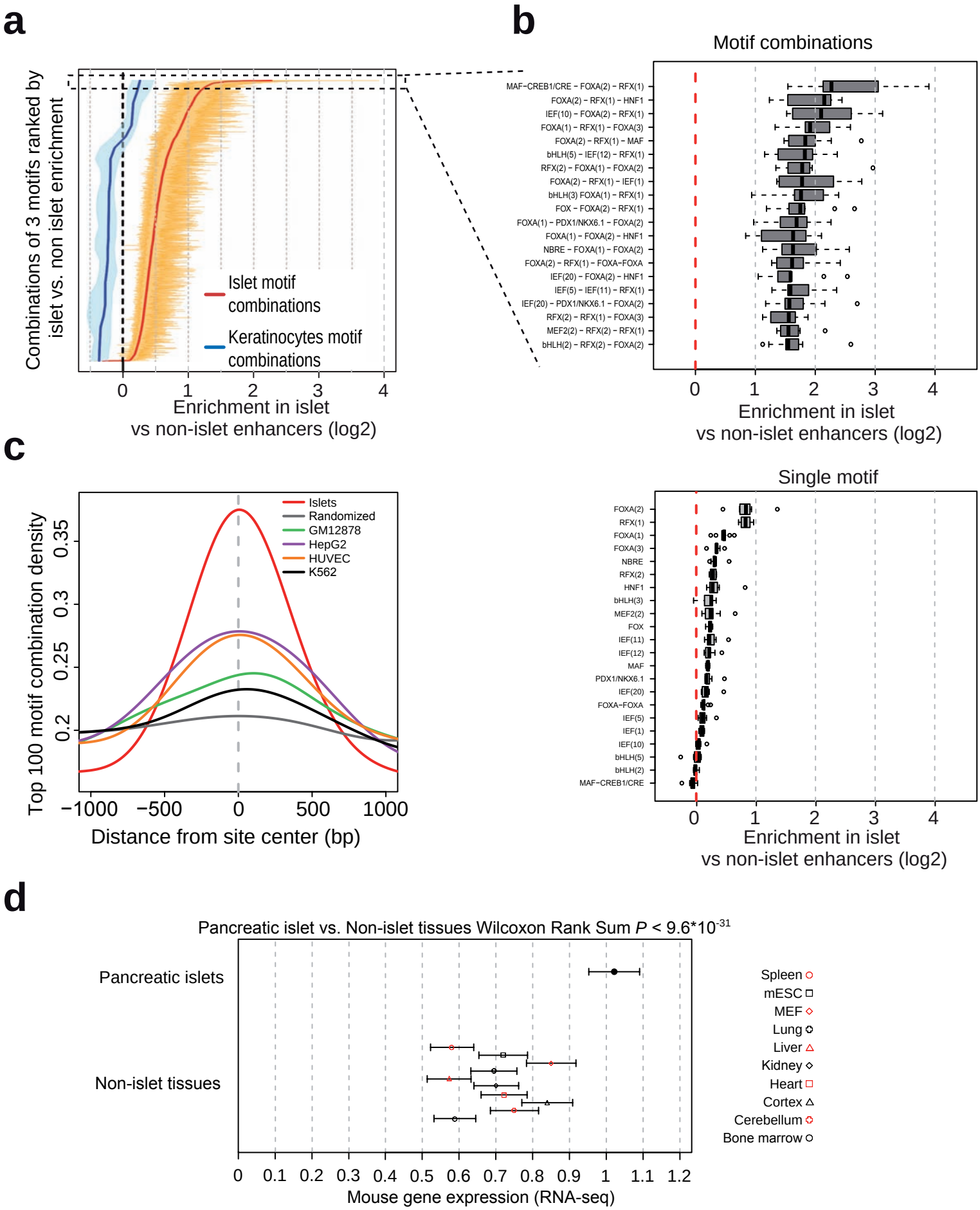
Supplementary Figure 8. In-vivo functional validation of islet enhancers. Human islet C3 (active enhancer) sites drove cell-specific enhancer activity in zebrafish. **(a)** Merged images of YFP and brightfield channels of zebrafish embryos injected with five human C3 (active enhancer) sites, or a control construct containing the *hsp70* promoter and a region lacking enhancer function. YFP expression is observed in the pancreatic islet (pi), neurons (ne), lens (lens), and floor plate (fp). We injected each construct in >200 eggs in at least 3 independent experiments. The quantitative analysis is provided in Supplementary Table 9. **(b)** Islet-specific patterns driven by sites C3-3, C3-4 and C3-5 were confirmed by establishing stable transgenic lines (white arrows point to pancreatic islet expression). C3-3 derived lines show expression in the islet, lens (lens), ventral neural tube (nt) including floor plate (fp) and hindbrain (hi). C3-4 derived lines show expression in the islet and a subset of midbrain neurons (ne). C3-5 derived lines show activity in the pancreatic islet, pronephric ducts (pd) and broad pancentral nervous system activity (cns) including hindbrain (hi) and olfactory bulbs (ob). Expression in the lens is ectopic activity from *hsp70* core promoter[3]. All embryos are 72 hpf, oriented dorsally or laterally anterior to the right. The table shown in **(c)** provides information on tested C3 sites.



Supplementary Figure 9. Functional MAFB targets. Human β cells (EndoC-βH1) were transduced with lentiviral vectors expressing two independent RNA hairpins that target *MAFB* RNA (*MAFB* shRNAs) or four independently transduced negative control non-targeting shRNA sequences (NT). **(a)** Quantitative PCR showed that *MAFB* knock-down with the two shRNAs (black and grey dots) led to 64% and 55% inhibition of *MAFB* mRNA, respectively, relative to the pooled data from NT shRNAs (white dots). The medians are indicated by an horizontal line, p-values were obtained with Student's t test. **(b-c)** Islet-enriched genes *ROBO2* and *G6PC2* are depicted as examples that support the observation that genes associated with *MAFB*-bound enhancer clusters were enriched among downregulated genes after knockdown of *MAFB*. Note that both loci contained prominent enhancer clusters, were bound by MAFB, and transcripts showed significant downregulation with two *MAFB* shRNAs, compared with NT shRNAs. **(d,e)** Examples of genes bound by MAFB at promoter open chromatin, which like other MAFB targets of this class showed no changes in expression upon *MAFB* knockdown. NS: $P > 0.05$. UNT: untreated cells. NT: non-targeting shRNA sequences.

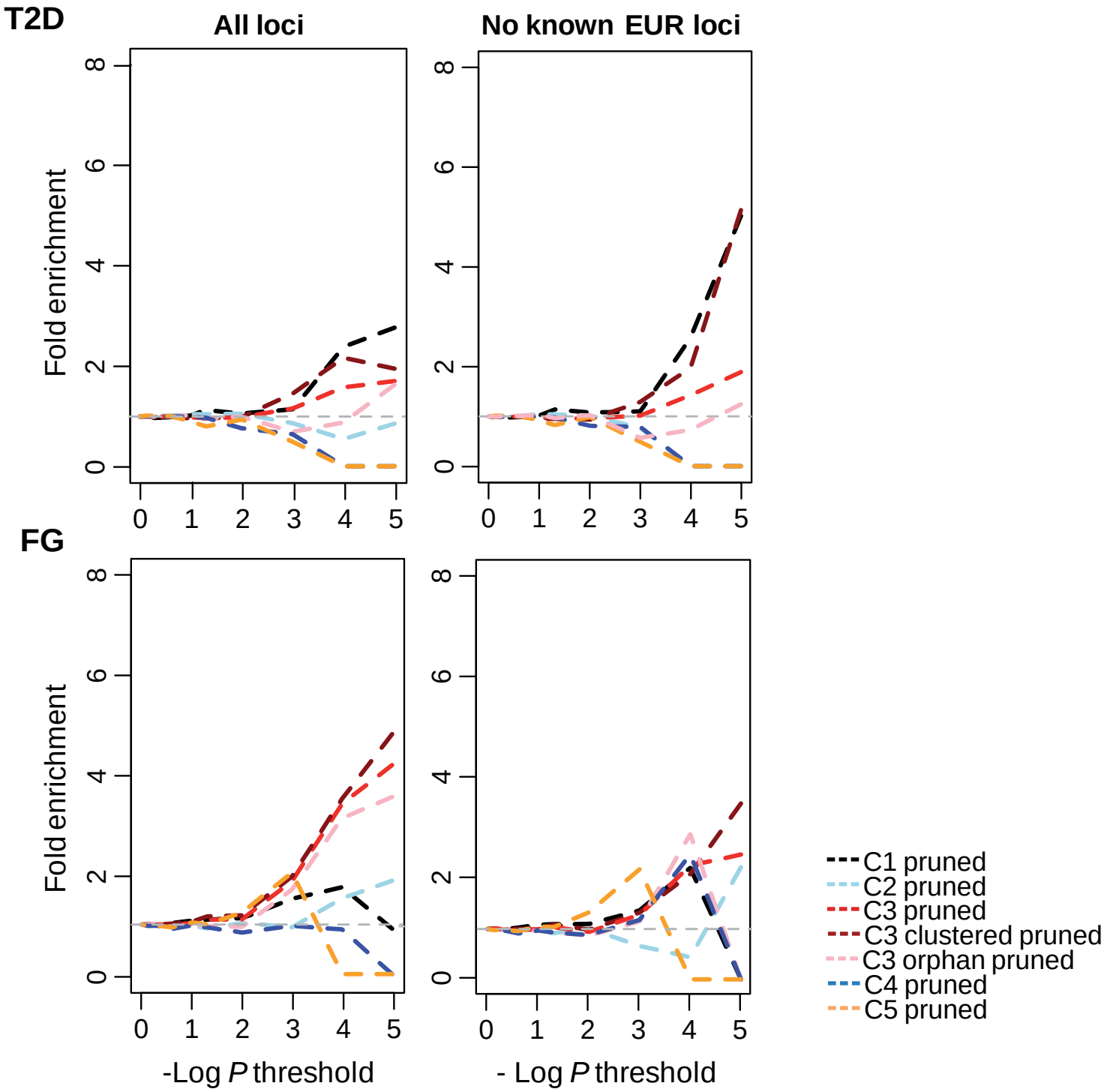


Supplementary Figure 10. Examples of 4C-Seq analysis. 4C-Seq analysis at different loci that contain enhancer clusters showing frequent interactions between distal C3 enhancers and promoters of islet-enriched genes. The red triangle indicates the promoter viewpoint for each 4C-Seq experiment. Genomic sites that interact with viewpoints are represented by a track labeled 4C-Seq sites. The islet regulome track depicts accessible chromatin sites following the color code shown in the upper right quadrant.

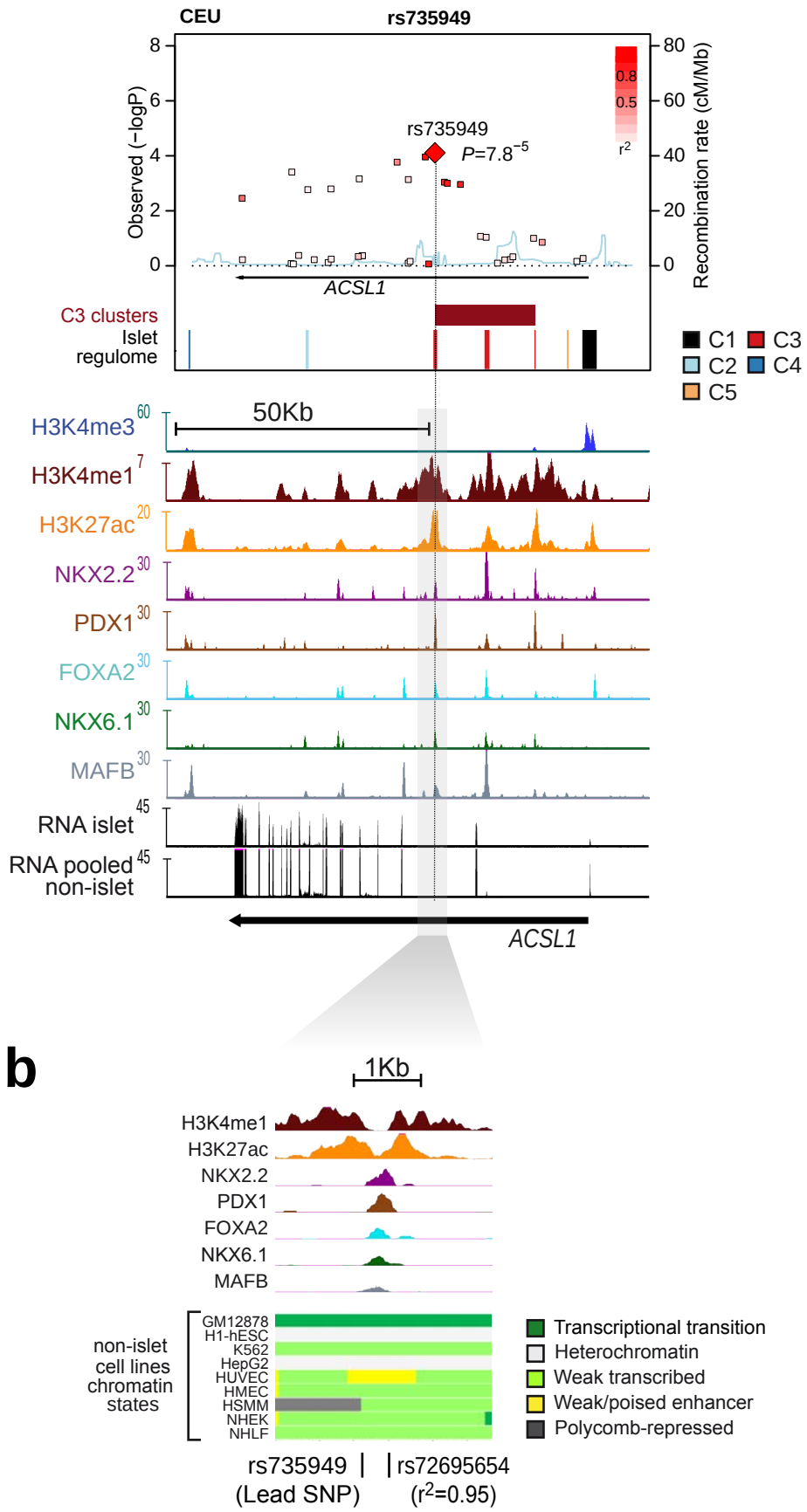


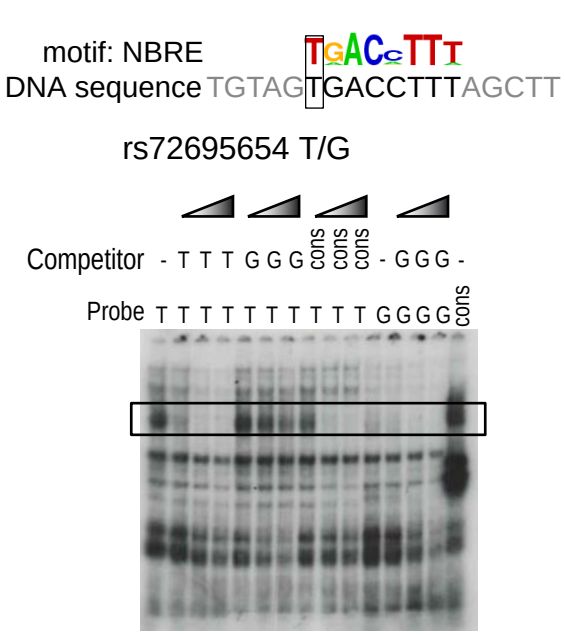
Supplementary Figure 11. Motif combinations enriched in human islet enhancers. (a) Enrichement of 3-motif combinations in clustered islet enhancers vs. non-islet enhancers derived from 9 non-islet cell types[4]. We examined clustered islet C3 sites (or *islet enhancers*) vs. non-islet enhancers and computed the frequency of all possible combinations of 3 motifs from the 46 islet-enriched motifs, restricting the motif search window to +/- 250 bp from the

center of the genomic site. The data is rank-ordered so that combinations that are most enriched in islet enhancers are shown on top. The red line represents the median enrichment, and orange lines depict the range of enrichment values for all nine islet/non-islet ratios. For comparison, we identified combinations of 3 motifs enriched in keratinocyte (NHEK) enhancers, and studied their enrichment in islet enhancers (blue line). **(b)** The most enriched 3-motif combinations in islet active enhancers showed markedly higher enrichment than the individual motifs that compose them. The box plots depict the motif enrichment distribution values in islet vs. active enhancer elements in the 9 non-pancreatic cell types. Motifs for RFX and/or pioneer FOXA factors are invariably present in the most enriched combinations. **(c)** Aggregation plot that represents the frequency density of the 100 most islet-enriched 3-motif combinations in enhancers from different cell types (islets and 4 four non-islet cell lines for which FAIRE and H2A.Z datasets were available to create plots centered on accessible chromatin with the same criteria). Randomized clustered islet enhancers are shown for comparison. **(d)** All instances of the 10 most enriched motif combinations in human islet enhancers where identified in the mouse genome without considering direct sequence orthology, and the resulting sites were associated to nearby genes with Genomic Regions Enrichment of Annotations Tool (GREAT) using default parameters[2]. The graph depicts the median mRNA expression in mouse islets and non-islet tissues for genes that were linked to the motif combinations. The notches surrounding each median line interval represent +/-1.58 IQR/sqrt(n) where IQR is the difference between the 1th and the 3rd quantiles[5]. When notches from two distributions do not overlap this is taken as strong evidence for a significant difference between two medians. On the x-axis gene expression is shown as RPKM values quantile-normalized across all tissues. MEF: mouse embryonic fibroblast. Wilcoxon Rank Sum $P < 9.6 \times 10^{-31}$ for the comparsion between pancreatic islets vs. non-islet tissues.

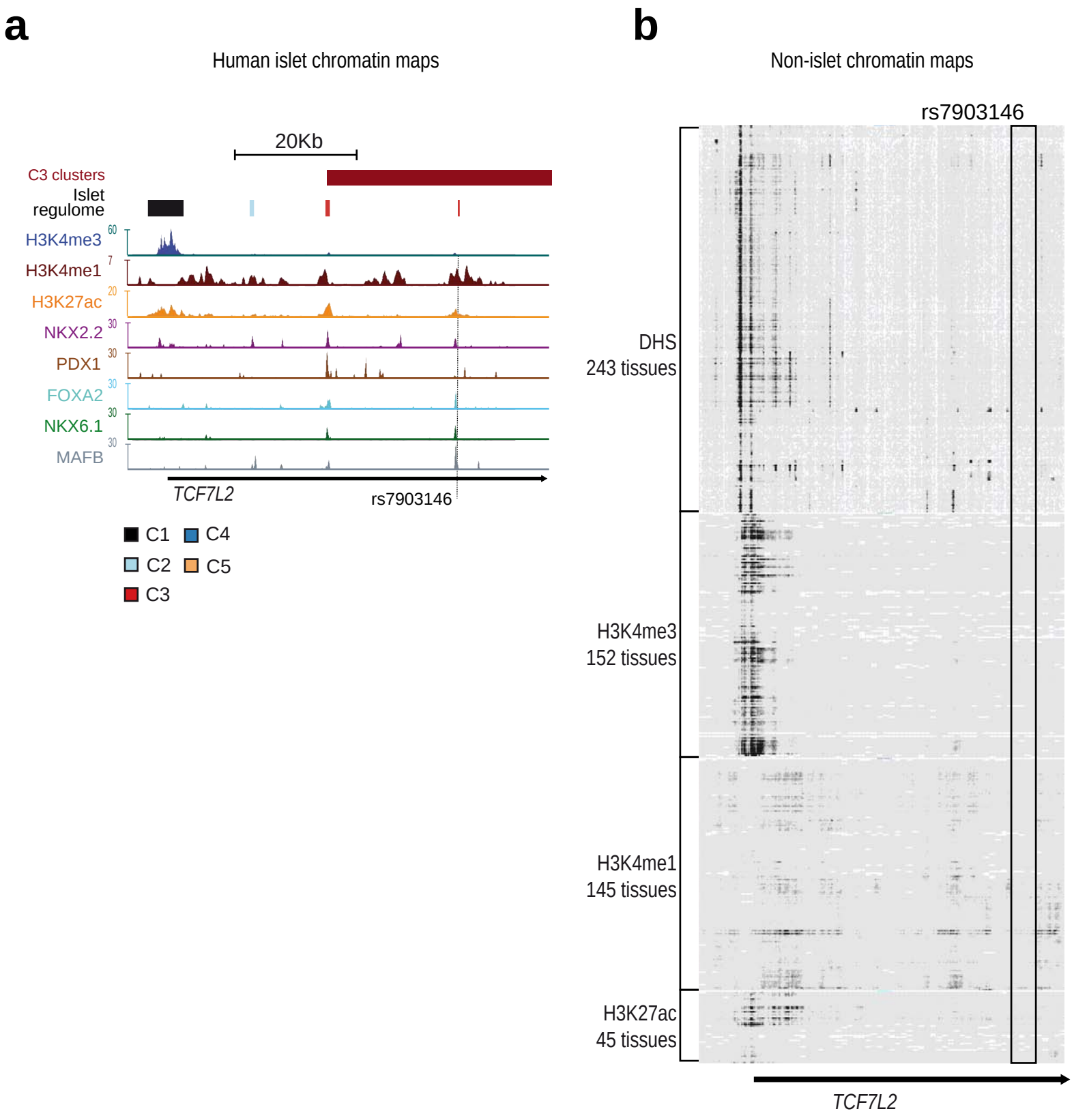


Supplementary Figure 12. Enrichment of pruned T2D and FG association in variants overlapping islet *regulome* sites. HapMap association data for T2D (DIAGRAMv3[6]) and FG level (MAGIC[7]) were pruned retaining only the most significant variant in each LD block (r^2>.2). Variants overlapping each class of islet sites (C1-C5, clustered C3, orphan C3) were then evaluated for fold enrichment over matched background variants at several p value thresholds. An increasing percentage of variants overlap C3 and clustered C3 sites at more significant p value thresholds for both T2D (top left panel) and FG (bottom left panel). Similar patterns of enrichment were observed for C1 sites in T2D data. When removing known European T2D/FG loci, variants overlapping C3, clustered C3 and C1 sites remain enriched at the most significant p-value thresholds for T2D only (right top panel).

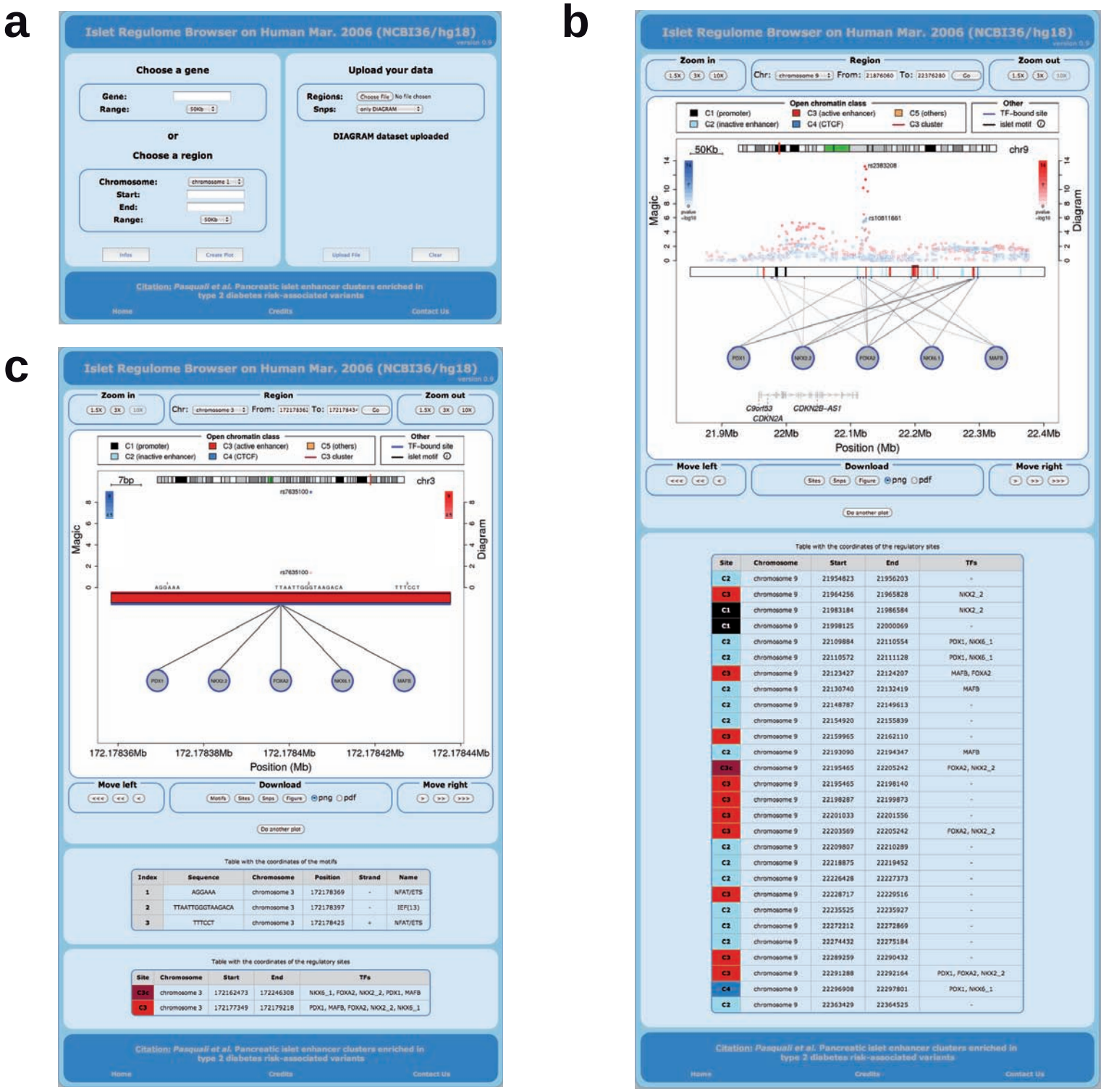




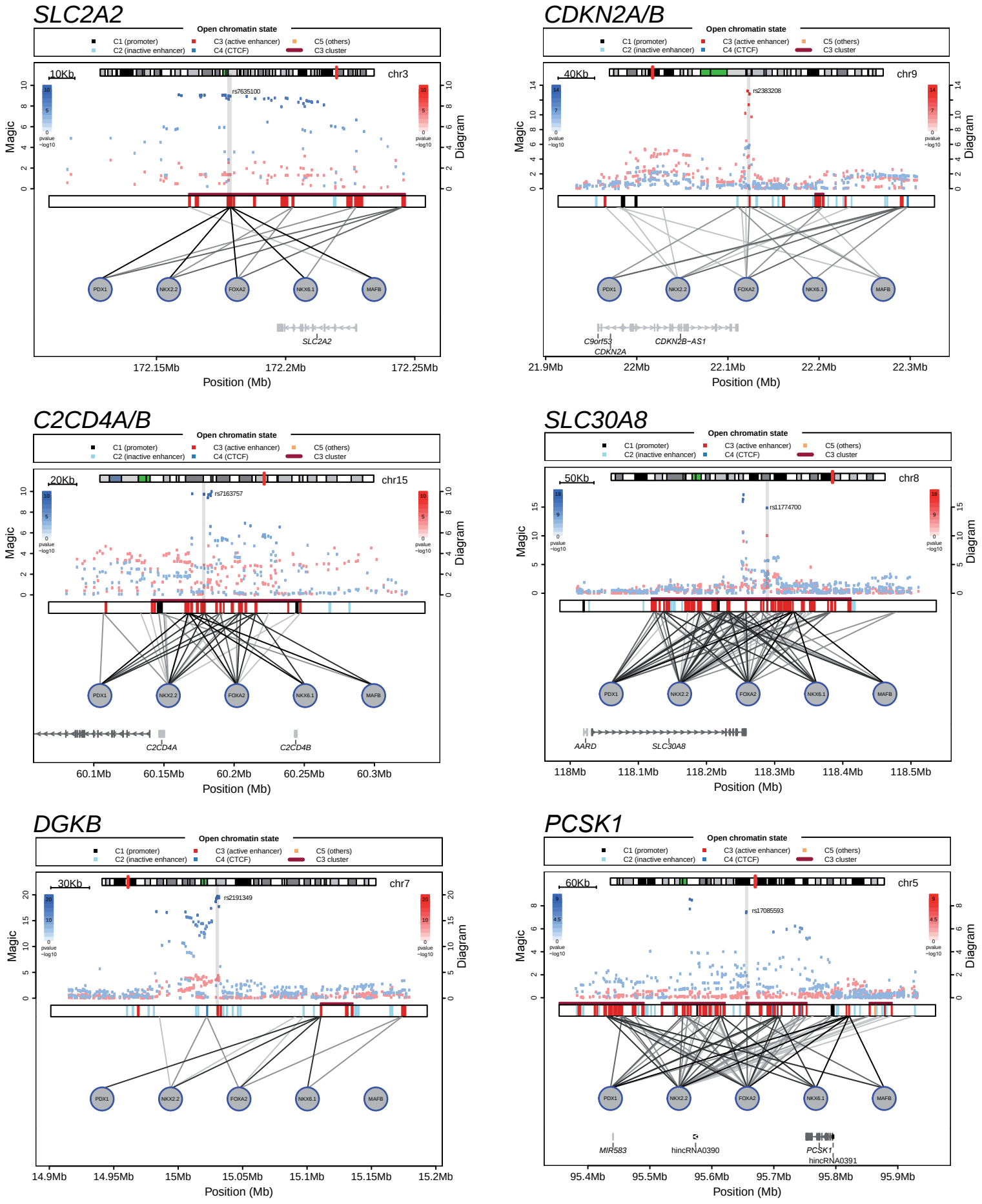
Supplementary Figure 13. *ACSL1*: a novel candidate locus harboring a common FG and T2D risk variant in a clustered enhancer. (a) Regional plot of DIAGRAM variants and islet regulome elements at *ACSL1* (r^2 values based on 1KG CEU with rs735949). SNP rs735949 is strongly associated with T2D (DIAGRAM[6] and Metabochip combined $P = 3.7x10^{-6}$) and independently with FG level (MAGIC[7] FG $P=1.6x10^{-5}$), yet does not surpass conventional GWA significance thresholds. **(b)** SNP rs735949 is the lead SNP in this locus and overlaps an intronic, islet-selective C3 enhancer that is bound by multiple islet transcription factors. Non-islet chromatin state data is taken from Ernst et al.[4]. This SNP is in strong linkage disequilibrium with another SNP, within the same enhancer, which overlaps a nuclear receptor-like motif that is highly enriched in islet enhancers (rs72695654; 1KG CEU r^2 = .94) **(c)**. **(d)** Electrophoretic mobility shift assay shows that the nucleotide change at rs72695654 abolishes sequence-specific binding of a protein complex in MIN6 β cells, supporting an islet regulatory function of this variant. Competition gradients identified by the grey triangle correspond to 5, 50, and 100-fold excess of “cold” competitor probe.



Supplementary Figure 14. Transcription factor and chromatin maps of the locus containing rs7903146 at *TCF7L2*. (a) T2D associated SNP rs7903146 at TCF7L2 locus, previously shown to affect islet chromatin accessible state and enhancer function[8] maps to a C3 active enhancer site bound by NKX2.2, FOXA2 and MAFB in human pancreatic islets. **(b)** Active enhancer chromatin at rs7903146 is specific to human pancreatic islets, as it is not observed in Roadmap Epigenomics datasets including >200 human tissues or cellular types[9].



Supplementary Figure 15. The human islet regulome browser. This open access browser enables viewing of human islet transcription factors binding, human islet open chromatin states, clusters of enhancers, islet motifs, and DIAGRAM T2D association p-values[5], or MAGIC FG p-values[6] at desired levels of resolution throughout the genome. In addition to these standard tracks, it is also possible for users to upload their own variants or regions sets for temporary display. **(a)** Front panel of the browser. **(b)** Example of a locus, depicting the T2D-associated region at *CDKN2B/CDKN2B-AS1* that has a highly associated SNP mapping to a transcription factor bound C3 site. Tables with information on the coordinates of the regulatory elements and transcription factor binding for the browsed regions are visualized on the bottom panel and are available for download. DIAGRAM T2D-association p-values are represented by red dots, whereas MAGIC FG-association p-values are represented by blue dots. For each variant the color intensity of the dot is proportional to –Log p-value of association. Vertical colored lines depict different chromatin states. Black lines point to transcription factors binding sites and their intensity is proportional to the number of bound transcription factors. Islet-specific genes are shown in dark grey. **(c)** At a zoom-in resolution of less than 1Kb per window, islet enriched motifs described in Supplementary Table 3 are shown in C3 sites and their sequence and genomic location are visualized in a table format in the middle panel. The human islet regulome browser is available at www.isletregulome.org.



Supplementary Figure 16. Examples of T2D and FG associated variants located in islet active enhancers. Screenshots from the human islet regulome browser showing T2D and/or FG-associated loci in which the associated variants directly map to C3 active enhancer sites (grey boxes). GWAs DIAGRAM T2D-associated (in red) and MAGIC FG-associated variants (in blue) are shown along with transcription factors binding and chromatin states as described in Supplementary Figure 15.

Supplementary tables:

Supplementary Table 1: Pancreatic expression patterns and genetic phenotypes of transcription factors examined in this study.

Transcription factor	Adult islet cell type expression	Mouse genetic phenotypes	Human Genetic Phenotypes
PDX1	β and δ-cells	-/-: Pancreas agenesis due to failure of expansion of pancreatic progenitors[10], hyperglycemia and early death.	-/-: Pancreas agenesis.
		+/-: normal pancreatic development, diabetic phenotype[11].	+/-: Autosomal dominant diabetes mellitus[13].
		β-cell-specific KO: oss of β-cells, diabetes mellitus[12].	
NKX6.1	β-cells	-/-: Reduced islet-cell number, β-cell differentiation defect[14].	None described.
FOXA2	β, α and δ-cells	β-cell specific knock-out: insulin secretion dysregulation[15,16].	None described.
MAFB	β, α and δ-cells	-/-: Reduced number of β-cells and α-cells[17].	-/-: None described.
			+/-: Multicentric carpotarsal osteolysis syndrome[18].
NKX2.2	β, α and δ-cells	-/-: Absence β-cells, reduced number of PP and α-cells, increased number of ghrelin cells. Hyperglycemia and early death[19].	None published.

Supplementary Table 2: Association with enhancer clusters for 65 genes important for islet cell identity and function.

Symbol	Common usage symbol	Gene Name	Associated with high TF occupancy islet enhancer cluster*	Associated with islet enhancer cluster*	Reference associating gene to islet identity or function
ABCC8	*SUR1*	ATP-binding cassette, sub-family C (CFTR/MRP), member 8	YES	YES	Science. 1995 Apr 21;268(5209):423-6.
ADCYAP1	*PACAP*	Adenylate cyclase activating polypeptide 1 (pituitary)	YES	YES	Regul Pept. 2003 May 15;113(1-3):31-9
CACNA1C		Calcium channel, voltage-dependent, L type, alpha 1C subunit	YES	YES	Am J Physiol Endocrinol Metab. 2005 Jan;288(1):E16-28.
CACNA1D		Calcium channel, voltage-dependent, L type, alpha 1D subunit	YES	YES	Diabetologia. 2013 Feb;56(2):340-9
CACNA1E		Calcium channel, voltage-dependent, R type, alpha 1E subunit	YES	YES	J Clin Invest. 2005 Jan;115(1):16-20
CHGA		Chromogranin A (parathyroid secretory protein 1)	YES	YES	Pancreas. 2003 Jul;27(1):38-46
CPE		Carboxypeptidase E	YES	YES	Proc Natl Acad Sci U S A. 2008 Jun 17;105(24):8452-7.
DACH1		Dachshund homolog 1 (Drosophila)	YES	YES	Dev Biol. 2010 Dec 15;348(2):143-52.
FFAR1	*GPR40*	Free fatty acid receptor 1	YES	YES	Cell Metab. 2005 Apr;1(4):245-58
FOXA2		Forkhead box A2	YES	YES	Mol Endocrinol. 2010 Aug;24(8):1594-604
FOXO1		Forkhead box O1	YES	YES	Nat Genet. 2002 Oct;32(2):245-53.
FXYD2		FXYD domain containing ion transport regulator 2	NO	YES	Diabetologia. 2010 Jul;53(7):1372-83
G6PC2		glucose-6-phosphatase, catalytic, 2	YES	YES	Diabetes. 1999 Mar;48(3):531-42.
GAD2	*GAD65*	Glutamate decarboxylase 2 (pancreatic islets and brain, 65kDa)	YES	YES	J Clin Invest. 1993 Jan;91(1):368-74
GCK		Glucokinase (hexokinase 4)	NO	YES	Proc Natl Acad Sci U S A. 1986 Apr;83(7):1998-2001.
GIPR		Gastric inhibitory polypeptide receptor	YES	YES	Diabetes. 1995 Oct;44(10):1202-8
GJD2	*Connexin36*	Gap junction protein, delta 2, 36kDa	YES	YES	J Clin Invest. 2000 Jul;106(2):235-43
GLIS3		GLIS family zinc finger 3	YES	YES	Mol Cell Biol. 2009 Dec;29(24):6366-79
GLP1R		glucagon-like peptide 1 receptor	YES	YES	Proc Natl Acad Sci U S A. 1992 Sep 15;89(18):8641-5.
HNF1A		HNF1 homeobox A	NO	NO	J Clin Invest. 1998 May 15;101(10):2215-22.
IAPP		Islet amyloid polypeptide	YES	YES	Proc Natl Acad Sci U S A. 1987 Jun;84(11):3881-5.
INS		insulin	NO	NO	Proc Natl Acad Sci U S A. 1969 Jan;62(1):278-85.
INSM1		Insulinoma-associated 1	YES	YES	EMBO J. 2006 Mar 22;25(6):1344-52
ISL1		ISL LIM homeobox 1	YES	YES	Nature. 1990 Apr 26;344(6269):879-82
KCNJ11	*KIR6.2*	Potassium inwardly-rectifying channel, subfamily J, member 11	YES	YES	Science. 1995 Nov 17;270(5239):1166-70.
MAFA		V-maf musculoaponeurotic fibrosarcoma oncogene homolog A (avian)	NO	NO	Mol Cell Biol. 2003 Sep;23(17):6049-62.
MNX1	*HLXB9*	Motor neuron and pancreas homeobox 1	YES	YES	Nat Genet. 1999 Sep;23(1):67-70.
MYT1		Myelin transcription factor 1	YES	YES	Development. 2004 Jan;131(1):165-79
NEUROD1		Neurogenic differentiation 1	YES	YES	Genes Dev. 1997 Sep 15;11(18):2323-34.
NKX2.2		NK2 homeobox 2	YES	YES	Development. 1998 Jun;125(12):2213-21.
NKX6.1		NK6 homeobox 1	YES	YES	Development. 1998 Jun;125(12):2213-21
PAX6		Paired box 6	YES	YES	Genes Dev. 1997 Jul 1;11(13):1662-73.
PCSK1		Proprotein convertase subtilisin/kexin type 1	YES	YES	Endocrinology. 1994 Oct;135(4):1651-60
PCSK2		Proprotein convertase subtilisin/kexin type 2	YES	YES	Endocrinology. 1994 Oct;135(4):1651-60
PDE3B		Phosphodiesterase 3B, cGMP-inhibited	YES	YES	J Clin Invest. 2006 Dec;116(12):3240-51.
PDX1		Pancreatic and duodenal homeobox 1	YES	YES	EMBO J. 1993 Nov;12(11):4251-9.
PRKCA		protein kinase C, alpha	YES	YES	Mol Cell Endocrinol. 1996 May 31;119(2):185-93
PRLR		Prolactin receptor	YES	YES	Endocrinology. 2002 Apr;143(4):1378-85.
PTPRN		Protein tyrosine phosphatase, receptor type, N	YES	YES	J Clin Invest. 1995 Sep;96(3):1506-11
PTPRN2		Protein tyrosine phosphatase, receptor type, N polypeptide 2	YES	YES	Proc Natl Acad Sci U S A. 1996 Mar 19;93(6):2307-11
RFX6		Regulatory factor X, 6	YES	YES	Nature. 2010 Feb 11;463(7282):775-80.
RGS4		Regulator of G-protein signaling 4	YES	YES	Proc Natl Acad Sci U S A. 2010 Apr 27;107(17):7999-8004
SCG2		Secretogranin II	YES	YES	Curr Mol Med. 2001 Dec;1(6):727-32.
SCG5		Secretogranin V (7B2 protein)	YES	YES	J Clin Endocrinol Metab. 1986 Sep;63(3):758-65.
SCGN		Secretagogin, EF-hand calcium binding protein	NO	YES	J Biol Chem. 2000 Aug 11;275(32):24740-51.
SLC2A2		Solute carrier family 2 (facilitated glucose transporter), member 2	YES	YES	Proc Natl Acad Sci U S A. 1990 Jun;87(11):4088-92.
SLC30A8		Solute carrier family 30 (zinc transporter), member 8	YES	YES	Proc Natl Acad Sci U S A. 2007 Oct 23;104(43):17040-5.
SNAP25		Synaptosomal-associated protein, 25kDa	NO	YES	J Cell Biol. 1995 Mar;128(6):1019-28
SOX4		SRY (sex determining region Y)-box 4	YES	YES	Diabetes. 2005 Dec;54(12):3402-9
ST18		Suppression of tumorigenicity 18 (breast carcinoma) (zinc finger protein)	YES	YES	Mech Dev. 2007 Nov-Dec;124(11-12):898-910
STX1A		Syntaxin 1A (brain)	NO	NO	Proc Natl Acad Sci U S A. 1994 Dec 20;91(26):12487-91.
SYT7		synaptotagmin VII	YES	YES	FASEB J. 2008 Jan;22(1):194-206. Epub 2007 Aug 20
TCF7L2		Transcription factor 7-like 2 (T-cell specific, HMG-box)	YES	YES	J Clin Invest. 2007 Aug;117(8):2155-63.
TM4SF4		Transmembrane 4 L six family member 4	NO	YES	Development. 2011 Aug;138(15):3213-24.
UCN3		Urocortin 3 (stresscopin)	YES	YES	Nat Biotechnol. 2012 Feb 26;30(3):261-4.
WNT4		Wingless-type MMTV integration site family, member 4	YES	YES	Genome Res. 2010 Jun;20(6):722-32.
SULT4A1		Sulfotransferase family 4A, member 1	YES	YES	Biochem J. 2000 Mar 15;346 Pt 3:857-64.
NCAM1		Neural cell adhesion molecule 1	YES	YES	J Cell Biol. 1999 Jan 25;144(2):325-37.
LMX1A		LIM homeobox transcription factor 1, alpha	NO	YES	Genomics. 1997 Dec 15;46(3):520-4
KCNK10		Potassium channel, subfamily K, member 10	YES	YES	Biochem Biophys Res Commun. 2004 Oct 8;323(1):323-31.
KCNB2		Potassium voltage-gated channel, Shab-related subfamily, member 2	YES	YES	J Biol Chem. 2013 Jun 20.
SYT4		Synaptotagmin IV	YES	YES	Curr Biol. 2003 Apr 1;13(7):563-7.
CDKAL1		CDK5 regulatory subunit associated protein 1-like 1	YES	YES	PLoS One. 2010 Dec 9;5(12):e15553
RFX3		Regulatory factor X, 3 (influences HLA class II expression)	YES	YES	Diabetes. 2007 Apr;56(4):950-9.
		Total Number with clusters	54	60	
		Fraction with clusters	0.83	0.92	

* Genes were associated with a cluster of islet enhancers using Genomic Regions Enrichment of Annotations Tool (GREAT) with default parameters[2]. High transcription factor occupancy enhancer clusters were defined as those showing the top two transcription factor occupancy score quartiles (**Supplementary Figure 7a**). The transcription factor occupancy score measures the frequency with which enhancers in a cluster are bound by islet transcription factors.

Supplementary Table 3: *De novo* sequence motifs enriched in islet clustered enhancers.

Name	p-value	Sequence Logo	Tomtom known matches (p<0.0005 q<0.5)	HOMER known matches (score>0.8)
AP1	1e-344		JUNDM2, CREB2, AP1, FOS, NF-E, NF32L2	AP-1, HIF1B, JUNDM2-2, FOS, NF-E2, BACH1, MAFK, NRF2
FOX	1e-320		FREAC-2, FOXD1, FOXO3	FOXO1, FOXP1, FOXD1, FOXO3, FOXA2, FOXA1
MAF	1e-274		NONE	MAFB, MAFA
PDX1/NKX6.1	1e-255		NONE	PDX1, NK6.1, PDX1
HNF1	1e-229		HNF1A, HNF1B, HNF-1, HOXA4, HOXB4	HNF1A, HNF1B, HNF1, FOXA2
RFX(1)	1e-199		RFX4, RFX3, RFXDC2, RFX1, RFX, MIF-1	RFX5, RFX4, RFXDC2, RFX3-1
RFX(2)	1e-194		RFX3, RFX4, NFACTC1, RFXDC2	RFX, X-BOX, RFX1
FOXA(1)	1e-192		FOXA1, FOXA2, FOXJ3, FOXD3	FOXA1,
FOXA(2)	1e-182		NONE	NONE
bHLH(1)	1e-171		MYF, NHLH1, NEUROD1.1, ASCL2, HLH-1, LIN-32, MYF6	ATOH1, TCF12, SCL, E2A, MYOG, MYOD, MYF5, NEUROD1, ASCL2.1, OLIG2
FOXA-FOXA	1e-161		NONE	NONE
IEF(1)	1e-143		NONE	NONE
NFAT/ETS	1e-141		NONE	NFATC2, ERG, NFAT, ETS1, ELF5, FLI1, FEV, SPIB
ETS	1e-131		FEV, C-ETS-2, ELF3, SPI1	ETS1, ERG, ETS1
FOXA(3)	1e-126		FOX2A, FOXO3, FOXA1, FREAC-2	NONE
IEF(2)	1e-110		NONE	NONE
IEF(3)	1e-107		NONE	NONE
HOMEODOMAIN	1e-106		RHOBX6, LHX3, PHOX2A, CART1, PROP1, POU2F1, ALX4, POU3F2, PAX7, OG2X, SHOX2, VAX2, EN1, LHX2, POU4F3, TCF1, PAX6, HOXC4, LBX2, POU6F1, PDX1, NANOG3, HOXC5, DBX2, HOXA3, PAX4, LHX9, LHX6, RAX, ARX, POU1F1	NONE
bHLH(2)	1e-103		NEUROD1.1, AP-4	NONE
IEF(4)	1e-101		NONE	NONE
bHLH(3)	1e-100		NONE	NEUROD1

Supplementary Table 3 (continued)

Name	p-value	Sequence Logo	Tomtom known matches (p<0.0005 q<0.5)	HOMER known matches (score>0.8)
bHLH(4)	1e-98		TFEB, SREBP-1, ARNT, MYCN, NEUROD1, BHLHB2, MAX, USF1, NEUROD1.3, CREB4, CREB6, BZIP-2, MYCN, MYC, NEUROD1.2, PAX-3	USF1, USF2, E-BOX, ATF3, BHLHE40, BHLHB2.1, USF1, MYCN, ARNT
bHLH(5)	1e-97		NEUROD1.1, AP-4	NONE
IEF(5)	1e-97		NONE	NONE
IEF(6)	1e-96		NONE	NONE
IEF(7)	1e-95		NONE	NONE
MAF-CREB1/CRE	1e-92		ATF3, CREB1, MAFB, JUNDM2, ATF1	NONE
IEF(8)	1e-89		NONE	NONE
IEF(9)	1e-85		NONE	NONE
PRDM/IRF	1e-84		PRDM1, IRF2, IRF1	PRDM1/BMI1, ISRE, IRF2
IEF(10)	1e-83		NONE	NONE
IEF(11)	1e-81		NONE	NONE
IEF(12)	1e-79		NONE	NONE
MEF2(1)	1e-78		MEF2A, RSRFC4	MEF2C, MEF2A
IEF(13)	1e-78		NONE	NONE
MEF2(2)	1e-78		MEF2A, RSRFC4, MEF-2	NONE
IEF(14)	1e-77		NONE	NONE
IEF(15)	1e-77		NONE	NONE
IEF(16)	1e-71		NONE	NONE
IEF(17)	1e-69		NONE	NONE
SP1	1e-68		SP1	NONE
IEF(18)	1e-67		NONE	NONE
NRBE	1e-66		RXRG, HBF4-7, NR1H2::RXRA, PPARG::RXRA, RARA, NR2F2, NR2F1, NR4A2, HNF4A, HNF4, ESRRA, RORA1 NR2F2, ESRRB	NUR77, NER2F2, RARA, ESRRB, ERRA NR4A2, ESRRA, RORA, NR2F1
CEBP/HLF	1e-65		HLF, CEBPE, CEBPA	HLF, CEBP, CEBPA, CEBP-LIKE, CEBPA
IEF(19)	1e-62		NONE	NONE
IEF(20)	1e-60		NONE	NONE

Notes: IEF= Islet Enriched Factor

Supplementary Table 4: Overlap of T2D and fasting glycemia level-associated common SNPs with different types of accessible chromatin sites or with motifs that are enriched in islet accessible chromatin sites.

Islet sites

Disease phenotype	# SNPs*	Total # loci *	SNPs overlapping C1 (# loci) **	SNPs overlapping C2 (# loci)	SNPs overlapping C3 (# loci)	SNPs overlapping C4 (# loci)	SNPs overlapping C5 (# loci)
Type 2 diabetes	1,914	66	25 (11)	17 (10)	63 (20)	4 (3)	6 (3)
Fasting glucose	698	38	19 (8)	5 (4)	49 (18)	2 (2)	2 (2)

Islet motifs

Disease phenotype	# SNPs*	Total # loci *	SNPs overlapping a motif in C1 (# loci) **	SNPs overlapping a motif in C2 (# loci)	SNPs overlapping a motif in C3 (# loci)	SNPs overlapping a motif in C4 (# loci)	SNPs overlapping a motif in C5 (# loci)
Type 2 diabetes	1,914	66	2 (2)	1 (1)	6 (6)	1 (1)	0 (0)
Fasting glucose	698	38	2 (2)	1 (1)	8 (6)	0 (0)	0 (0)

*All 1000 Genomes Project pilot 1 SNPs in LD r2>0.8 in CEU with index variants defined from Morris et. al. 2012[6] (for T2D) and Scott et al.[7] (for FG).
** Positions of both islet sites and SNPs were lifted over to hg19 before calculating overlap

Supplementary Table 5: List of T2D-associated common SNPs overlapping islet enhancers.

SNP	Locus	Enhancer type	Motifs*
rs74382177	*THADA*	clustered C3	(C5_M6-s)
rs78681698	*THADA*	clustered C3	
rs79629200	*THADA*	clustered C3	
rs74469180	*THADA*	clustered C3	
rs28700209	*THADA*	clustered C3	
rs7567685	*THADA*	clustered C3	
rs7581586	*THADA*	clustered C3	
rs6752448	*THADA*	clustered C3	
rs76282560	*THADA*	clustered C3	
rs6728106	*THADA*	clustered C3	(C2_M17-l)
rs17039133	*THADA*	clustered C3	
rs7600657	*THADA*	clustered C3	
rs998768	*THADA*	clustered C3	(C3_M24-s)
rs74821293	*THADA*	orphan C3	
rs72865297	*THADA*	orphan C3	(C4_M11-s)
rs74600494	*THADA*	orphan C3	
rs11936387	*MAEA*	orphan C3	(C2_M10-s)
rs4689388	*WFS1*	orphan C3	
rs4320200	*WFS1*	orphan C3	
rs13107806	*WFS1*	orphan C3	
rs13127445	*WFS1*	orphan C3	
rs4273545	*WFS1*	orphan C3	
rs6830765	*WFS1*	orphan C3	
rs4457054	*ZBED3*	clustered C3	
rs7708285	*ZBED3*	clustered C3	
rs7732130	*ZBED3*	clustered C3	RFX(1); (C3_M4-l); (C2_M7-s)
rs9348441	*CDKAL1*	clustered C3	
rs77114369	*ZFAND3*	clustered C3	
rs58692659	*ZFAND3*	clustered C3	bHLH(3); (C3_M10-l);(C2_M35-l); (C4_M2-s)
rs61332486	*ZFAND3*	clustered C3	(C2_M16-s)
rs57995712	*ZFAND3*	orphan C3	(C5_M7-s)
rs2908286	*GCK*	clustered C3	
rs4607517	*GCK*	clustered C3	
rs508419	*ANK1*	clustered C3	
rs9694034	*ANK1*	clustered C3	
rs6989203	*ANK1*	clustered C3	
rs11774700	*SLC30A8*	clustered C3	
rs4237150	*GLIS3*	clustered C3	
rs10814915	*GLIS3*	clustered C3	
rs10811660	*CDKN2A/B*	orphan C3	
rs10811661	*CDKN2A/B*	orphan C3	
rs703977	*ZMIZ1*	orphan C3	IEF(7)
rs7903146	*TCF7L2*	clustered C3	(C3_M22-s); (C3_M6-l)
rs231361	*KCNQ1*	clustered C3	
rs3862791	*ARAP1*	clustered C3	(C2_M16-s)
rs12581729	*KLHDC5*	clustered C3	
rs10842991	*KLHDC5*	clustered C3	
rs10771372	*KLHDC5*	clustered C3	(C2_M16-l)
rs3751239	*KLHDC5*	clustered C3	
rs10842992	*KLHDC5*	clustered C3	
rs10842993	*KLHDC5*	clustered C3	
rs11049161	*KLHDC5*	clustered C3	
rs7163757	*C2CD4A*	clustered C3	(C1_M17-l); (C3_M8-s)
rs1357335	*ZFAND6*	orphan C3	
rs1357336	*ZFAND6*	orphan C3	
rs72804106	*BCAR1*	orphan C3	
rs11670462	*GIPR*	clustered C3	
rs55872740	*GIPR*	clustered C3	
rs10403962	*GIPR*	clustered C3	
rs10404142	*GIPR*	clustered C3	
rs10404527	*GIPR*	clustered C3	
rs10409882	*GIPR*	clustered C3	
rs8104845	*GIPR*	clustered C3	

* Motifs enriched in clustered islet C3 enhancers are defined in Supplementary Table 3, additional motifs enriched in different types of islet accessible chromatin but not specifically in clustered islet C3 enhancers are listed in brackets.

Supplementary Table 6: List of fasting glycemia level-associated SNPs overlapping islet enhancers.

SNP	Locus	Enhancer type	Motifs*
rs13431652	*G6PC2*	clustered C3	(C1_M11-s)
rs4625	*AMT*	orphan C3	
rs1905506	*SLC2A2*	clustered C3	(C2_M15-s)
rs1905504	*SLC2A2*	clustered C3	(C3_M8-l);
rs7635100	*SLC2A2*	clustered C3	IEF(13)
rs7635470	*SLC2A2*	clustered C3	
rs11923694	*SLC2A2*	clustered C3	
rs11920090	*SLC2A2*	clustered C3	
rs11924648	*SLC2A2*	clustered C3	
rs61169219	*SLC2A2*	clustered C3	(C3_M4-s); (C2_M8-s)
rs7638998	*SLC2A2*	clustered C3	
rs5393	*SLC2A2*	clustered C3	
rs4457054	*ZBED3*	clustered C3	
rs7708285	*ZBED3*	clustered C3	
rs7732130	*ZBED3*	clustered C3	RFX(1); (C3_M4-l); (C2_M7-s)
rs56198733	*PCSK1*	clustered C3	
rs4869273	*PCSK1*	clustered C3	
rs12186664	*PCSK1*	clustered C3	
rs17085593	*PCSK1*	clustered C3	FOX; (C2_M13-l)
rs59139497	*PCSK1*	clustered C3	
rs2882298	*PCSK1*	clustered C3	
rs9348441	*CDKAL1*	clustered C3	
rs10244051	*DGKB*	orphan C3	
rs10950550	*DGKB*	orphan C3	
rs10228456	*DGKB*	orphan C3	
rs10228561	*DGKB*	orphan C3	
rs10228796	*DGKB*	orphan C3	
rs10258074	*DGKB*	orphan C3	
rs2191348	*DGKB*	orphan C3	(C4_M12-s)
rs2191349	*DGKB*	orphan C3	
rs2908286	*GCK*	clustered C3	
rs4607517	*GCK*	clustered C3	
rs11774700	*SLC3A8*	clustered C3	
rs6476842	*GLIS3*	clustered C3	
rs10814916	*GLIS3*	clustered C3	
rs10811660	*CDKN2A/B*	orphan C3	
rs10811661	*CDKN2A/B*	orphan C3	
rs7903146	*TCF7L2*	clustered C3	(C3_M22-s); (C3_M6-l)
rs7945565	*CRY2*	clustered C3	(C2_M18-l)
rs7945689	*CRY2*	clustered C3	
rs1401419	*CRY2*	clustered C3	ETS; NFAT; (C3_M24-s)
rs10501320	*MADD*	orphan C3	RFX(2); (C3_M13-s)
rs3862791	*ARAP1*	clustered C3	(C2_M16-s)
rs35369009	*PDX1*	orphan C3	
rs3783346	*WARS*	orphan C3	(C2_M22-l)
rs73300993	*FOXA2*	clustered C3	
rs6048202	*FOXA2*	clustered C3	
rs1203898	*FOXA2*	clustered C3	
rs1203899	*FOXA2*	clustered C3	

* Motifs enriched in clustered islet C3 enhancers are defined in Supplementary Table 3, additional motifs enriched in different types of islet accessible chromatin but not specifically in clustered islet C3 enhancers are listed in brackets.

Supplementary Table 7: Characteristics of human islet donors and samples.

Sample ID	Sex	Age (years)	Cause of death	BMI (kg/m2)	Premortem diagnosis of Diabetes mellitus	Premortem non-fasting glycemia (mmol/l)	Cold ischemia (hours)	Islet purity (%)	Experiments
HI 10	Male	53	Cerebral hemorrhage	23.5	No	6.9	6.5	90	FAIRE-Seq
HI 21	Male	70	NA	26.0	No	NA	NA	98	H3K4me1, H3K4me3
HI 22	Male	63	NA	30.2	No	NA	NA	94	H2A.Z, CTCF
HI 25	Male	59	Ischemic brain injury	24.2	No	11.3	6.4	93	H3K4me1, H3K4me3
HI 32	Male	38	Trauma	22.9	No	5.1	8.5	94	H3K4me1, H3K4me3, H3K27ac, FAIRE-Seq, H2A.Z, CTCF, PDX1, FOXA2, Input DNA
HI 34	Male	62	Cerebral trauma	27.5	No	6.4	5	95	H2A.Z, CTCF
HI 45	Female	38	Cerebral hemorrhage	21.6	No	4.6	6.5	80	PDX1, H3K27ac
HI 81	Female	32	Trauma	26.9	No	11.0	8	80	MAFB
HI 87	Female	63	Cerebral bleeding	23.4	No	NA	5	75	MAFB, NKX2.2
HI 88	Female	58	Vascular complication	22.0	No	NA	9.5	80	NKX2.2
HI 101	Male	34	Anoxia	27.7	No	NA	8	85	FOXA2
HI 102	Male	64	Cerebral bleeding	24.2	No	4.8	7	80	NKX6.1
HI 118	Female	46	Cerebral bleeding	33.1	No	NA	9	75	NKX6.1

Supplementary Table 8: Summary of ChIP-Seq alignments.

Sample	Library	Read Length	# Uniquely aligned reads
Human islets (HI 32)	H3K4me1 ChIP-Seq	51 bp	73×10^6
Human islets (HI 21)	H3K4me1 ChIP-Seq	36 bp	37×10^6
Human islets (HI 25)	H3K4me1 ChIP-Seq	36 bp	38×10^6
Human islets (HI 32)	H3K4me3 ChIP-Seq*	51 bp	24×10^6
Human islets (HI 21)	H3K4me3 ChIP-Seq*	36 bp	21×10^6
Human islets (HI 25)	H3K4me3 ChIP-Seq*	36 bp	18×10^6
Human islets (HI 32)	H3K27ac ChIP-Seq	49 bp	19×10^6
Human islets (HI 45)	H3K27ac ChIP-Seq	51 bp	34×10^6
Human islets (HI 32)	FAIRE-Seq	51 bp	74×10^6
Human islets (HI 10)*	FAIRE-Seq**	35 bp	59×10^6
Human islets (HI 32)	H2A.Z ChIP-Seq	51 bp	45×10^6
Human islets (HI 22)	H2A.Z ChIP-Seq	51 bp	46×10^6
Human islets (HI 34)	H2A.Z ChIP-Seq	51 bp	43×10^6
Human islets (HI 32)	CTCF ChIP-Seq	51 bp	22×10^6
Human islets (HI 22)	CTCF ChIP-Seq	51 bp	23×10^6
Human islets (HI 34)	CTCF ChIP-Seq	51 bp	21×10^6
Human islets (HI 32)	PDX1 ChIP-Seq	51 bp	43×10^6
Human islets (HI 45)	PDX1 ChIP-Seq	51 bp	32×10^6
Human islets (HI 32)	FOXA2 ChIP-Seq	51 bp	32×10^6
Human islets (HI 101)	FOXA2 ChIP-Seq	49 bp	21×10^6
Human islets (HI 81)	MAFB ChIP-Seq	49 bp	20×10^6
Human islets (HI 87)	MAFB ChIP-Seq	49 bp	21×10^6
Human islets (HI 88)	NKX2.2 ChIP-Seq	49 bp	21×10^6
Human islets (HI 87)	NKX2.2 ChIP-Seq	49 bp	22×10^6
Human islets (HI 102)	NKX6.1 ChIP-Seq	49 bp	18×10^6
Human islets (HI 118)	NKX6.1 ChIP-Seq	49 bp	19×10^6
Human islets (HI 32)	Input DNA	51 bp	56×10^6
CD133+ Umbilical Cord Blood	MEIS1 ChIP-Seq***	36 bp	29×10^6

* Previously reported in[20]
** Previously reported in[8] as sample 3
*** Publicly available datasets GSM638314, GSM638315, GSM638316[21]

Supplementary Table 9: Frequency of transgene expression in 3 dpf injected zebrafish embryos.

Name	Nearby gene upstream (Kb from TSS)	Nearby gene downstream (Kb from TSS)	Stable line available	Domains of expression at 72 hpf						Number of replicates
				Lens	Pancreatic islet	Neurons	Hindbrain	Floor plate	Pronephric duct	
C3-1	*AGPAT9* (663)	*NKX6-1* (299)	no	**49.87%** (372/746)	0.00% (0/746)	**18.63%** (139/746)	0.00% (0/746)	0.00% (0/746)	0.00% (0/746)	4
C3-3	*ISL1* (1,108)	*PELO* (296)	yes	**94.89%** (334/352)	**38.359%** (135/352)	0.00% (0/352)	0.00% (0/352)	0.00% (0/352)	0.00% (0/352)	3
C3-4	*PROX1* (14)	*SMYD2* (277)	yes	0.00% (0/230)	**28.7%** (66/230)	0.00% (0/230)	0.00% (0/230)	0.00% (0/230)	0.00% (0/230)	3
C3-5	*LOC100128568* (88)	*RFX2* (43)	yes	0.00% (0/228)	**32.89%** (75/228)	0.00% (0/228)	**94.74%** (216/228)	**86.84%** (198/228)	**44.74%** (102/228)	3
C3-6	*TCF7L2* (27)	*HABP2* (573)	no	**41.28%** (205/496)	0.00% (0/496)	**46.57%** (231/496)	0.00% (0/496)	0.00% (0/496)	0.00% (0/496)	4
G2	*KIF26B* (990)	*SMYD3* (271)	no	0.00% (0/466)	0.00% (0/466)	0.00% (0/466)	0.00% (0/466)	0.00% (0/466)	0.00% (0/466)	4
G5	*LINC00499* (481)	*CCRN4L* (224)	no	0.00% (0/407)	0.00% (0/407)	0.00% (0/407)	0.00% (0/407)	0.00% (0/407)	0.00% (0/407)	4

Supplementary Table 10: List of oligonucleotides used in this study.

Oligonucleotides used for 4C experiments

4c_TM4SF4_DpnII	GACCACCTTTCCCAAGAGATC
4c_TM4SF4_Csp6I	TGCACATTCTGGGGTGGTA
4c_TM4SF1_DpnII	GATGGTTTACCCAGCCAGATC
4c_TM4SF1_Csp6I	AGCCAGAGCCTGCCATTAG
4c_PDX1_DpnII	CCCTGCCTGGGCATGATC
4c_PDX1_Csp6I	TCTGTGAATGCTTCAGAAGTTACC
4c_MAFB_DpnII	GCGTTCCTGTTTCTGGAGATC
4c_MAFB_Csp6I	GGTGATTTGGCCTAGAGGTG
4c_ISL1_DpnII	GCCTCTTACTTTTTGGTGGATC
4c_ISL1_Csp6I	TGCATGCTTACTGTCGGTTC
4c_GNAS_DpnII	AGCCCGGGACCTCCGATC
4c_GNAS_Csp6I	GAGGCAGACCTTGAGCTGTC
4c_G6PC2_DpnII	TGTGACTTCGGCATAATGATC
4c_G6PC2_Csp6I	CCCTCCTCTGCTGAATAGCTC
4c_C2CD4B_DpnII	CACTGAGCTTGCAACCAGATC
4c_C2CD4B_Csp6I	GAGGCGCCACCTTCAGTA
4c_C2CD4A_DpnII	GGGAACATTTCATGTCTTGATC
4c_C2CD4A_Csp6I	GGTATTAGGCAAGAATAAGACAAACC

Oligonucleotides used for cloning into TOPO vector (for Luciferase assays)

Cis element	Forward primer (5' to 3')	Reverse primer (5' to 3')
C3-1	CACCAGAGAACAACAGGCAGGT	TCATTGAACTTCCCCGATGG
C3-2	CACCCAAAACACCTTTGAAAAACA	CCCCATCTCATACTATTGGGGTTC
C3-7	CACCTGAGAGCGGATTTTACAGAT	AGCAAAAGCTGCCATGCCTA
C3-8	CACCGAGCATGGTGAGATGGGTTT	GCTTTGACCCCAGGGGGAGACC
C3-9	CACCACATCCCTTACCCTTACTGGA	GGCAATGCGGGCTCTTTTT
C3-10	CACCTTCATGTTTCCCCCGTATGT	TCCTGCCCCAAGTTGCACAG
C3-11	CACCCCATCACCTGTGACCTTTT	ACAGGGGCTGGGCAACATTC
C3-12	CACCCTGTGGACAGGTCCCATTCT	CTTGGGCCCTGCCCTTCTCA
C3-13	CACCGGAGGTCAGAGTTGCCTACA	CATGGATTCAGGCCATTCCTGTCA
C3-14	CACCGTGTGGGAGAAAAGTCTTCA	TGCCTTGTGAATGGCAGAGGAG
C3-15	CACCGAAGGGCAGAGAGGAGCA	TCTTCTGTGGCACATGGTGGGCTA
C3-16	CACCGCGCTGTATGCAAAGTGAGG	TGTCCCCAAAACTGCTCCCACA
C2_1	CACCGGAAATGTATCAAGATTCCAG	CGGTTACGTATGTCATCAGATTGAGC
C2_2	CACCAGTGGGTTGTTTGTGTTTTT	TTGCTATGGTTAAGGTAAATGGCACA
C2_3	CACCCATCTGAGCTTCACTGGATT	TGAGTTTTGAAGTCAGACAGACCTGGA
C2_4	CACCCGAAGGAGTCTGTTGTCAC	GCCCCTAAAATCAAGGCAATTGAG
C2_5	CACCTTCTGCAATGTTTAATCTGC	GGTCTCAGCTCAGTGGTAGGG
C5_1	CACCAGCCCTTGGTCAAGAATAGT	TGTAAGCGGTTTGGGCAATCTT
C5_2	CACCATGGTCGGATATGACAGTTT	CCCATGGGAATAAGGCCTGTAGA
C5_3	CACCGACAGTATAAGTGATTTGTCAGGA	CCAACCCCTCCCTGGAGTTAG

Oligonucleotides used for cloning contructs for zebrafish

Cis element	Forward primer (5' to 3')	Reverse primer (5' to 3')
C3-1	TGCAGTCACATGCACAAAG	AGAAACTAGGGCTGTGTTTA
C3-3	TTAAGGTCCCTCTGCCATGT	AACTCTTCCCAAGCCTCATT
C3-4	AATTTTCTTCCTCCGCTTTC	CATTCCCTTTAATATCCCATGC
C3-5	GAAAAGCGCTCCAGAAATTG	AGTTCCCTTTGCACTTGTT
C3-6	CCAAGGCTTGAAAATGGATG	AGAGCTTTTTCTAGGCCTCC
Fugu rubripes control region	GTGTGTCATCCTCATCCACG	CATTCCATGATGGTGCTCTG
Zebrafish hsp70 promoter	TTGATTGGTCGAACATGCTG	CAGTCCGCTCGCTGTCTC
M13 universal primers	TGTAAAACGACGGCCAGT	CAGGAAACAGCTATGACC
attB3/attB5 Gateway adaptors	ATAAAGTAGGCT	CAAAAGTTGGGT
attB3/attB5 Gateway primers	GGGGACAAGTTTGTATAATAAAGTAGGCT	GGGGACCACTTTGTATACAAAAGTTGGGT

Supplementary Table 10 (continued)

Oligonucleotides used for EMSA

Probe	Forward (5' to 3')	Reverse (5' to 3')
rs58692659 G	GATCGCCGGGAGAACAGATGGCAGCTGC	GATCGCAGCTGCCATCTGTTCTCCCGGC
rs58692659 A	GATCGCCGGGAGAATAGATGGCAGCTGC	GATCGCAGCTGCCATCTATTCTCCCGGC
NEUROD1 consensus	GATCAGCCCCCAGCCATCTGCCGACCCC	GATCGGGGTCGGCAGATGGCTGGGGGCT
rs72695654 T	GATCGTATCTGTAGTGACCTTTAGCTTTTT	GATCAAAAAGCTAAAGGTCACTACAGATAC
rs72695654 G	GATCGTATCTGTAGGGACCTTTAGCTTTTT	GATCAAAAAGCTAAAGGTCCCTACAGATAC
NBRE consensus	GATCGCGCAGGTCAAAGGTCACCTC	GATCGAGGTGACCTTTGACCTGCGC

MIR Based Knockdown inserts

Flanking sequences for MIR recognition

5´MIR	CTGGAGGCTTGCTGAAGGCTGTA
3´ MIR	CAGGACACAAGGCCTGTTACTAGCACTCACATGGAACAAATGGCC

Target	Insert Sequence (AS-target-loop-mismatched target)
MAFB sh1	TGCTGTTCACGTCGAACTTGAGCAGGGTTTTGGCCACTGACTGACCCTGCTCATTCGACGTGAA
MAFB sh2	TGCTGAACTGATGAGATTTGGGCGCCGTTTTGGCCACTGACTGACGGCGCCCATCTCATCAGTT
Non-Targetting 1	TGCTGAAATGTACTGCGCGTGGAGACGTTTTGGCCACTGACTGACGTCTCCACGCAGTACATTT
Non-Targetting 2	TGCTGaaTTGaTGTGTTTaGTCGCTaGTTTTGGCCACTGACTGACtAGCGACtACACAtCAAtt
Non-Targetting 3	TGCTGAATCTTTGGACGCAAATCTGGGTTTTGGCCACTGACTGACCCAGATTTGTCCAAAGATT
Non-Targetting 4	TGCTGTTAACTGCTAGGCCAACCCATGTTTTGGCCACTGACTGACATGGGTTGCTAGCAGTTAA

Supplementary References:

1 Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**, 576-89 (2010).
2. McLean, C.Y. et al. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* **28**, 495-501 (2010).
3. Blechinger, S.R. et al. The heat-inducible zebrafish hsp70 gene is expressed during normal lens development under non-stress conditions. *Mech Dev* **112**, 213-5 (2002).
4. Ernst, J. et al. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**, 43-9 (2011).
5. Chambers, J.M., Cleveland, W.S., Kleiner, B. & Tukey, P.A. Graphical Methods for Data Analysis (vol 17, pg 180, 1983).
6. Morris, A.P. et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet* **44**, 981-90 (2012).
7. Scott, R.A. et al. Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. *Nat Genet* **44**, 991-1005 (2012).
8. Gaulton, K.J. et al. A map of open chromatin in human pancreatic islets. *Nat Genet* **42**, 255-9 (2010).
9. Bernstein, B.E. et al. The NIH Roadmap Epigenomics Mapping Consortium. *Nat Biotechnol* **28**, 1045-8 (2010).
10. Jonsson, J., Carlsson, L., Edlund, T. & Edlund, H. Insulin-promoter-factor 1 is required for pancreas development in mice. *Nature* **371**, 606-9 (1994).
11. Dutta, S., Bonner-Weir, S., Montminy, M. & Wright, C. Regulatory factor linked to late-onset diabetes? *Nature* **392**, 560 (1998).
12. Ahlgren, U., Jonsson, J., Jonsson, L., Simu, K. & Edlund, H. beta-cell-specific inactivation of the mouse Ipf1/Pdx1 gene results in loss of the beta-cell phenotype and maturity onset diabetes. *Genes Dev* **12**, 1763-8 (1998).
13. Stoffers, D.A., Ferrer, J., Clarke, W.L. & Habener, J.F. Early-onset type-II diabetes mellitus (MODY4) linked to IPF1. *Nat Genet* **17**, 138-9 (1997).
14. Sander, M. et al. Homeobox gene Nkx6.1 lies downstream of Nkx2.2 in the major pathway of beta-cell formation in the pancreas. *Development* **127**, 5533-40 (2000).
15. Sund, N.J. et al. Tissue-specific deletion of Foxa2 in pancreatic beta cells results in hyperinsulinemic hypoglycemia. *Genes Dev* **15**, 1706-15 (2001).
16. Lee, C.S. et al. Foxa2 controls Pdx1 gene expression in pancreatic beta-cells in vivo. *Diabetes* **51**, 2546-51 (2002).
17. Artner, I. et al. MafB is required for islet beta cell maturation. *Proc Natl Acad Sci U S A* **104**, 3853-8 (2007).
18. Zankl, A. et al. Multicentric carpotarsal osteolysis is caused by mutations clustering in the amino-terminal transcriptional activation domain of MAFB. *Am J Hum Genet* **90**, 494-501 (2012).
19. Sussel, L. et al. Mice lacking the homeodomain transcription factor Nkx2.2 have diabetes due to arrested differentia tion of pancreatic beta cells. *Development* **125**, 2213-21 (1998).
20 Moran, I. et al. Human beta cell transcriptome analysis uncovers lncRNAs that are tissue-specific, dynamically regulated, and abnormally expressed in type 2 diabetes. *Cell Metab* **16**, 435-48 (2012).
21. Novershtern, N. et al. Densely interconnected transcriptional circuits control cell states in human hematopoiesis. *Cell* **144**, 296-309 (2011).