

Supplementary Information

A map of open chromatin in human pancreatic islets

Kyle J. Gaulton¹, Takao Nammo^{2,3}, Lorenzo Pasquali^{2,3}, Jeremy M. Simon^{1,4}, Paul G. Giresi⁴, Marie P. Fogarty¹, Tami M. Panhuis¹, Piotr Mieczkowski¹, Antonio Secchi⁵, Domenico Bosco⁶, Thierry Berney⁶, Eduard Montanya^{3,7}, Karen L. Mohlke^{1,8}, Jason D. Lieb^{4,8}, Jorge Ferrer^{2,3,9}

¹Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA.

²Genomic Programming of Beta Cells, Institut d'Investigacions Biomediques August Pi i Sunyer, 08036 Barcelona, Spain

³CIBER de Diabetes y Enfermedades Metabólicas Asociadas (CIBERDEM), 08036 Barcelona, Spain

⁴Department of Biology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA

⁵Clinical Transplant Unit, San Raffaele Scientific Institute, 20132 Milano, Italy

⁶Cell Isolation and Transplantation Center, CH-1211, Geneva, Switzerland

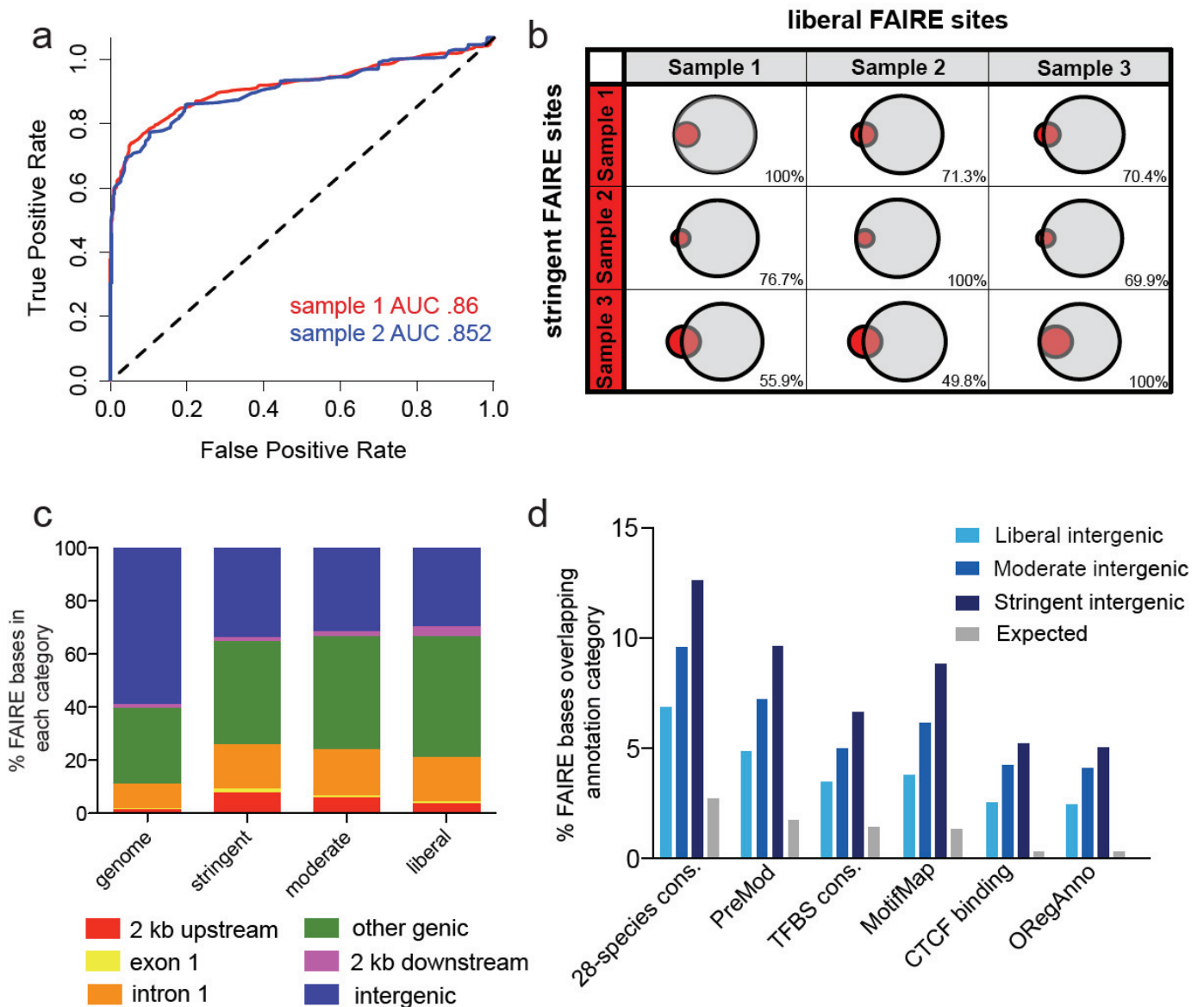
⁷Laboratory of Diabetes and Experimental Endocrinology, Endocrine Unit, IDIBELL-Hospital Universitari Bellvitge, University of Barcelona, Spain

⁸Carolina Center for Genome Sciences and Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA.

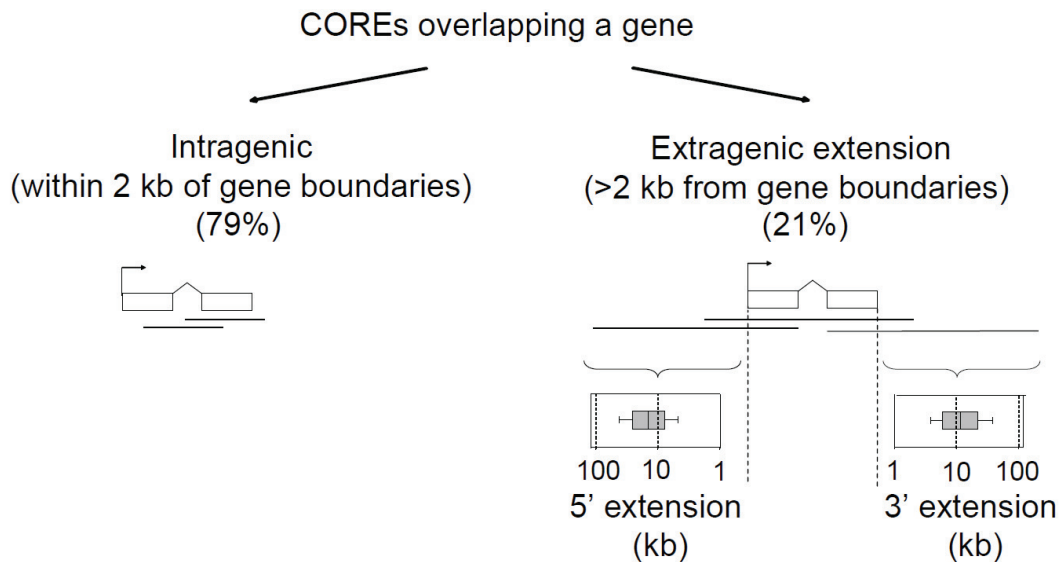
⁹Department of Endocrinology, Hospital Clínic de Barcelona, 08036 Barcelona, Spain

Table of Contents

Supplementary Figure 1. Characteristics of FAIRE-Seq in human pancreatic islets	Page 2
Supplementary Figure 2. Distribution of islet-selective COREs relative to gene boundaries	Page 3
Supplementary Figure 3. Long-range regulatory maps of selected loci	Page 4
Supplementary Figure 4. Additional functional analysis of the genomic region surrounding rs7903146	Page 12
Supplementary Table 1. Donor profiles of islet samples	Page 13
Supplementary Table 3. Referenced list of genes with preferential islet FAIRE enrichment that are expressed selectively in islets	Page 14
Supplementary Table 6. Functional annotation of genes overlapping islet COREs	Page 16
Supplementary Table 8. Islet FAIRE enrichment at T2D susceptibility and fasting glycemia loci	Page 17
Supplementary Table 9. Oligonucleotides for genomic region surrounding rs7903146	Page 18

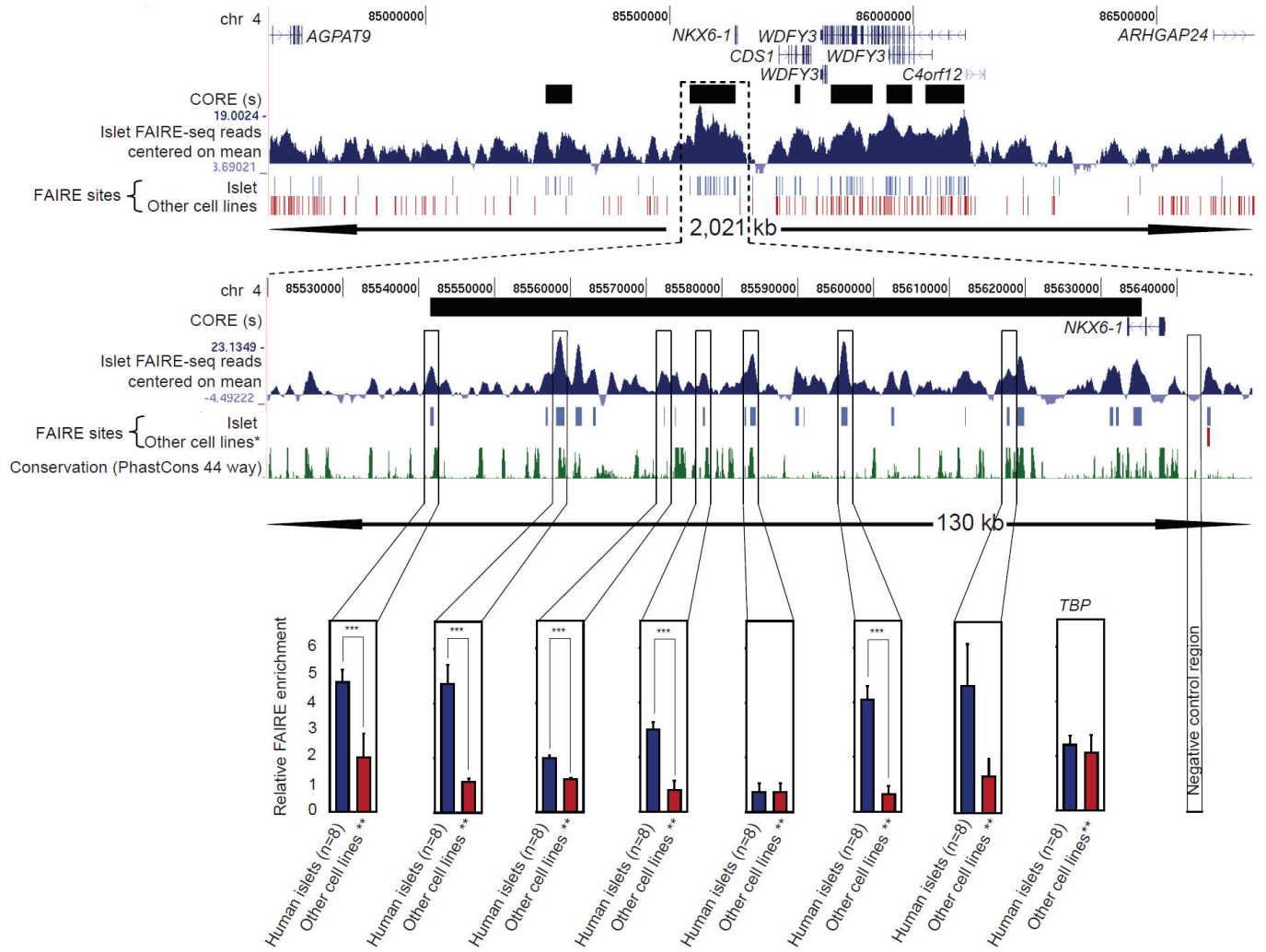


Supplementary Figure 1. Characteristics of FAIRE-Seq in human pancreatic islets. (a) Comparison of FAIRE-chip and FAIRE-seq. Receiver Operating Characteristic (ROC) curve analysis using positive regions called from ENCODE tiling DNA microarrays at a stringent threshold ($P < 1 \times 10^{-12}$) for both samples 1 and 2, and negative regions called from contiguous regions of low intensity. The percentage of positive and negative regions captured at increasing sequence read density thresholds was then plotted. The sequence data from both samples 1 and 2 captures positive tiling array data while rejecting negative tiling array data at a much higher rate than random. (b) Comparison of FAIRE-seq regions across three samples. Stringent islet FAIRE sites (in columns) from all three samples were compared against liberal islet FAIRE sites (in rows). Overlap is reported as the percentage of stringent sites that are captured with liberal sites from another sample. (c) Genome-wide distribution of islet FAIRE sites. Genomic regions enriched for FAIRE are preferentially located upstream and through the body of known genes. More than 25% of FAIRE regions at most stringent enrichment threshold are located proximal to known transcription start sites and in first introns. (d) Intergenic open chromatin sites are enriched for functional annotations. FAIRE sites significantly ($P < 0.001$) overlap sequence conservation (28-species most conserved elements), transcription factor binding sites (TFBS) (TFBS cons. and MotifMap), predicted regulatory modules (PreMod) and the ORegAnno database of regulatory elements compared to random regions. Stringent regions are the most enriched across all functional annotations.



Supplementary Figure 2. Distribution of islet-selective COREs relative to gene boundaries. In most islet-selective COREs that are associated with genes, the entire CORE is located within 2 kb of gene boundaries. However, 21% of islet-selective COREs that are associated with genes extend further than 2 kb from the transcription start site, termination site, or both. Among these COREs that extend far away from gene boundaries, in 54% of the cases the extension is greater in the 5' rather than in the 3' direction of the gene, whereas in 46% the extension is greater towards the 3' end of the gene. The box plots represent the median (line), 25-75th (shaded) and 5-95th percentile (whiskers) of the sizes of COREs extending in either 5' or 3' directions from genes.

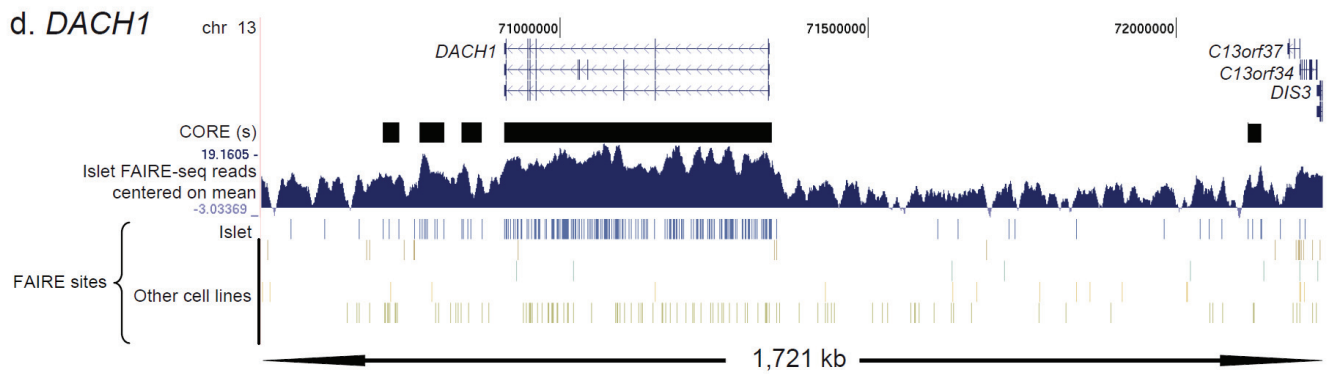
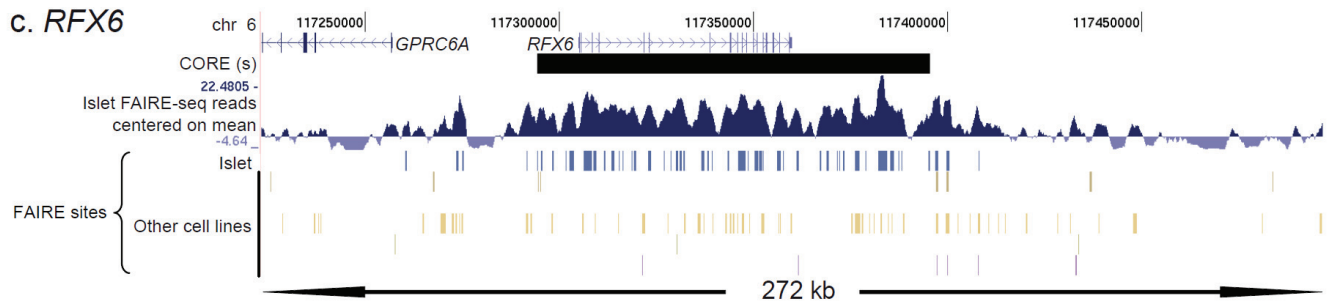
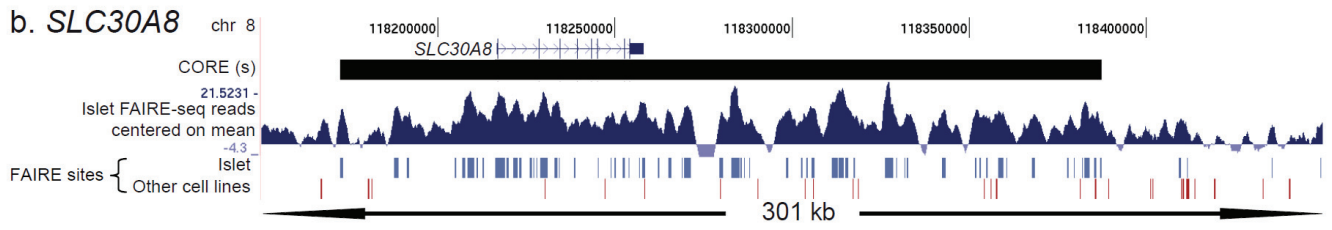
a. *NKX6-1*

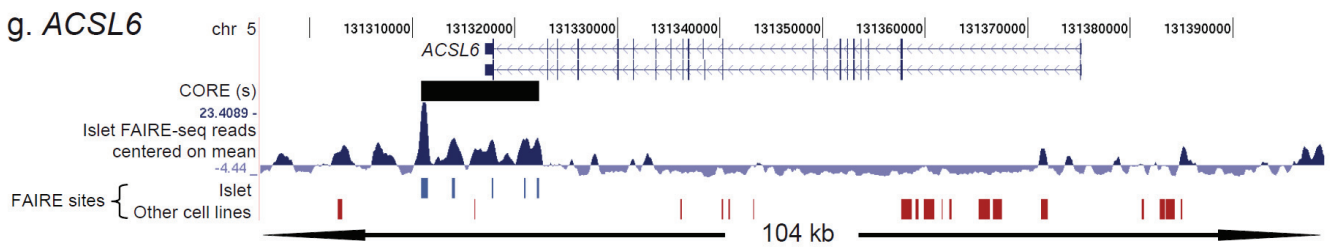
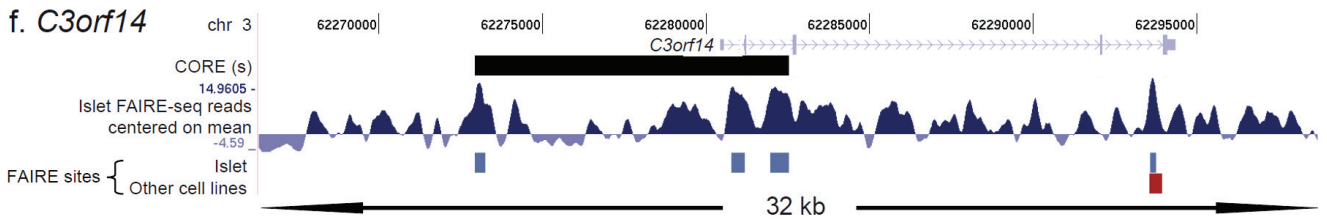
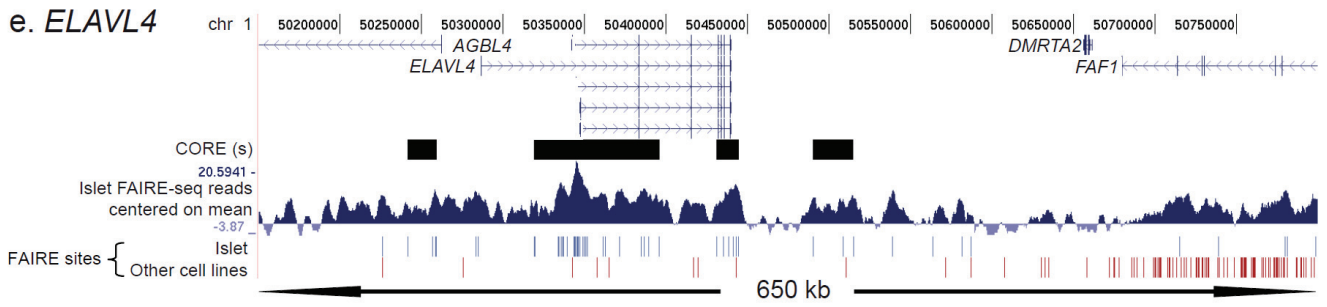


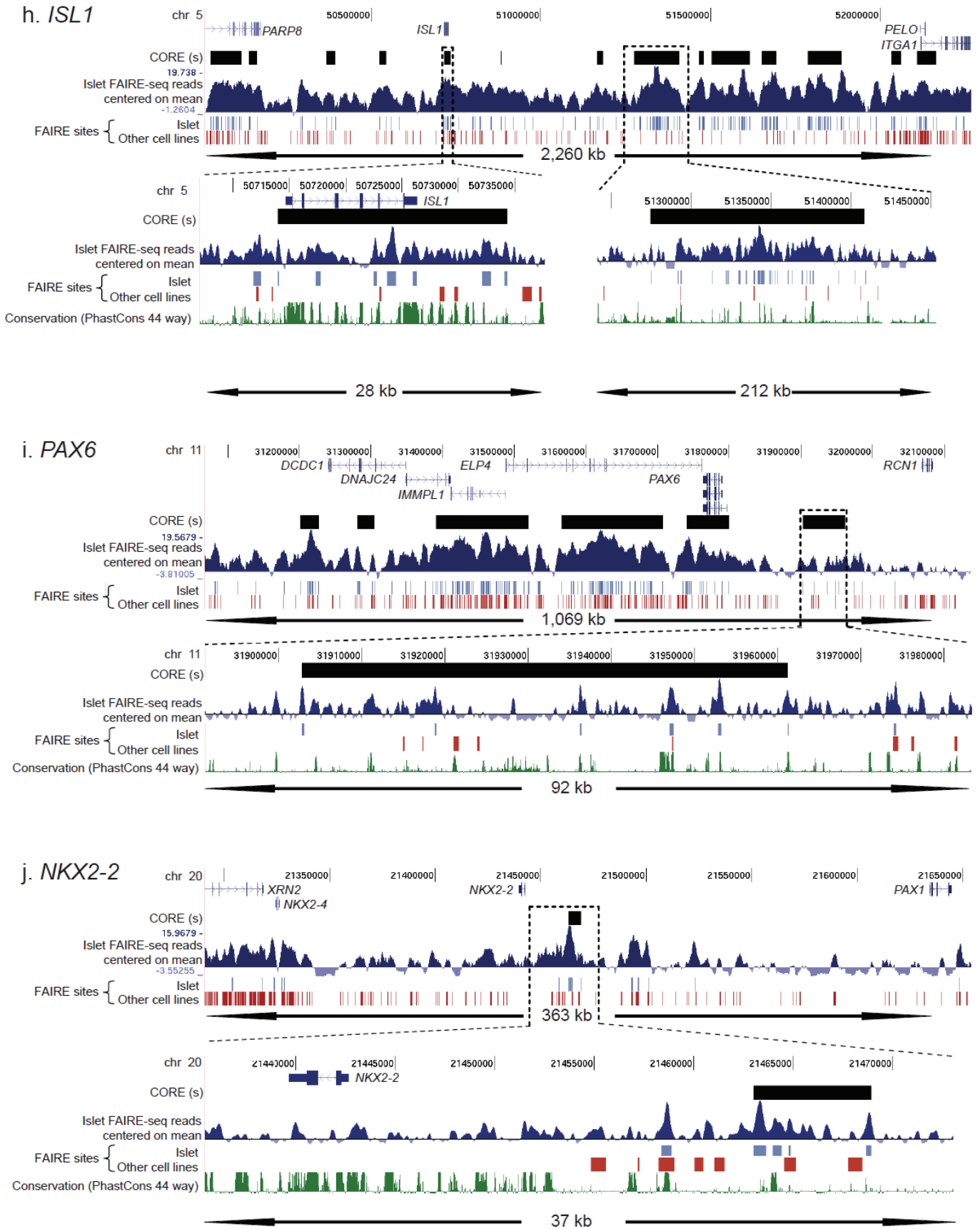
* HeLa-S3, HUVEC, GM12878, HepG2 and K562

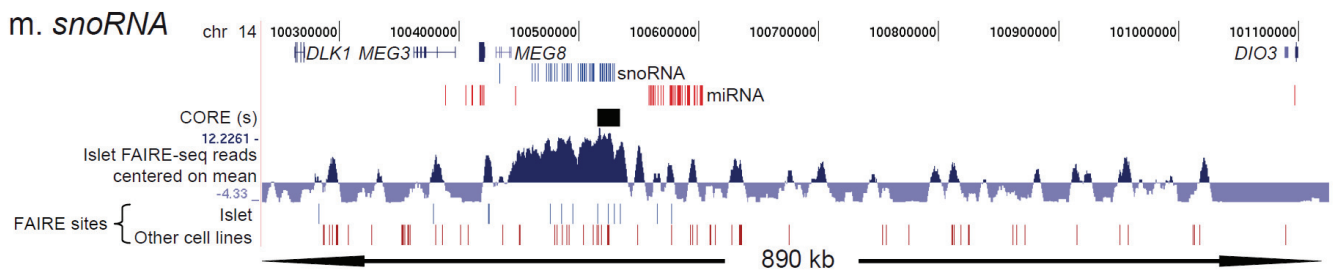
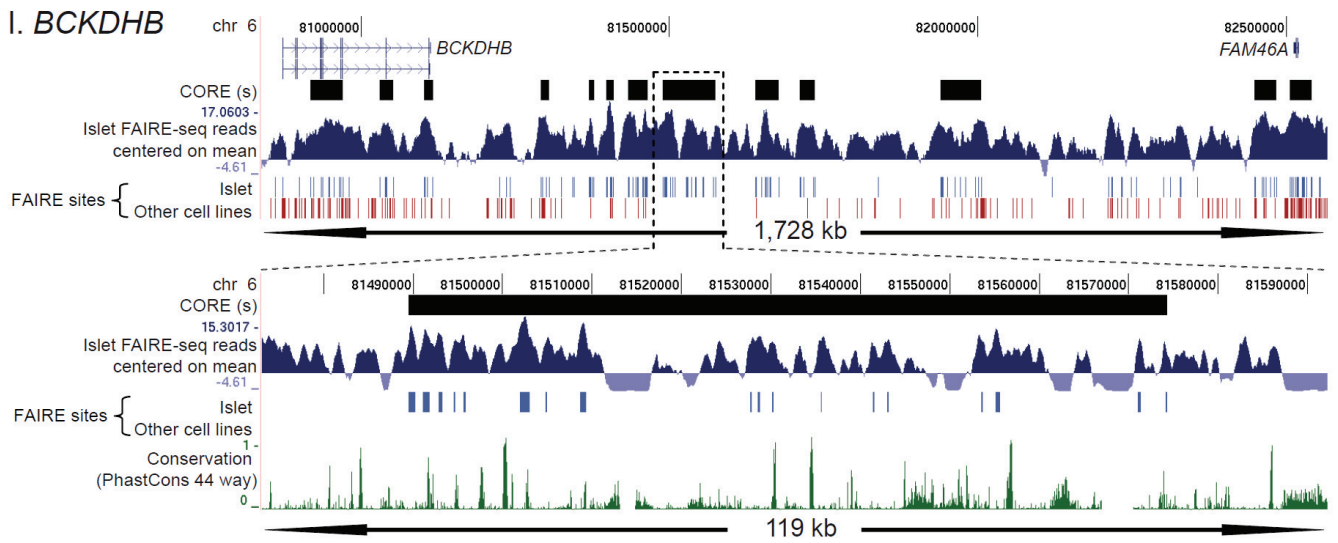
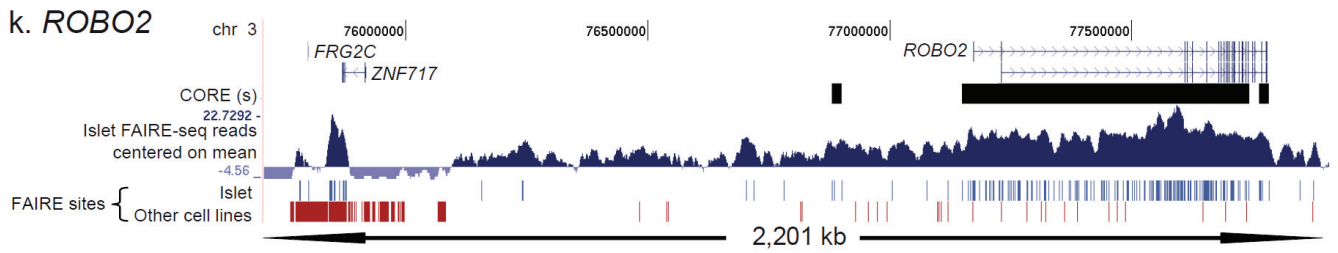
** SW480, C33A, HEK293T and MCF-7

*** p<0.05

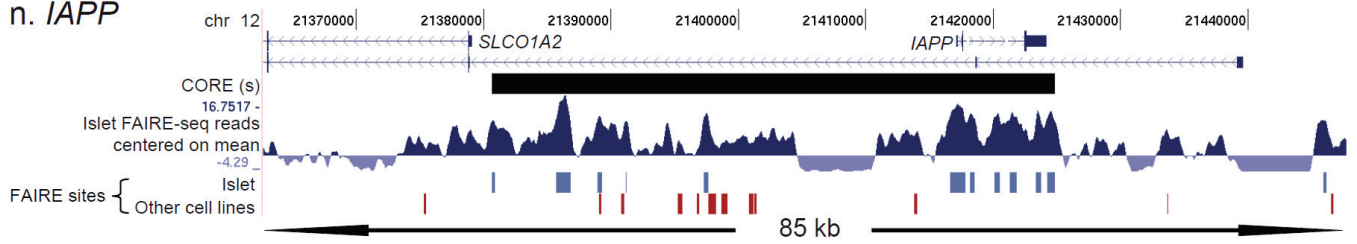




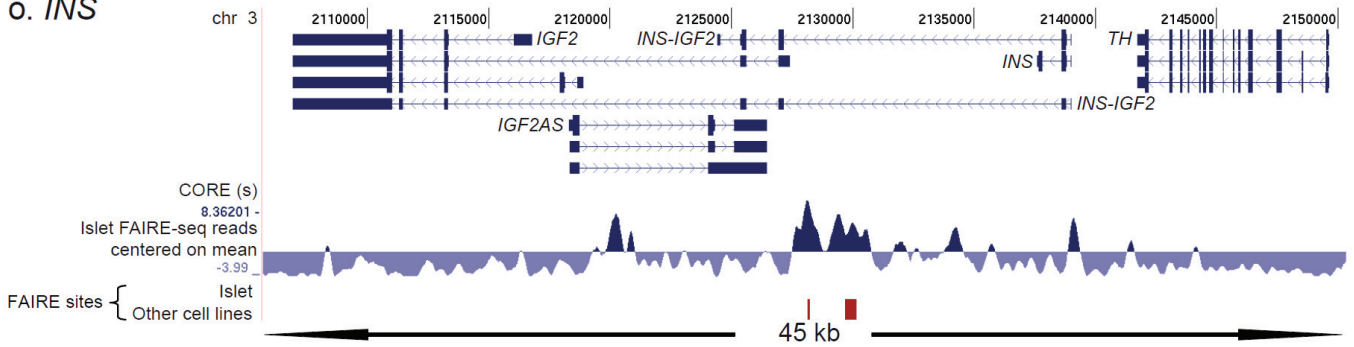




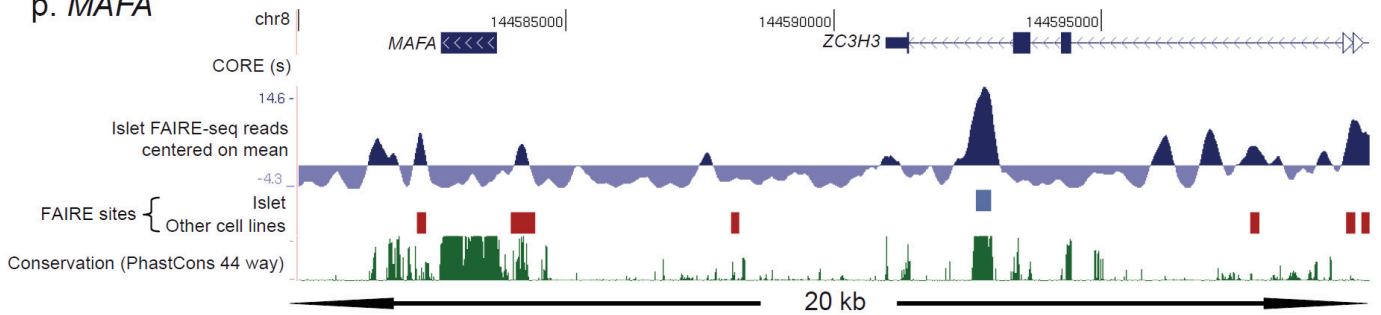
n. *IAPP*



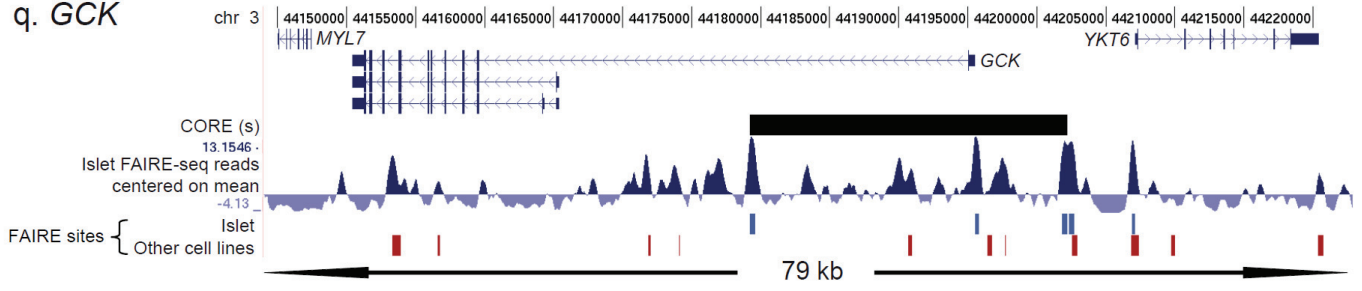
o. *INS*



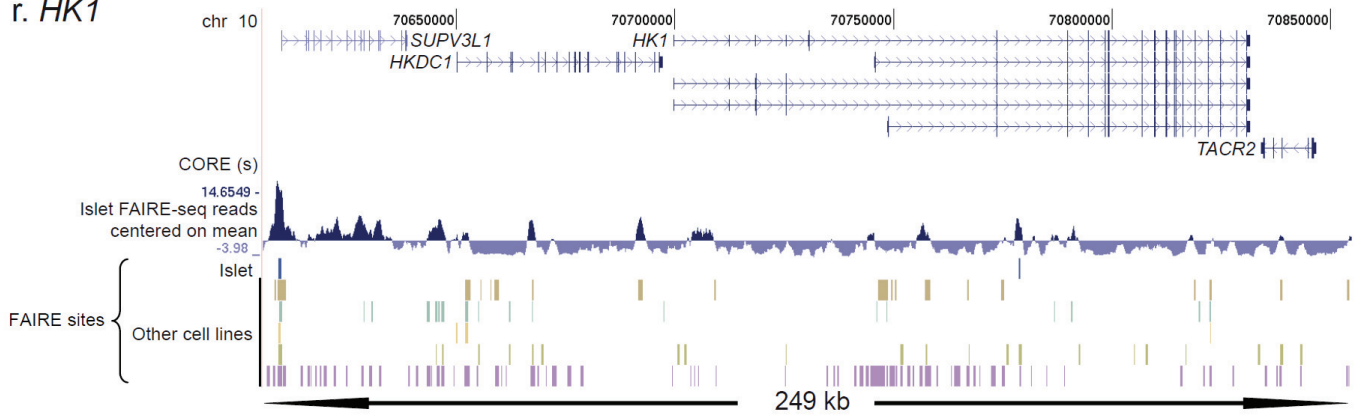
p. *MAFA*



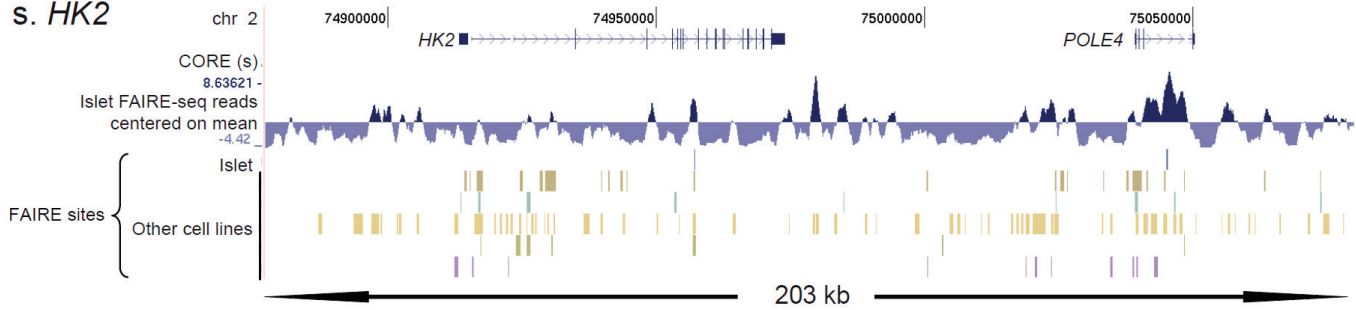
q. *GCK*



r. *HK1*



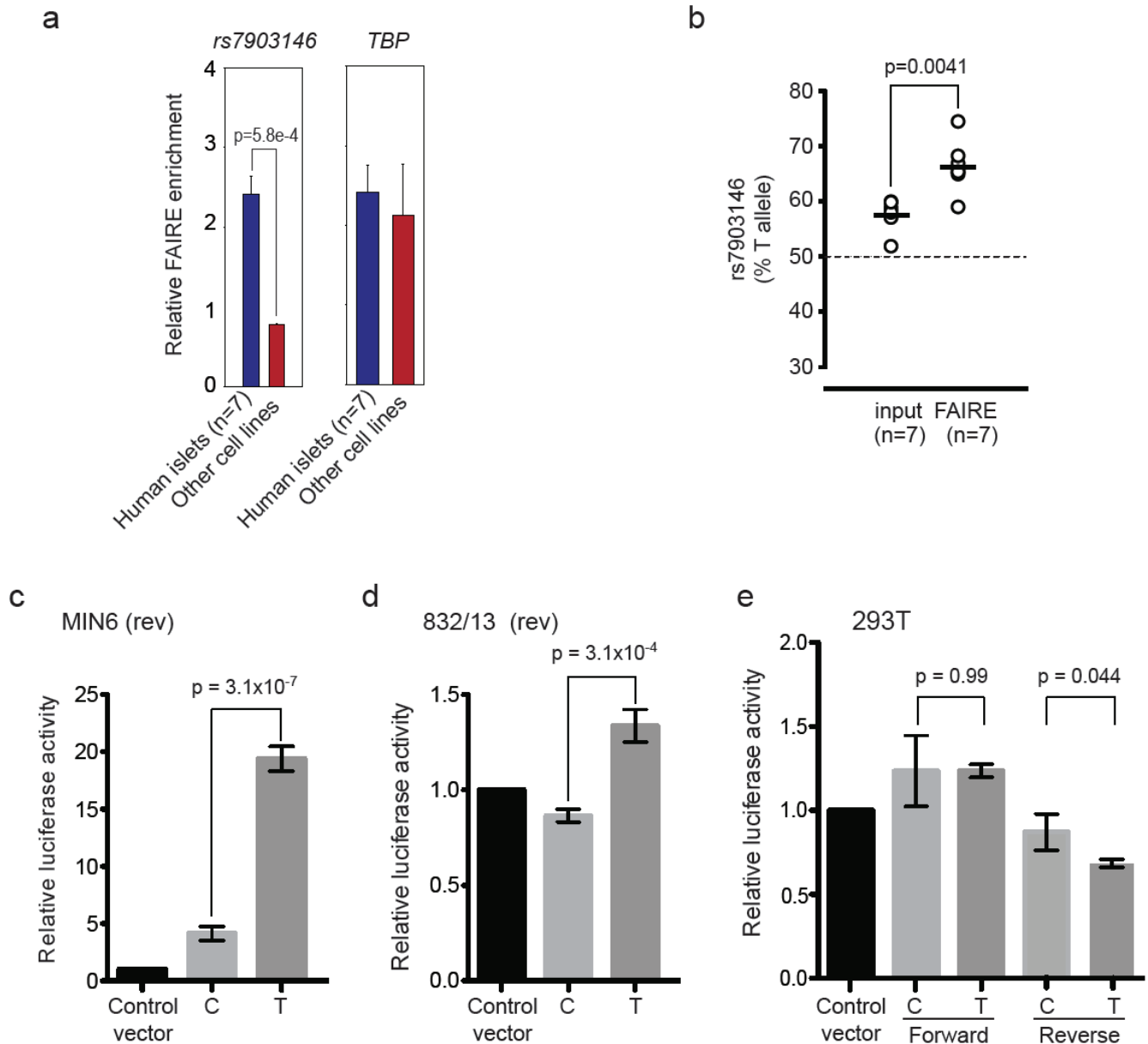
s. *HK2*



Supplementary Figure 3. Long-range regulatory maps of selected loci. Islet-selective COREs are depicted as black horizontal stripes and labeled as COREs. Blue vertical lines are moderate stringency FAIRE sites in islets, red vertical lines are FAIRE sites present in any of the 5 non-islet cells. When FAIRE sites are observed in only one non islet-cell line, we separate non-islet cell line tracks to show the islet-cell selectivity of the CORE. For some loci we show zoomed images with tracks depicting evolutionary conserved sequences (PhastCons 44-way alignments). (a) A CORE spanning 94 kb located in the 3' region of *NKX6.1*. The lower panel shows qPCR verifications of islet-selective FAIRE enrichment in 6/7 sites from this CORE in 8 additional human islet samples and 4 additional non-islet cell lines. No differences in enrichment between the two groups of samples were found at the TBP promoter region. FAIRE enrichment was normalized to a local negative control region located 5' of *NKX6.1* that lacks apparent enrichment in islet FAIRE-Seq. P values correspond to non-paired t tests, error bars are S.E.M. (b-d) COREs overlapping *SLC30A8*, *RFX6*, and *DACHI*. (e-g) COREs extending 5' and/or 3' of *ELAVL4*, *C3orf14* and *ACSL6*. (h,i) COREs overlapping *ISL1* and *PAX6*. Distant 5' or 3' COREs are also present and shown as zoomed images. (j) A CORE in a distant region 5' of *NKX2.2*. (k) A 602 kb CORE overlapping *ROBO2*. (l) A CORE located at a gene desert near *BCKDHB*. (m) A CORE in a noncoding RNA cluster on chromosome 14. (n) A CORE at *IAPP* with a 30-kb 5' extension. (o) Islet FAIRE-seq map of the *INS* locus, showing a small accumulation of islet FAIRE-seq reads surrounding the *INS* promoter. (p) Islet FAIRE-Seq map of *MAFA*, showing an islet-selective site at the previously reported distant upstream evolutionary conserved *MAFA* enhancer sequence located in an intronic region of *ZC3H3*¹ (q,r,s) *GCK*, encoding for the β -cell low-affinity hexokinase, exhibits a CORE at the islet-cell promoter, whereas *HK1* and *HK2*, encoding for high-affinity hexokinases that are inactive in β -cells to enable *GCK* function, are devoid of FAIRE sites in islets but not in other cell types².

References

1. Raum, J.C. et al. FoxA2, Nkx2.2, and PDX-1 regulate islet beta-cell-specific *mafA* expression through conserved sequences located between base pairs -8118 and -7750 upstream from the transcription start site. *Mol Cell Biol* **26**, 5735-43 (2006).
2. Schuit, F., Moens, K., Heimberg, H. & Pipeleers, D. Cellular origin of hexokinase in pancreatic islets. *J Biol Chem* **274**, 32803-9 (1999).



Supplementary Figure 4. Additional functional analysis of the genomic region surrounding rs7903146 (a) Confirmation by qPCR of islet-cell selective FAIRE enrichment in the region surrounding rs7903146. FAIRE enrichment in 7 human islet samples and four non islet cell lines (SW480, C33A, HEK293T and MCF-7) was normalized to the negative control region 5' of NKX6.1 described in Supplementary Figure 3a. Similar levels of enrichment were found at the TBP promoter region in the two groups of cells. P values correspond to non-paired t tests, error bars are S.E.M. (b) Sanger sequencing of input and FAIRE DNA surrounding rs7903146 from human islets from 7 heterozygous individuals. Area under the sequence curves of each allele was quantified using ImageJ (input: 57.5 +/- 2.7% T allele, FAIRE: 66.2 +/- 4.6% T allele). (c-e) Luciferase reporter data for (c,d) reverse orientation in MIN6 and 832/13 cell lines, and (e) both orientations for 293T cells; standard deviations represent four independent clones for each allele.

Supplementary Table 1. Donor profiles of islet samples

	Sample 1	Sample 2	Sample 3
Ethnicity	European descent	European descent	N/A
Sex	Male	Male	Male
Age (years)	26	53	53
BMI (kg/m ²)	27.3	32.7	23.5
Premortem non fasting glycemia (mmol/l)	NA	6.9 mmol/l	6.4 mmol/l
Premortem diagnosis of diabetes	No	No	No
Cause of death	Head trauma	Cerebral hemorrhage	Cerebral hemorrhage
Cold ischemia (hours)	9	18	6.5
Islet purity (%)	55	55	85
Culture duration before shipment (days)	4	0	3
3-day reculture after arrival	No	No	Yes
In vitro insulin secretory response	NA	2.8 mM glucose: 0.21 ng/islet/60 min* 20 mM glucose: 2.33 ng/islet/60 min	2.8 mM glucose: 3.44 ± 0.52 ng/10 islets/30 min** 16.8 mM glucose: 6.83 ± 1.68 ng/10 islets/30min

NA: Not available

* Static incubation insulin release studies performed as described in Téllez N, Montolio M, Biarnés M, Castaño E, Soler J, Montanya E. Adenoviral overexpression of interleukin-1 receptor antagonist protein increases beta-cell replication in rat pancreatic islets. Gene Ther. 2005 Jan;12(2):120-8. PMID: 15578044

** Static incubation insulin release studies performed as described by Dr Pierre Maechler, University of Geneva Medical Center. Performed as described in Diabetologia. 2009 Nov 12. [Epub ahead of print]; PMID: 19908022

Supplementary Table 3. Referenced list of genes with preferential islet FAIRE enrichment that are expressed selectively in islets

Gene Symbol	Gene name	References
<i>IAPP</i>	Islet amyloid polypeptide	Nishi, M et al. . Islet amyloid polypeptide. A new beta cell secretory product related to islet amyloid deposits. J Biol Chem 265, 4173-6 (1990).
<i>NKX6-1</i>	NK6 homeobox 1	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).
<i>FOXD3</i>	Forkhead box D3	Perera, H.K. et al. Expression and shifting subcellular localization of the transcription factor, Foxd3, in embryonic and adult pancreas. Gene Expr Patterns 6, 971-7 (2006).
<i>PDX1</i>	Pancreatic duodenal homeobox 1	Ohlsson, H et al. IPF1, a homeodomain-containing transactivator of the insulin gene. EMBO J 12, 4251-9 (1993). Stoffers, D.A. et al. Early-onset type-II diabetes mellitus (MODY4) linked to IPF1. Nat Genet 17, 138-9 (1997).
<i>SULT4A1</i>	Sulfotransferase family 4A, member 1	Falany, C.N., et al. Molecular cloning and expression of novel sulphotransferase-like cDNAs from human and rat brain. Biochem J 346 Pt 3, 857-64 (2000).
<i>SCG2</i>	Secretogranin II	Karlsson, E. The role of pancreatic chromogranins in islet physiology. Curr Mol Med 1, 727-32 (2001).
<i>GAD2</i>	GAD65, Glutamate decarboxylase 2	Kaufman, D.L. et al. Autoimmunity to two forms of glutamate decarboxylase in insulin-dependent diabetes mellitus. J Clin Invest 89, 283-92 (1992).
<i>NEUROD1</i>	Neurogenic differentiation 1	Naya, F.J. et al. Diabetes, defective pancreatic morphogenesis, and abnormal enteroendocrine differentiation in BETA2/neuroD-deficient mice. Genes Dev 11, 2323-34 (1997). Malecki, M.T. et al. Mutations in NEUROD1 are associated with the development of type 2 diabetes mellitus. Nat Genet 23, 323-8 (1999).
<i>ISL1</i>	ISL LIM homeobox 1	Ahlgren, U. et al. Independent requirement for ISL1 in formation of pancreatic mesenchyme and islet cells. Nature 385, 257-60 (1997).
<i>DACH1</i>	Dachshund homolog 1 (Drosophila)	Miyatsuka, T., et al. Chronology of islet differentiation revealed by temporal cell labeling. Diabetes 58, 1863-8 (2009).
<i>ST18</i>	Suppression of tumorigenicity 18	Wang, S. et al. Loss of Myt1 function partially compromises endocrine islet cell differentiation and pancreatic physiological function in the mouse. Mech Dev 124, 898-910 (2007).
<i>NCAM1</i>	Neural cell adhesion molecule 1	Esni, F. et al. Neural cell adhesion molecule (N-CAM) is required for cell type segregation and normal ultrastructure in pancreatic islets. J Cell Biol 144, 325-37 (1999).
<i>CDK5R2</i>	Cyclin-dependent kinase 5, regulatory subunit 2 (p39)	Lilja, L. et al. Cyclin-dependent kinase 5 associated with p39 promotes Munc18-1 phosphorylation and Ca(2+)-dependent exocytosis. J Biol Chem 279, 29534-41 (2004).
<i>RFXDC1</i>	regulatory factor X, 6, RFX6	Miyatsuka, T., et al. Chronology of islet differentiation revealed by temporal cell labeling. Diabetes 58, 1863-8 (2009).
<i>PAX6</i>	Paired box 6	Sander, M. et al. Genetic analysis reveals that PAX6 is required for normal transcription of pancreatic hormone genes and islet development. Genes Dev 11, 1662-73 (1997).
<i>ABCC8</i>	ATP-binding cassette, sub-family C(CFTR/MRP),8	Aguilar-Bryan, L. et al. Cloning of the beta cell high-affinity sulfonylurea receptor: a regulator of insulin secretion. Science 268, 423-6 (1995).
<i>GLP1R</i>	Glucagon-like peptide 1 receptor	Thorens, B. Expression cloning of the pancreatic beta cell receptor for the gluco-incretin hormone glucagon-like peptide 1. Proc Natl Acad Sci U S A 89, 8641-5 (1992).

Gaulton, Nammo, Pasquali et al. Supplementary Table 3

Gene Symbol	Gene name	References
<i>CPE</i>	Carboxypeptidase E	Naggert, J.K. et al. Hyperproinsulinaemia in obese fat/fat mice associated with a carboxypeptidase E mutation which reduces enzyme activity. <i>Nat Genet</i> 10, 135-42 (1995).
<i>LMX1A</i>	LIM homeobox transcription factor 1, alpha	Iannotti, C.A. et al. Identification of a human LMX1 (LMX1.1)-related gene, LMX1.2: tissue-specific expression and linkage mapping on chromosome 9. <i>Genomics</i> 46, 520-4 (1997).
<i>KCNK10</i>	Potassium channel, subfamily K, member 10	Kang, D., et al. Functional expression of TREK-2 in insulin-secreting MIN6 cells. <i>Biochem Biophys Res Commun</i> 323, 323-31 (2004).
<i>G6PC2</i>	Glucose-6-phosphatase, catalytic, 2	Hutton, J.C. & Eisenbarth, G.S. A pancreatic beta-cell-specific homolog of glucose-6-phosphatase emerges as a major target of cell-mediated autoimmunity in diabetes. <i>Proc Natl Acad Sci U S A</i> 100, 8626-8 (2003).
<i>KCNB2</i>	Potassium voltage-gated channel, Shab-related subfamily, member 2	Wolf-Goldberg, T. et al. Target soluble N-ethylmaleimide-sensitive factor attachment protein receptors (t-SNAREs) differently regulate activation and inactivation gating of Kv2.2 and Kv2.1: Implications on pancreatic islet cell Kv channels. <i>Mol Pharmacol</i> 70, 818-28 (2006).
<i>SYT4</i>	Synaptotagmin IV	Gauthier, B.R. et al. Synaptotagmin VII splice variants alpha, beta, and delta are expressed in pancreatic beta-cells and regulate insulin exocytosis. <i>FASEB J</i> 22, 194-206 (2008).
<i>SLC30A8</i>	Solute carrier family 30 (zinc transporter), member 8	Sladek, R. et al. A genome-wide association study identifies novel risk loci for type 2 diabetes. <i>Nature</i> 445, 881-5 (2007). Nicolson, T.J. et al. Insulin storage and glucose homeostasis in mice null for the granule zinc transporter ZnT8 and studies of the type 2 diabetes-associated variants. <i>Diabetes</i> (2009).

Supplementary Table 6. Functional annotation of genes overlapping islet COREs*

PANTHER Biological Process	P-Value	GO Biological Process	P-Value
BP00044:mRNA transcription regulation	2.25E-24	GO:0051056~regulation of small GTPase mediated signal transduction	5.72E-08
BP00071:Proteolysis	1.45E-22	GO:0016192~vesicle-mediated transport	5.85E-07
BP00063:Protein modification	4.95E-21	GO:0006512~ubiquitin cycle	9.13E-07
BP00040:mRNA transcription	4.32E-20	GO:0009966~regulation of signal transduction	3.44E-06
BP00104:G-protein mediated signaling	7.57E-16	GO:0007264~small GTPase mediated signal transduction	7.57E-06
BP00143:Cation transport	8.83E-14	GO:0006366~transcription from RNA polymerase II promoter	9.15E-06
BP00064:Protein phosphorylation	1.49E-12	GO:0046903~secretion	1.05E-05
BP00142:Ion transport	1.96E-11	GO:0008104~protein localization	2.43E-05
BP00286:Cell structure	9.20E-11	GO:0032940~secretion by cell	5.91E-05
BP00289:Other metabolism	3.87E-10	GO:0006468~protein amino acid phosphorylation	5.95E-05
BP00102:Signal transduction	8.08E-10	GO:0046578~regulation of Ras protein signal transduction	5.99E-05
BP00060:Protein metabolism and modification	1.17E-08	GO:0007265~Ras protein signal transduction	7.37E-05

*Twelve most enriched PANTHER and GO biological process terms

Supplementary Table 8. Islet FAIRE enrichment at T2D susceptibility and fasting glycemia loci

Locus	Reference SNP	Position ^a	Locus size (kb)	# FAIRE-enriched regions ^b			# SNPs in dbSNP ^c			# T2D-associated SNPs ^d				
				Liberal	Moderate	Stringent	Liberal	Moderate	Stringent	Liberal	Moderate	Stringent		
<i>TCF7L2</i>	rs7903146 ^e	chr10:114,700,000-114,830,000	130	10	7	2	106	3	2	1	4	1	1	-
<i>CDK4L1</i>	rs4712523 ^e	chr6:20,720,000-20,850,000	130	38	14	5	242	44	8	2	18	5	-	-
<i>CDKN2A/CDKN2B</i>	rs2383208 ^e	chr9:21,970,000-22,135,000	165	25	10	2	324	23	8	-	3	2	1	-
<i>IGF2BP2</i>	rs4402960 ^e	chr3:186,960,000-187,050,000	900	3	3	2	82	4	2	2	29	1	1	1
<i>JAZF1</i>	rs864745 ^e	chr7:28,100,000-28,250,000	150	22	10	6	157	11	4	2	6	-	-	-
<i>CDC123/CAMK1D</i>	rs12779790 ^e	chr10:12,280,000-12,380,000	100	8	3	2	135	9	5	2	4	1	1	-
<i>TSPAN8/LGR5</i>	rs7961581 ^e	chr12:69,636,000-70,000,000	364	34	11	5	623	22	8	5	7	-	-	-
<i>THADA</i>	rs7578597 ^e	chr2:43,290,000-43,750,000	460	100	57	26	638	85	50	24	109	23	12	6
<i>ADAMTS9</i>	rs4607103 ^e	chr3:64,667,000-64,710,400	43.4	1	-	-	76	-	-	-	8	-	-	-
<i>NOTCH2/ADAM30</i>	rs10923931 ^e	chr1:120,200,000-120,350,000	150	5	1	1	223	2	-	-	41	-	-	-
<i>FTO</i>	rs8050136 ^e	chr16:52,350,000-52,410,000	60	15	7	3	114	6	4	1	37	2	-	-
<i>SLC30A8</i>	rs13266634 ^e	chr8:118,100,000-118,350,000	250	80	55	38	377	122	63	21	4	1	1	-
<i>HNF1B</i>	rs7501939 ^f	chr17:33,040,000-33,250,000	210	18	9	2	284	11	7	2	3	1	1	1
<i>WFS1</i>	rs10010131 ^g	chr4:6,310,000-6,375,000	65	3	2	-	115	-	-	-	45	-	-	-
<i>MTNR1B</i>	rs10830963 ^h	chr11:92,300,000-92,370,000	70	3	2	1	96	2	1	1	1	-	-	-
<i>HHEX/IDE</i>	rs1111875 ^e	chr10:94,150,000-94,500,000	350	22	6	4	266	13	5	4	4	-	-	-
<i>KCNQ1</i>	rs2237892 ⁱ	chr11:2,640,000-2,824,000	184	4	3	2	285	2	-	-	3	-	-	-
<i>PPARG/SYN2</i>	rs1801282 ^e	chr3:12,000,000-12,390,000	390	12	1	-	557	1	-	-	13	-	-	-
<i>KCNJ11</i>	rs5215 ^e	chr11:17,329,000-17,385,000	56	7	5	3	88	7	5	3	7	-	-	-
<i>G6PC2</i>	rs560887 ^h	chr2:169,457,500-169,532,500	75	28	21	11	177	52	30	20	3	1	1	1

a. Coordinates defined by identifying recombination hotspots flanking the reference SNP

b. Number of FAIRE-enriched sites in sample 3 at liberal, moderate and stringent threshold located within locus coordinates

c. Number of SNPs in dbSNP v129 with a reported average heterozygosity > 1% located within locus coordinates

d. Number of SNPs in HapMap CEU $r^2 > 0.8$ with reference SNP

e. Mohlke, et al. Hum Mol Genet. 2008 Oct 15;17(R2):R102-8. f. Gudmundsson, et al. Nat Genet. 2007 Aug;39(8):977-83. g. Sandhu, et al. Nat Genet. 2007 Aug;39(8):951-3. h. Prokopenko, et al. Nat Genet. 2007 Jan;41(1):77-81. i. Yasuda, et al. Nat Genet. 2008 Sep;40(9):1092-7.

Supplementary Table 9. Oligonucleotides for genomic region surrounding rs7903146

rs7903146 Sequencing 1	5' GCCTCAAACCTAGCACAGC
rs7903146 Sequencing 2	5' GTGAAGTGCCCAAGCTTCTC
Luciferase construct forward 1	5' <u>GCGCGGTACCAATTCATGGGCTTTCTCTGC</u>
Luciferase construct forward 2	5' <u>GCGCCTCGAGGTGAAGTGCCCAAGCTTCTC</u>
Luciferase construct reverse 1	5' <u>GCGCCTCGAGAATTCATGGGCTTTCTCTGC</u>
Luciferase construct reverse 2	5' <u>GCGCGGTACCGTGAAGTGCCCAAGCTTCTC</u>

Underlined sequences represent introduced restriction enzyme site