nature genetics

ARTICLES

https://doi.org/10.1038/s41588-019-0457-0

Human pancreatic islet three-dimensional chromatin architecture provides insights into the genetics of type 2 diabetes

Irene Miguel-Escalada[1,2,3,4,33], Silvia Bonàs-Guarch[1,2,3,4,33], Inês Cebola[1,33], Joan Ponsa-Cobas[1], Julen Mendieta-Esteban[5], Goutham Atla[1,2,3,4], Biola M. Javierre[6,7], Delphine M. Y. Rolando[1], Irene Farabella[5], Claire C. Morgan[1,2], Javier Garcîa-Hurtado[2,3,4], Anthony Beucher[1], Ignasi Morán[1,8], Lorenzo Pasquali[3,7,9], Mireia Ramos-Rodríguez[9], Emil V. R. Appel[10], Allan Linneberg[11,12], Anette P. Gjesing[10], Daniel R. Witte[13,14], Oluf Pedersen[10], Niels Grarup[10], Philippe Ravassard[15], David Torrents[8,16], Josep M. Mercader[8,17,18], Lorenzo Piemonti[19,20], Thierry Berney[21], Eelco J. P. de Koning[22,23], Julie Kerr-Conte[24], François Pattou[24], Iryna O. Fedko[25,26], Leif Groop[27], Inga Prokopenko[28,29], Torben Hansen[10], Marc A. Marti-Renom[5,16,30,31], Peter Fraser[6,32] and Jorge Ferrer[1,2,3]*

Genetic studies promise to provide insight into the molecular mechanisms underlying type 2 diabetes (T2D). Variants associated with T2D are often located in tissue-specific enhancer clusters or super-enhancers. So far, such domains have been defined through clustering of enhancers in linear genome maps rather than in three-dimensional (3D) space. Furthermore, their target genes are often unknown. We have created promoter capture Hi-C maps in human pancreatic islets. This linked diabetes-associated enhancers to their target genes, often located hundreds of kilobases away. It also revealed >1,300 groups of islet enhancers, super-enhancers and active promoters that form 3D hubs, some of which show coordinated glucose-dependent activity. We demonstrate that genetic variation in hubs impacts insulin secretion heritability, and show that hub annotations can be used for polygenic scores that predict T2D risk driven by islet regulatory variants. Human islet 3D chromatin architecture, therefore, provides a framework for interpretation of T2D genome-wide association study (GWAS) signals.

Type 2 diabetes affects more than 400 million people worldwide[1], and is a classic example of a polygenic disease in which the genetic susceptibility is largely driven by noncoding variants[2,3]. T2D susceptibility variants are enriched in active islet enhancers that cluster in linear genome maps—variously defined as super-enhancers, clusters of open regulatory elements (COREs), enhancer clusters or stretch enhancers[4–7]. Enhancer clusters from other tissues or cell types are similarly enriched in risk variants for

[1]Section of Epigenomics and Disease, Department of Medicine, and National Institute for Health Research Imperial Biomedical Research Centre, Imperial College London, London, UK. [2]Regulatory Genomics and Diabetes, Centre for Genomic Regulation, The Barcelona Institute of Science and Technology, Barcelona, Spain. [3]CIBER de Diabetes y Enfermedades Metabólicas Asociadas, Madrid, Spain. [4]Genomic Programming of Beta Cells Laboratory, Institut d'Investigacions Biomediques August Pi i Sunyer, Barcelona, Spain. [5]CNAG-CRG, Centre for Genomic Regulation, Barcelona Institute of Science and Technology, Barcelona, Spain. [6]Nuclear Dynamics Programme, The Babraham Institute, Cambridge, UK. [7]Josep Carreras Leukaemia Research Institute, Campus ICO-Germans Trias i Pujol, Barcelona, Spain. [8]Barcelona Supercomputing Center, Joint BSC-CRG-IRB Research Program in Computational Biology, Barcelona, Spain. [9]Endocrine Regulatory Genomics Laboratory, Germans Trias i Pujol University Hospital and Research Institute, Barcelona, Spain. [10]Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. [11]Center for Clinical Research and Disease Prevention, Bispebjerg and Frederiksberg Hospital, Copenhagen, Denmark. [12]Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. [13]Department of Public Health, Aarhus University, Aarhus, Denmark. [14]Danish Diabetes Academy, Odense, Denmark. [15]Université Sorbonne, UPMC Univ Paris 06, Inserm, CNRS, Institut du cerveau et de la moelle—Hôpital Pitié-Salpêtrière, Boulevard de l'Hôpital, Paris, France. [16]Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain. [17]Programs in Metabolism and Medical & Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA. [18]Diabetes Unit and Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA. [19]Diabetes Research Institute, IRCCS San Raffaele Scientific Institute, Milan, Italy. [20]Vita-Salute San Raffaele University, Milan, Italy. [21]Cell Isolation and Transplantation Center, University of Geneva, Geneva, Switzerland. [22]Department of Medicine, Leiden University Medical Center, Leiden, the Netherlands. [23]Hubrecht Institute/KNAW, Utrecht, the Netherlands. [24]European Genomic Institute for Diabetes, Lille, France. [25]Department of Biological Psychology, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands. [26]Amsterdam Public Health Research Institute, Amsterdam, the Netherlands. [27]Genomics, Diabetes and Endocrinology, Department of Clinical Sciences, Clinical Research Centre, Lund University, Malmö, Sweden. [28]Section of Genomics of Common Disease, Department of Medicine, Imperial College London, London, UK. [29]Department of Clinical and Experimental Medicine, University of Surrey, Guildford, UK. [30]Universitat Pompeu Fabra, Barcelona, Spain. [31]Gene Regulation, Stem Cells and Cancer, Centre for Genomic Regulation, The Barcelona Institute of Science and Technology, Barcelona, Spain. [32]Department of Biological Science, Florida State University, Tallahassee FL, USA. [33]These authors contributed equally: Irene Miguel-Escalada, Silvia Bonàs-Guarch, Inês Cebola. *e-mail: jorge.ferrer@crg.eu

various common diseases[5,7–11]. So far, however, genome-wide maps of enhancer clusters have been largely defined with unidimensional epigenomic maps, which do not necessarily reflect the capacity of enhancers to cluster in 3D space, as shown for well-characterized loci such as *Hbb* (β-globin) and *Hoxd*[12,13]. In addition, linear maps do not reveal the target genes of enhancers, which are often separated by hundreds of thousands of base pairs. Therefore, there is a need to obtain accurate representations of enhancer domains, and to connect them to the target genes that underpin disease mechanisms.

Here, we used promoter capture Hi-C (pcHi-C)[14] to generate a genome-scale map of interactions between gene promoters and their regulatory elements in human pancreatic islets. This uncovered ~1,300 hubs of islet enhancers that cluster in 3D space. We show that islet enhancer hubs are connected with key islet gene promoters and exhibit properties of regulatory domains. We use genome/epigenome editing to demonstrate the functional connectivity of hubs and we validate functional interactions between enhancers bearing T2D risk variants and their target genes. Finally, we show that islet hubs not only are enriched for T2D association signals, but also can be used to partition polygenic scores to identify T2D genetic susceptibility driven by pancreatic islet regulatory variation.

Results

The promoter interactome of human islets. To create a genome-wide, high-resolution map of long-range interactions between gene promoters and distant regulatory elements in human pancreatic islets, we prepared Hi-C libraries from four human islet samples, and then performed hybridization capture of 31,253 promoter-containing HindIII fragment baits and their ligated DNA fragments. These were then sequenced and processed with the CHiCAGO algorithm to define 175,784 high-confidence interactions (CHiCAGO score > 5) between annotated promoters and distal promoter-interacting DNA fragments[14,15] (Fig. 1a,b and Supplementary Fig. 1). These high-confidence interactions were called with pooled samples, but for 89% of interactions all individual samples showed CHiCAGO scores above the 95% confidence interval of random distance-matched regions (Supplementary Fig. 1d–g). We also validated pcHi-C landscapes by 4C-seq analysis in the EndoC-βH1 human β cell line in two selected loci (Supplementary Fig. 1h,i).

To define the chromatin landscape of interacting regions, we refined existing human islet epigenome annotations by generating human islet ATAC-seq maps and 30 new chromatin immunoprecipitation (ChIP)-seq datasets (Fig. 1b–d and Supplementary Table 1). This enabled a subclassification of active enhancers according to Mediator, cohesin and H3K27ac occupancy patterns (Fig. 1b–d and Supplementary Dataset 1). As expected, promoter-interacting genomic regions were enriched in active enhancers, promoters and CTCF-bound regions (Fig. 1e and Supplementary Fig. 2a–c). pcHi-C interactions observed in pcHi-C maps from distant cell types were enriched in CTCF-binding sites and active promoters, whereas islet-selective interacting regions were enriched in active enhancers (particularly those with strongest Mediator occupancy, which we term class I enhancers) and were connected with genes showing islet-specific expression (Supplementary Fig. 2d–f). This genome-scale map of the human pancreatic islet promoter interactome is accessible for visualization along with pcHi-C maps of other human tissues (www.chicp.org)[16], or as virtual 4C representations of all genes along with islet regulatory annotations (isletregulome.org)[17].

Identification of target genes for islet enhancers. Long-range chromatin interactions are largely constrained within topologically associating domains (TADs), which typically span hundreds of kilobases (kb) and are often invariant across tissues (Supplementary Fig. 3a–e)[18,19]. TADs, however, define broad genomic intervals that do not necessarily inform on the specific interactions that take place in each tissue between individual *cis*-regulatory elements and their target genes. Human islet pcHi-C maps identified high-confidence pcHi-C interactions (CHiCAGO score > 5) between gene promoters and 18,031 different islet enhancers (Fig. 2a). Remarkably, 42.2% of enhancers that showed interactions with gene promoters had high-confidence interactions with more than one gene, thereby illustrating an unexpected complexity of islet enhancer–promoter interactions (Supplementary Fig. 3f).

We used pcHi-C maps to further expand the number of enhancers that could be assigned to target genes. We reasoned that interactions between enhancers and their target genes can be missed due to the stringency of detection thresholds, the strong bias of Hi-C methods against proximal interactions or their dependence on specific environmental conditions. To impute additional enhancer–promoter assignments, we considered promoter-associated three-dimensional spaces (PATs). A PAT space was defined as the space containing all pcHi-C interactions that stem from a promoter bait (Supplementary Fig. 3g,h). We observed that PATs that had one high-confidence enhancer–promoter interaction were more likely to show other enhancer–promoter interactions, and they exhibited chromatin features that distinguished them from other PATs (Supplementary Fig. 3i–k). This prompted us to leverage PAT features to impute plausible target promoter(s) of an additional 18,633 islet enhancers that did not show high-confidence interactions (Fig. 2a; see Supplementary Fig. 3l and Methods for a detailed description of the imputation pipeline). Imputed promoter–enhancer pairs showed higher CHiCAGO scores than distance-matched regions (Kruskall–Wallis $P < 10^{-16}$), suggesting that many imputed assignments represent physical interactions that do not reach our stringent significance thresholds (Supplementary Fig. 3m). In total, we used high-confidence interactions and imputations to assign 36,664 human islet active enhancers (80% of all enhancers) to at least one target gene (Fig. 2a and Supplementary Dataset 2).

We validated these enhancer-to-gene assignments with complementary approaches. First, we calculated normalized H3K27ac signals in assigned enhancer–promoter pairs across human tissues and human islet samples, and found that assigned pairs had distinctly higher correlation values than enhancers paired with distance-matched promoters from the same TAD or an overlapping PAT (Fig. 2b). Importantly, this was true for both high-confidence and imputed assignments (Fig. 2b). As expected, islet-selective expression was enriched in enhancer-assigned genes but not in unassigned genes from the same TAD (Supplementary Fig. 3n). Furthermore, we determined 1,091 expression quantitative trait loci (eQTL) genes (eGenes) from 183 human islet samples (Supplementary Table 2), and found that eQTLs were enriched in enhancer-to-gene assignments determined through either high-confident interactions or imputations, compared with distance-matched regions (odds ratio 3.18 and 4.36; $P = 3.05 \times 10^{-9}$ and 9.01×10^{-23}, respectively) (Fig. 2c).

We further tested enhancer–promoter assignments in a dynamic perturbation model. We exposed human islets from seven donors to moderately low (4 mM) or high (11 mM) glucose for 72 h, which correspond to quasi-physiological glucose concentrations. This led to glucose-dependent H3K27ac changes in 3,850 enhancers at adjusted $P < 0.05$, most of which showed increased activity at high glucose (Supplementary Fig. 3o). This result, therefore, showed that changes in glucose concentrations elicit quantitative chromatin changes in a large number of human islet enhancers. We next reasoned that glucose-regulated enhancers should tend to cause glucose-regulated expression of their target genes. Indeed, we observed that glucose-induced enhancers were preferentially assigned to genes showing glucose-induced messenger RNA, compared with distance-matched active control genes from the same TAD (odds ratio 2.7 and 2.6, Fisher's $P = 4.9 \times 10^{-16}$ and 6.4×10^{-12}, for high-confidence or imputed assignments, respectively) (Fig. 2d). Likewise, genes assigned to glucose-induced enhancers showed greater glucose induction of promoter H3K27ac than of distance-matched promoters in the same TAD (Fig. 2e). Collectively, these

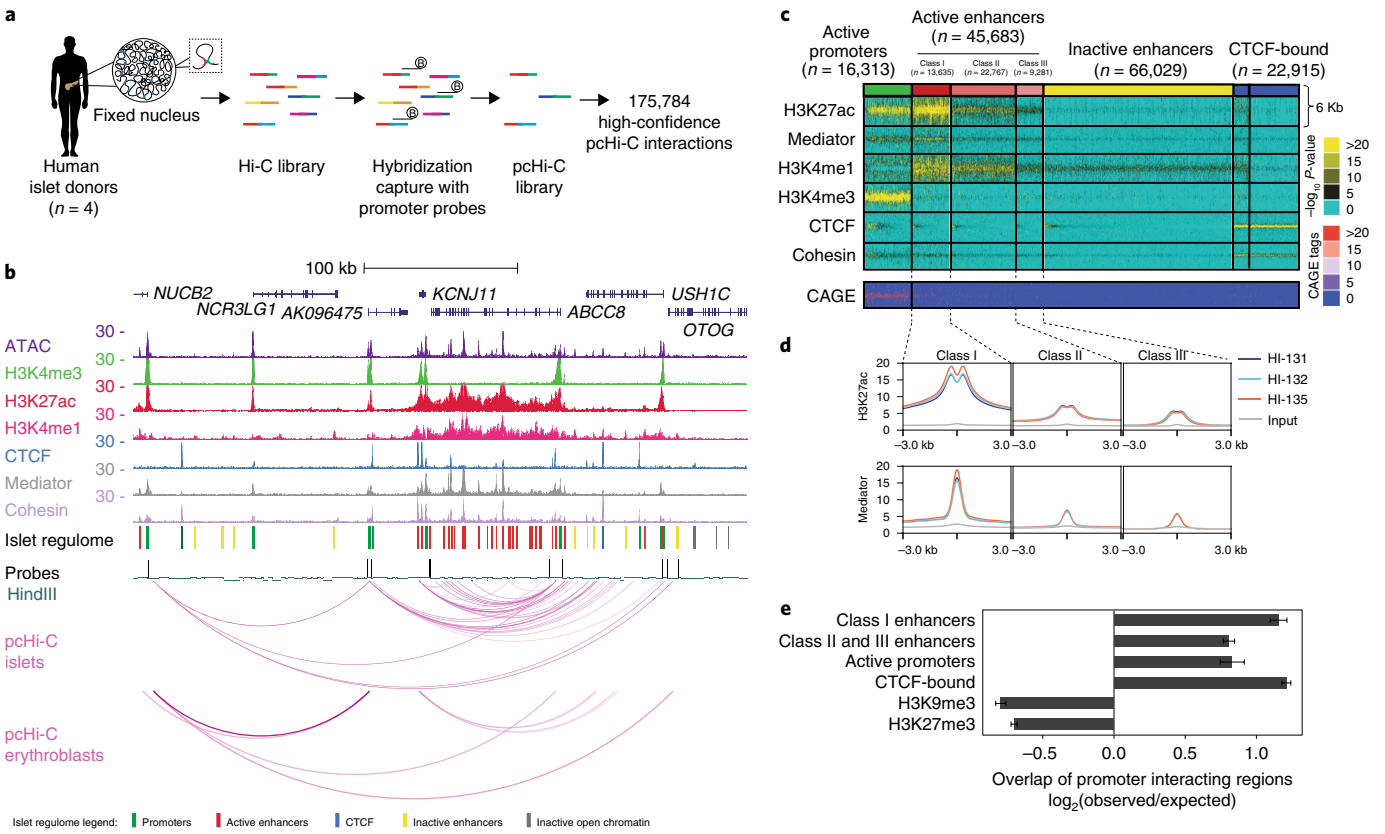


Fig. 1 | The promoter interactome of human pancreatic islets. a, Overview of pcHi-C in human islets. **b**, Integrative map of the *KCNJ11–ABCC8* locus, showing human islet ATAC-seq and ChIP-seq, HindIII bait fragments and arcs representing high-confidence pcHi-C interactions in human islets and erythroblasts. **c**, High-resolution annotations of islet open chromatin. ATAC-seq data from 13 islet samples were used to define consistent open chromatin regions, which were classified with *k*-medians clustering on the basis of epigenomic features. Mediator and H3K27ac binding patterns allowed subclassification of active enhancer classes I–III. Post-hoc analysis of islet CAGE tags confirmed that transcription start sites are highly enriched in promoters and weakly in class I enhancers. These islet regulome annotations are hereafter Supplementary Dataset 1. CAGE, cap-analysis gene expression. **d**, Average H3K27ac and Mediator signal centered on open chromatin regions for active enhancer subtypes in three human islet (HI) samples and input DNA. **e**, Overlap of promoter-interacting regions with epigenomic features, expressed as average $\log_2$ ratios (and 95% confidence intervals) over the overlaps obtained with 100 sets of distance-matched fragments. Error bars show s.d. across control sets.

studies validated pcHi-C maps for the identification of functional target genes of transcriptional enhancers in human pancreatic islets.

Genome editing of T2D-relevant enhancers. A fundamental challenge to translate GWAS data into biological knowledge lies in identifying the target genes of noncoding elements that carry disease-associated regulatory variants. To link noncoding variants to their target genes, we compiled T2D- and fasting glycemia (FG)-associated variants from 109 loci, most of which have been fine-mapped to a credible set (Supplementary Fig. 4a and Supplementary Dataset 3). For fine-mapped loci, variants with a high posterior probability (PP > 0.1) of being causal were most enriched in active islet enhancers ($Z = 20.9$ relative to control regions in the same locus) and promoters ($Z = 7.2$) ($Z < 2$ for other accessible chromatin regions) (Supplementary Fig. 4b). In 61 loci we identified T2D- and/or FG-associated variants overlapping islet enhancers, and assigned one or more candidate target genes for 53 (87%) of these (Fig. 3a and Supplementary Table 3). Some of these target genes were expected, based on their linear proximity to the variants (for example, *ADCY5*, *TCF7L2*, *ZFAND3*, *PROX1*, *FOXA2*), but for 75% of loci we identified more distant candidate genes. Examples of unexpected distal target genes, sometimes in addition to previously nominated proximal genes, include *SOX4* (in the *CDKAL1* locus), *OPTN* (*CDC123/CAMK1D*), *TRPM5* (*MIR4686*), *PDE8B* (*ZBED3*), *SLC36A4* (*MTNR1B*), *POLR3A* and *RPS24* (*ZMIZ1*), *MDGA1* (*ZFAND3*) and *PHF21A* (*CRY2*) (Fig. 3a and Supplementary Table 3; see isletregulome.org or www.chicp.org). Selected unexpected targets, including *ABCB9* and *STARD10*, were additionally supported by concordant eQTLs (Supplementary Fig. 4c,d).

We used genome editing to validate target genes of ten enhancers bearing T2D- or FG-associated variants from eight loci (Fig. 3b and Supplementary Table 4). We performed these experiments in EndoC-βH3 cells, a glucose-responsive human β cell line[20].

In the *CDC123/CAMK1D* locus, only one SNP from a small set of fine-mapped T2D-associated variants is located in an islet enhancer (Fig. 3c, Supplementary Fig. 5a,b and Supplementary Table 3). This variant was previously proposed to be a regulatory variant on the basis of plasmid reporter studies[21], allele-specific chromatin accessibility[22] and as an eQTL for *CAMK1D*[23,24] (Supplementary Table 2). The enhancer showed moderate-confidence interactions (CHiCAGO = 4.42) with *CAMK1D*, but, more surprisingly, showed high-confidence pcHi-C interactions with a more distant gene, *OPTN* (Fig. 3c and Supplementary Fig. 5a). Accordingly, deletion of this enhancer (but not an adjacent region), or silencing with KRAB-dCas9, led to selectively decreased expression of both *OPTN* and *CAMK1D*, whereas targeted activation of the enhancer stimulated their expression (Fig. 3d and Supplementary Fig. 5c,d). These results, therefore, confirm functional relationships predicted by pcHi-C maps. Although the role of *OPTN* and *CAMK1D* as mediators of this T2D-associated genetic signal remains to be defined, the

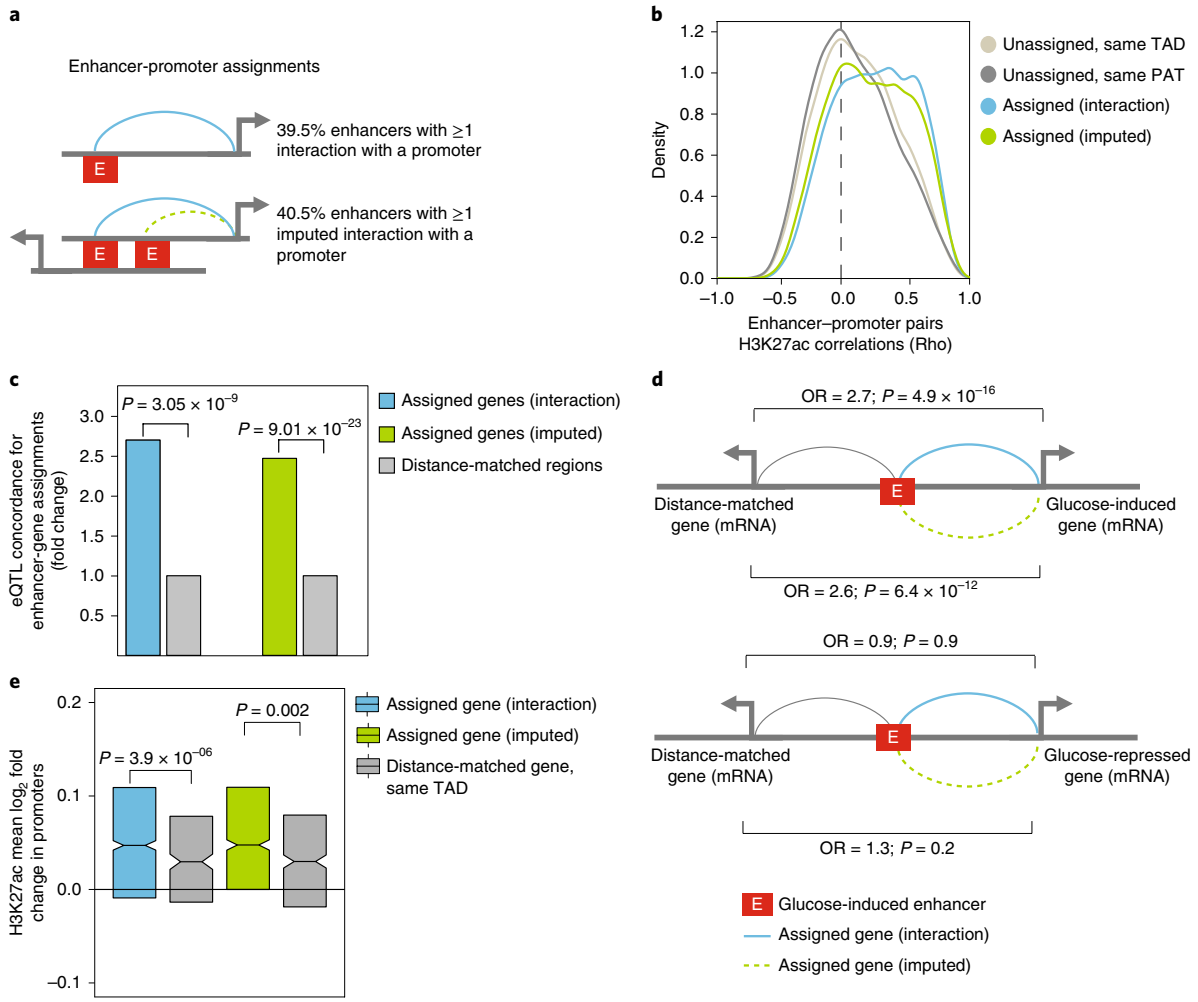


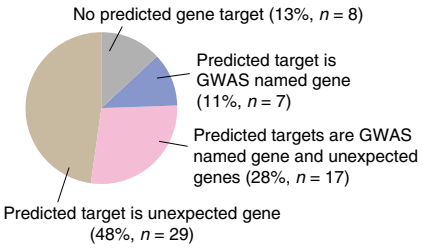
Fig. 2 | Identification of target genes of islet enhancers. a, We assigned target genes to 39.5% of all 45,683 active enhancers through high-confidence interactions. PAT features allowed imputing the assignment of promoters to another 40% of all active enhancers (see Supplementary Fig. 3l,m for further details and evidence that imputed assignments are enriched in subthreshold interactions). **b**, Functional correlation of enhancer–gene pairs assigned through high-confidence interactions ($n = 18,637$ pairs) or imputations ($n = 28,695$ pairs). Spearman's rho values for normalized H3K27ac signal in enhancer–promoter pairs across 14 human islet samples and 51 Roadmap Epigenomics tissues. Control enhancer–gene pairs were enhancers that overlapped a PAT in linear maps but were not assigned to the PAT promoter ($n = 9,770$ pairs), or other unassigned gene-enhancer pairs from the same TAD ($n = 20,186$ pairs). **c**, Concordance of enhancer eQTL–eGene pairs and enhancers-gene pairs assigned through high-confidence interactions ($n = 351$ pairs) or imputations ($n = 293$ pairs) relative to distance-matched control regions ($n = 579$ and 593 pairs, respectively), shown as a fold change. *P* values were derived from one-sided Fisher's exact test. **d**, Genes assigned to glucose-induced enhancers showed concordant glucose-induced expression. Glucose-induced enhancers showed enriched high-confidence ($n = 439$) or imputed ($n = 640$) assignments to glucose-induced genes, compared with distance-matched genes from the same TAD (top). Glucose-induced enhancers showed no enrichment for assignments to genes that were inhibited by high glucose concentrations ($n = 196$ interacting and $n = 218$ imputed pairs) (bottom). OR, odds ratio. *P* values were calculated with chi-square tests. **e**, Genes assigned to glucose-induced enhancers through high-confidence interactions ($n = 275$) or imputations ($n = 321$ pairs) were enriched for glucose-induced promoter H3K27ac, compared with control genes from the same TAD. Box plots represent interquartile ranges (IQRs), notches are 95% confidence intervals of median, *P* values are from Wilcoxon's two-sided signed ranked tests. See also Supplementary Dataset 2.

findings highlight an example of a diabetes-relevant enhancer with multiple target genes.

We also examined rs7903146, a plausible causal SNP in the *TCF7L2* locus. This is the strongest known genetic signal for T2D, and it is known to influence islet cell traits in non-diabetic individuals[2,25,26]. SNP rs7903146 lies in a class I enhancer with unusually high Mediator occupancy (Supplementary Fig. 6a). The SNP alters allele-specific accessibility and episomal enhancer activity[6], and has been associated with differences in *TCF7L2* mRNA[27]. However, deletion of this enhancer in human colon cancer cells affects *ACSL5* rather than *TCF7L2* (ref. [28]), thereby questioning the true target gene(s) of this enhancer in islet cells. We found that the rs7903146-bearing enhancer has imputed and moderate-confidence pcHi-C interactions with *TCF7L2*, but no evidence of proximity with any other gene in human islets (Supplementary Fig. 6a). Consistently, targeted deletion, functional inhibition or stimulation of the enhancer caused selective changes in *TCF7L2* mRNA (Supplementary Fig. 6b,c). Therefore, the enhancer that harbors rs7903146 regulates *TCF7L2* in human β cells. Regardless of the possible metabolic role of this locus in other cell types[29], this finding indicates that *TCF7L2* is a likely mediator of the genetic association between rs7903146 and islet-related traits.

For all eight tested loci, at least one of the genes assigned by pcHi-C to an enhancer showed gene expression changes, and four showed changes in expression of more than one gene (Fig. 3b, Supplementary Table 4 and Supplementary Dataset 4). This included

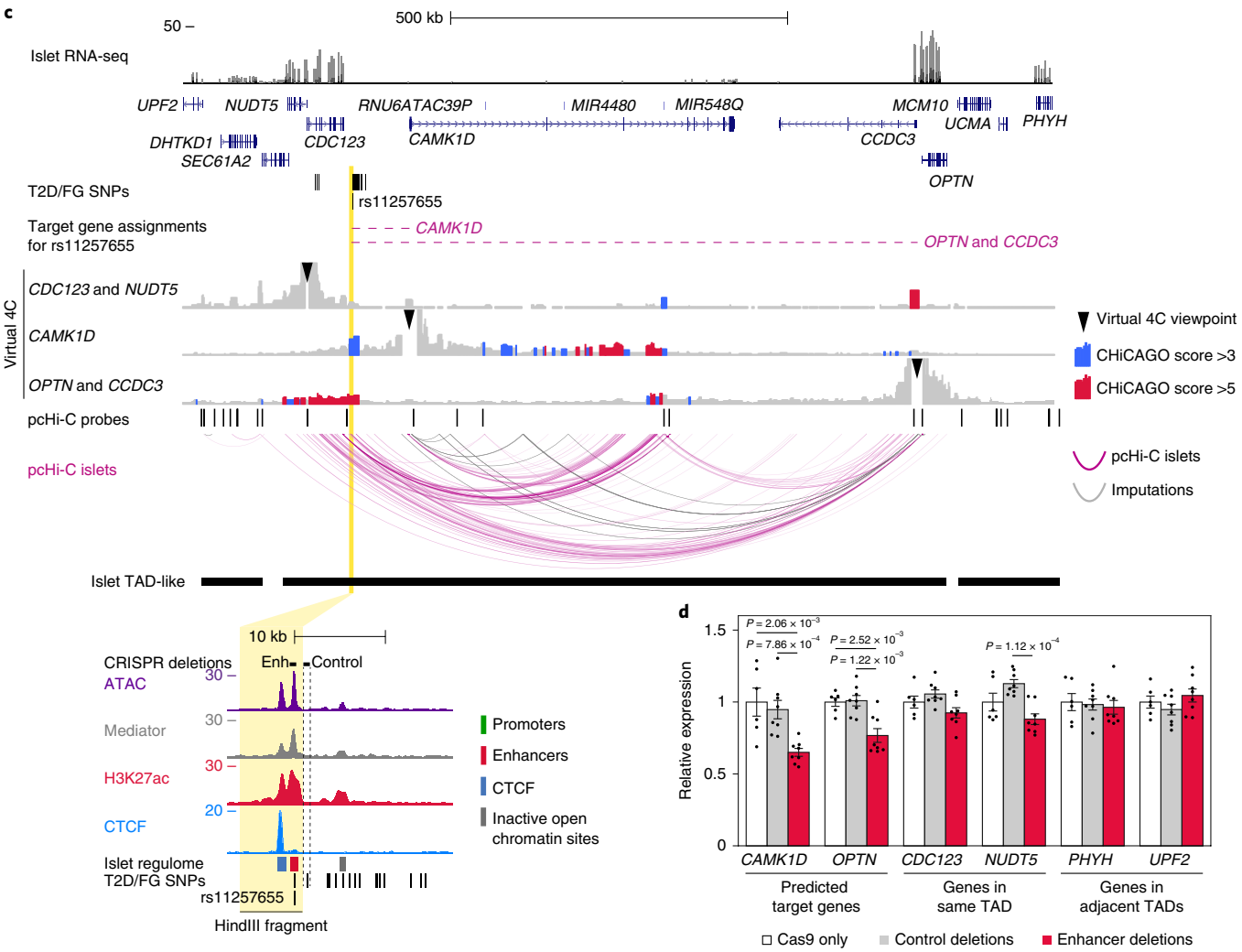
a



b

Enhancer SNP	Assigned genes	Genes affected by CRISPR	CRISPR/Cas9 validations CRISPRi	CRISPRa	Deletion
rs11257655	*CAMK1D, OPTN*	*CAMK1D, OPTN*	✓	✓	✓
rs4237150	*RFX3, RFX3-AS1, GLIS3*	*GLIS3*	✓	✓	✓
rs7903146	*TCF7L2*	*TCF7L2*	✓	✓	✓
rs7163757	*C2CD4A, VPS13C*	*C2CD4A, VPS13C, C2CD4B*	✓	✓	NT
rs1401419	*CRY2, PHF21A*	*CRY2*	✓	✓	NT
rs12189774	*VEGFA*	*VEGFA*	✓	✓	NT
rs7732130	*S100Z, ZBED3, snoRA47, ZBED3-AS1, PDE8B, WDR41*	*ZBED3, snoRA47, ZBED3-AS1, PDE8B, WDR41*	✓	✓	NT
rs58692659	*MDGA, ZFAND3*	*MDGA1, ZFAND3*	✓	✓	NT

c



d

Fig. 3 | Identification of gene targets of T2D-relevant enhancers. a, We assigned gene targets through high-confidence interactions or imputations for 53 (87%) out of 61 T2D/FG-associated loci with genetic variants in islet enhancers (Supplementary Table 3). **b**, Summary of T2D-associated enhancer perturbations presented in this study (see also Supplementary Table 4). NT, not tested. **c**, Islet pcHi-C analysis defines gene targets of an enhancer bearing T2D-associated variants near *CDC123/CAMK1D*. The only T2D risk credible set variant that maps to an islet enhancer in the locus (rs11257655, zoomed inset) is assigned to *CAMK1D* and *OPTN* (dashed horizontal lines). Islet pcHi-C virtual 4C representations from pooled samples show interactions stemming from both *CAMK1D* and *OPTN* promoters towards rs11257655 with CHiCAGO > 3, but not from *CDC123*. **d**, *CAMK1D* and *OPTN* mRNA are regulated by the rs11257655-containing enhancer. We deleted the rs11257655-containing enhancer and a nearby control region with a T2D-associated variant (rs33932777) that lacked active chromatin marks in human islets. Cas9 only: $n = 6$ (two independent experiments with triplicates). Deletions: $n = 8$ (two guide RNA (gRNA) pairs in two independent experiments with biological duplicates). Bars are mean ± s.e.m., normalized by *TBP* and expressed relative to mean levels of the Cas9 only controls. Statistical significance: two-tailed Student's *t*-test.

functionally validated imputed target genes, such as *VEGFA*, as well as *MDGA1* and *ZFAND3* (Supplementary Fig. 7). These functional studies, therefore, underscore the complexity of enhancer–promoter interactions, with long-range interactions that cannot be predicted from linear genome maps, interactions that are not functionally essential and frequent target gene multiplicity. Importantly, the results validate the use of human pcHi-C maps to connect regulatory elements that harbor T2D-associated

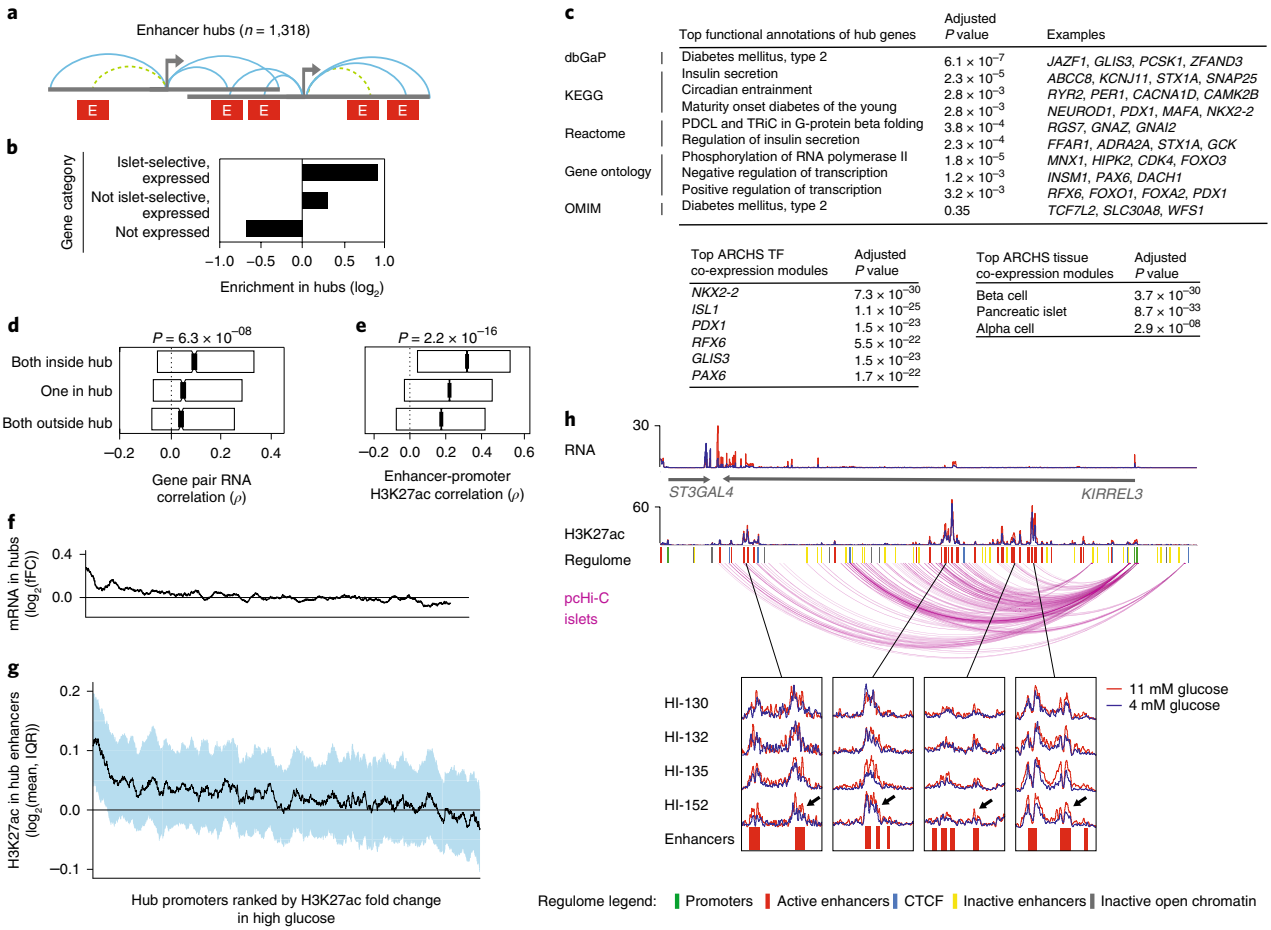


Fig. 4 | Tissue-specific enhancer hubs regulate key islet genes. a, Hubs are composed of one or more enhancer-rich PATs (≥3 class I enhancers) connected through at least one common interacting enhancer. Turquoise and dashed green lines depict high-confidence and imputed assignments, respectively. Descriptive features of hubs are summarized in Supplementary Fig. 8c. **b**, Islet hubs are enriched in genes showing islet-selective expression. Ratios were calculated relative to all annotated genes. **c**, Islet hub genes are enriched in annotations important for islet differentiation, function and diabetes. Benjamini–Hochberg-adjusted *P* values from EnrichR are shown (see complete lists in Supplementary Table 5). **d**, Gene pairs from the same hub show higher RNA correlations across human islet samples and 15 control tissues than gene pairs from the same TAD in which only one gene or neither gene is in a hub. *P* values were derived with Kruskall–Wallis analysis of variance. **e**, Enhancer–promoter pairs from the same hub show high H3K27ac correlations across 14 human islet samples and 51 Epigenome Roadmap tissues, compared with pairs from the same TAD in which only one element or neither are in a hub. *P* values were derived with a Kruskall–Wallis test. **f**,**g**, Culture of seven human islet donor samples at 4 versus 11 mM glucose shows concerted changes in H3K27ac in hub enhancers connected with glucose-dependent genes. Hub promoters were ranked by their median fold change (FC) in H3K27ac at high glucose, so that glucose-induced promoters are on the left of the *x* axis. Median mRNA for genes associated with each hub (**f**). Median glucose-dependent fold change of H3K27ac in enhancers from hubs connecting with each promoter, IQR values in blue shading (**g**). In both graphs values are shown as running averages (window = 50). **h**, Coordinated glucose-induced H3K27ac in enhancers of a hub connected to *KIRREL3*. Top tracks show RNA and H2K27ac in one representative sample. Bottom insets highlight H2K27ac at 11 mM glucose (red) versus 4 mM (blue) in regions showing coordinated glucose-induced changes in most hub enhancers, highlighted with black arrows ($n = 4$ human islet samples). See also Supplementary Table 6 and Supplementary Dataset 5.

variants with the genes that can mediate disease susceptibility mechanisms.

Islet-specific transcription is linked to enhancer hubs. Earlier studies demonstrated that risk variants for common diseases such as T2D are enriched in clusters of enhancers that regulate key cell identity genes[4–7]. However, spatial clustering of enhancers is not necessarily apparent from linear genome maps. To identify 3D enhancer clusters, we again considered promoter-associated 3D spaces, or PATs, and empirically defined enhancer-rich PATs as those containing three or more class I enhancers (enhancers with high H3K27ac and Mediator occupancy, Fig. 1c). This definition of enhancer-rich PATs was supported by a multivariate analysis of genomic and epigenomic PAT features that were most predictive of islet-specific gene expression (Supplementary Fig. 8a and Methods). In total, we identified 2,623 enhancer-rich PATs (Supplementary Fig. 8b). As noted above, many active enhancers (~40%) had interactions with one or more promoters (Supplementary Fig. 3f). Thus, separate enhancer-rich PATs were often connected. We therefore merged enhancer-rich PATs with other PATs connected through enhancer-mediated high-confidence interactions, yielding 1,318 islet enhancer hubs (Fig. 4a and Supplementary Fig. 8c). Compared to alternative enhancer hub definitions, this definition maximized the enrichment of islet cell functional annotations and the number of mapped hubs (Supplementary Fig. 9). The 1,318 islet enhancer hubs are, in essence, 3D chromatin domains that contain a median of 18 enhancers, two active promoters and two shared enhancer interactions (Supplementary Fig. 8d). They are often

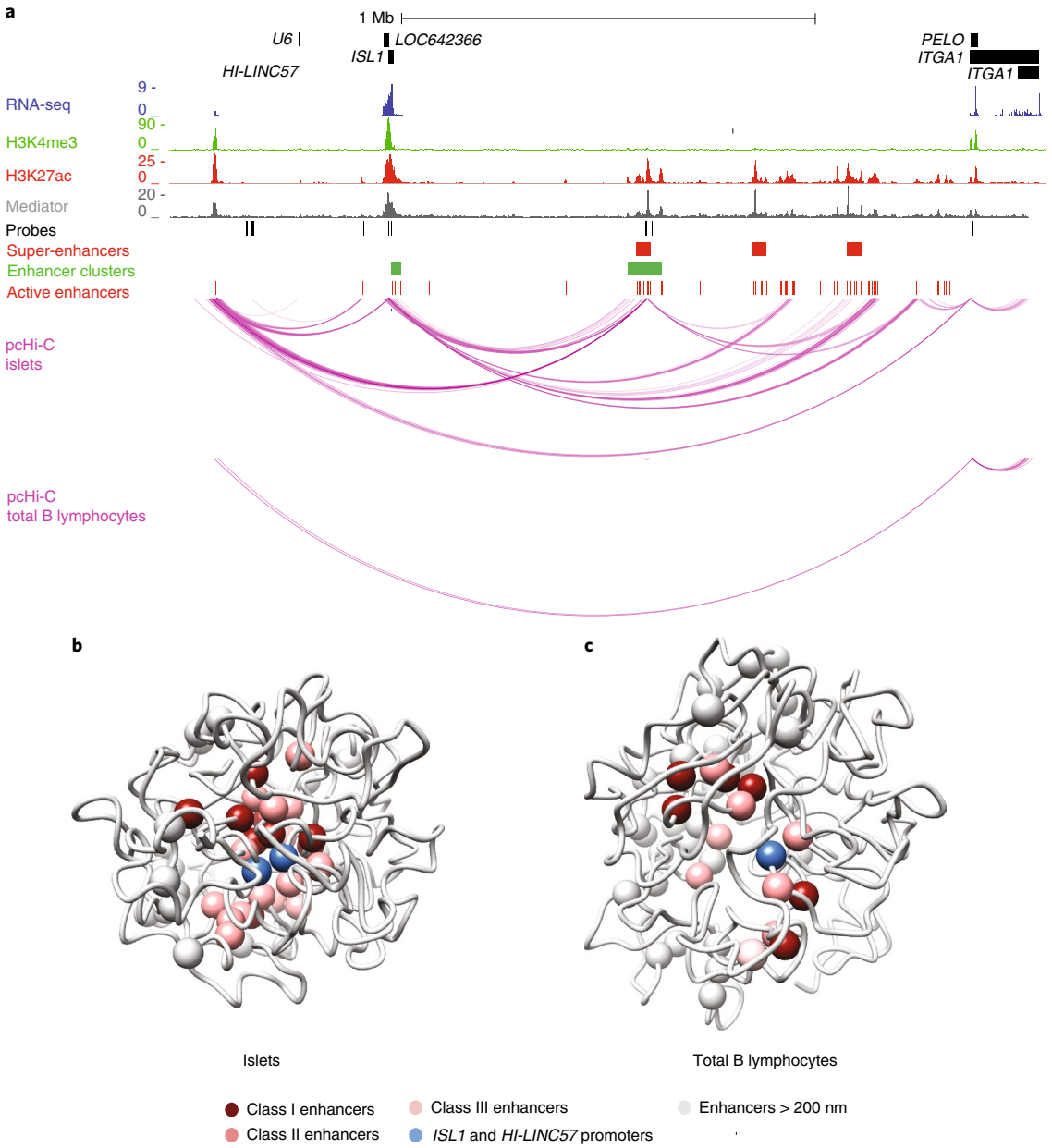


Fig. 5 | Tissue-specific topology of the *ISL1* enhancer hub. a, Epigenomic annotations and high-confidence pcHi-C interactions from pooled islet samples and total B lymphocytes are shown to illustrate active enhancers, super-enhancers and enhancer clusters distributed across a TAD, while sharing islet-selective 3D interactions with *ISL1* and *HI-LNC57*. **b**,**c**, 3D chromatin conformation models of the *ISL1* enhancer hub generated from pcHi-C libraries from human islets (**b**) and total B lymphocytes (**c**). Images represent the top-scoring model from the ensemble of structures that best satisfied spatial restraints. Class I, II and III enhancers are colored in dark to light red and promoters in blue if they are within 200 nm of the *ISL1* promoter, or as white spheres if they are further away than 200 nm. Note the proximity of lncRNA *HI-LNC57* and *ISL1* promoters in islets. The models show that active islet regulatory elements interact in a restricted 3D space in islet nuclei. See also Supplementary Fig. 10b,c and Supplementary Videos 1 and 2.

tissue-selective interaction domains, because hub promoters had a 2.8-fold higher fraction of islet-selective interactions than did non-hub promoters (Wilcoxon's $P = 2.8 \times 10^{-36}$) (Supplementary Fig. 8e; examples in Figs. 1b and 5a and Supplementary Figs. 1h,i and 10a). Furthermore, the genes that form part of enhancer hubs were enriched in islet-selective transcripts, and in functional annotations that are central to islet cell identity, differentiation and diabetes (Fig. 4b,c, Supplementary Table 5 and Supplementary Dataset 5).

Hubs exhibit domain-level chromatin changes. Consistent with the high internal connectivity of hubs, gene pairs from the same hub showed increased RNA expression correlation values across tissues and islet samples, as compared to control active gene pairs in the same TAD as the hubs ($P = 6.3 \times 10^{-8}$) (Fig. 4d). Moreover, hub enhancers showed higher H3K27ac correlations with their target promoters than when they were paired with non-hub promoters from the same TAD ($P = 2.2 \times 10^{-16}$) (Fig. 4e). These findings are consistent with enhancer interaction hubs as functional regulatory domains.

To further explore the behavior of hubs as functional domains, we again examined islets exposed to moderately low versus high glucose concentrations. Glucose-induced enhancers and mRNAs were highly enriched in hubs, compared with their non-hub counterparts (Fisher's $P = 1.1 \times 10^{-7}$ and 2.2×10^{-16}, respectively). Of 297

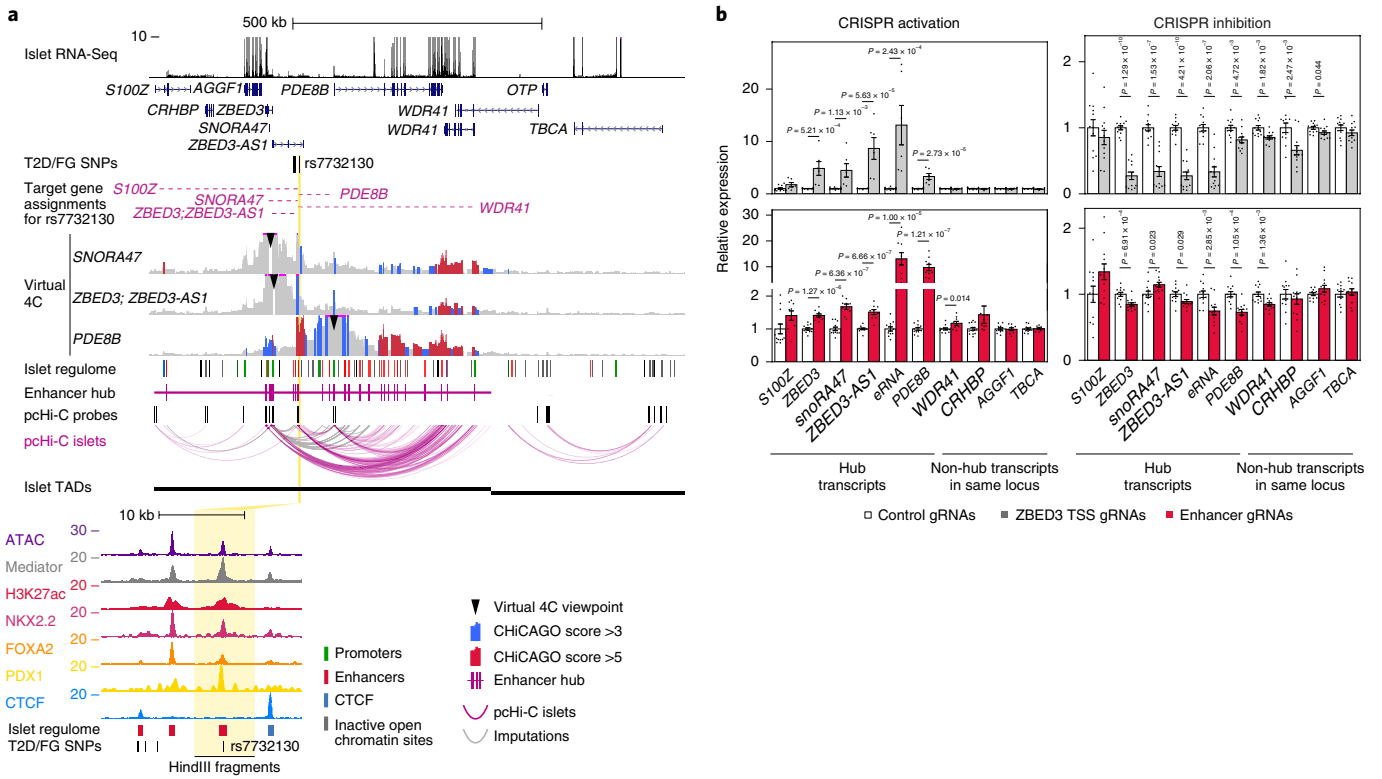


Fig. 6 | The *ZBED3* enhancer hub links an enhancer bearing a T2D SNP with multiple target genes. a, pcHi-C and virtual 4C representations from pooled islet samples for three viewpoints (see also Supplementary Fig. 10). The variant with highest posterior probability in this locus (rs7732130) maps to a class I islet enhancer (yellow line, and zoomed inset) that shows interactions with *PDE8B* (CHiCAGO > 5) and *ZBED3, ZBED3-AS1, SNORA47* and *S100Z* (CHiCAGO > 3, see also Supplementary Fig. 11). *WDR41* is assigned to rs7732130 by imputation. Dashed horizontal lines show all targets assigned through imputation or high-confidence interactions. **b**, Analysis of hub and non-hub transcripts after CRISPR activation or inhibition of the transcriptional start site of *ZBED3* or the rs7732130 enhancer in EndoC-βH3 cells. Data are presented as mean ± s.e.m. of all gRNAs combined per target region (enhancer CRISPRa: three gRNAs, CRISPRi: four gRNAs, all $n = 3$ independent experiments). Statistical significance: two-tailed Student's *t*-test.

promoters that showed glucose-induced H3K27ac, 94 were contained in hubs, and 65% of these showed glucose-induced mRNA (Supplementary Tables 6 and 7). We predicted that if hubs are functional regulatory domains, hub enhancers connected to glucose-induced genes should tend to show coordinated glucose-dependent changes. Our analysis showed that hub enhancers assigned to glucose-induced promoters showed a widespread parallel increase in H3K27ac (Fig. 4f–h and Supplementary Table 8). Thus, varying glucose concentrations elicit chromatin changes in human islets at the level of broad regulatory domains. Taken together, our findings indicate that enhancer hubs have properties of functional units.

Enhancer hubs contain super-enhancers and enhancer clusters. We compared islet enhancer hubs with previously defined islet enhancer domains, such as linear enhancer clusters and super-enhancers (Supplementary Fig. 8f). This showed that hubs have at least some spatial overlap with 70% of enhancer clusters[7], and with 87% of super-enhancers defined with a standard algorithm[4] (Supplementary Fig. 8g–i). Hubs, however, differ in that they can be connected with their target genes. Furthermore, enhancer hubs capture spatial clusters of Mediator-bound (class I) enhancers that do not cluster in the linear genome and therefore do not fulfill definitions of super-enhancers and enhancer clusters (Supplementary Fig. 8j–l)[4,7]. In fact, many hubs contained several interconnected enhancer clusters or super-enhancers (Supplementary Fig. 8m–o). This is illustrated by the *ISL1* locus, which has several enhancer clusters and super-enhancers distributed across an entire TAD, whereas pcHi-C points to a single hub that connects dozens of enhancers with *ISL1* and long noncoding RNA *HI-LNC57* (Fig. 5a). Thus, enhancer hubs are 3D domains that often include one or more enhancer clusters or super-enhancers and their target gene(s).

Tissue-specific architecture of the *ISL1* enhancer hub. To gain insight into the 3D conformation of enhancer hubs, we built 3D models of hubs using islet pcHi-C interaction data (Fig. 5a). We focused on the *ISL1* locus because it contains a single hub within a TAD-like domain, with few other annotated genes. We used islet pcHi-C data to build interaction matrices at 5-kb resolution, and transformed the frequency of interactions between genomic segments into spatial restraints[30,31]. We then used molecular dynamic optimization to generate an ensemble of 500 models that best satisfied the imposed restraints. This showed colocalization of islet enhancers and target genes in a constrained space of the TAD, whereas models built from B lymphocyte pcHi-C libraries showed decreased aggregation of these regions (Fig. 5b,c, Supplementary Fig. 10b,c and Supplementary Videos 1 and 2). Quantitative analysis of *ISL1* and six other T2D-relevant hubs showed analogous tissue-specific aggregation of hub enhancers and promoters (Supplementary Figs. 10d–I and 13f–h). These models, which capture the average topology in a population of cells, serve to highlight that whereas TADs are defined as single intervals in linear genome maps, hubs are formed by multiple interspersed regions that occupy a shared 3D subspace within a TAD.

Epigenome editing of T2D-associated islet hubs. We used enhancer perturbations to test the functional connectivity of selected enhancer hubs. In the *ZBED3* locus, we targeted a class I enhancer that contains a variant with highest posterior probability

for causality in T2D fine-mapping studies (PP = 0.461) (Fig. 6a, Supplementary Fig. 11a and Supplementary Table 4). Targeted epigenomic activation or inhibition of this single enhancer led to significant changes in the expression of five of the six genes connected with this hub, but not of non-hub genes from the same TAD (Fig. 6b). In three other hubs we perturbed single enhancers containing candidate T2D susceptibility causal variants, which led to expression changes in *CRY2* and *PHF21A* (Supplementary Fig. 11b,c), *VPS13C, C2CD4A* and *C2CD4B* (Supplementary Fig. 12) and *GLIS3* (Supplementary Fig. 13). These findings highlight a remarkable functional connectivity of enhancer hubs.

Islet hub variants impact insulin secretion. Previous evidence that T2D susceptibility variants are enriched in islet enhancer clusters[5–7,24,32] prompted us to examine the enrichment of diabetes-associated variants in our newly defined annotations. T2D/FG-associated SNPs were enriched in islet pcHi-C interaction regions (Fig. 7a) and in islet enhancer hub class I enhancers, rather than in other active enhancers (Fig. 7b, Supplementary Figs. 9 and 14a–f and Supplementary Table 9). This indicates that hub class I enhancer variants are important for T2D susceptibility.

A major portion of the heritability of common diseases is driven by many variants that, individually, have not achieved genome-wide significance, yet exert a large aggregate effect[33–35]. Consistent with this notion, common variants that have so far not shown genome-wide significance for T2D association, but are located in pcHi-C interacting regions or hub class I enhancers, showed more significant association *P* values than expected distributions (Fig. 7c,d). This observation prompted us to quantify the overall contribution of common variants in islet hubs to the heritability of T2D. We used stratified linkage disequilibrium (LD) score regression[36], and found that hub class I enhancers showed the most significantly increased per-SNP T2D heritability coefficient ($q = 1.64 \times 10^{-2}$) compared with various islet and non-islet genomic annotations (Fig. 7e, Supplementary Fig. 15a and Supplementary Table 10).

Although islet dysfunction is central to the pathophysiology of T2D, other tissues (liver, adipose, muscle, brain, among others) are also critically important[37]. Genetic variation in islet hub enhancers should, therefore, predominantly impact on the heritability of pancreatic islet function. Indeed, islet hub variants showed higher heritability enrichment estimates for islet cell traits than for T2D (Fig. 7e, Supplementary Fig. 15a–f and Supplementary Table 10). Consequently, common variation in hub class I enhancers (0.26% of genomic SNPs) explained 9.9% of observed genetic heritability for T2D, 21.9% for acute insulin secretory response in intravenous glucose tolerance tests[26], 17.2% for HOMA-B models of β-cell function and 31.2% for an insulinogenic index on the basis of oral glucose tolerance tests (OGTT) [38] (Supplementary Table 10). In sharp contrast, islet hub variants showed no enrichment for HOMA-IR, an estimate of insulin resistance (Supplementary Fig. 15e). Of note is that significant heritability enrichments were generally also observed for enhancer clusters, stretch enhancers or super-enhancer annotations, yet estimates were consistently larger for hub enhancers (Fig. 7e and Supplementary Fig. 15a–d). These results indicate that enhancer hubs define genomic spaces that play a prominent role in the heritability of T2D and insulin secretion.

Hub variants provide tissue-specific risk scores. Recent studies suggest that polygenic risk scores (PRS) that integrate effects of a very large number of variants, including many that lack genome-wide significant association, can identify individuals with extreme levels of risk for polygenic diseases, including T2D[33,35,39–41]. We assessed whether islet hub variants could be harnessed to more specifically identify individuals in whom variation in islet function plays a preponderant role in T2D susceptibility.

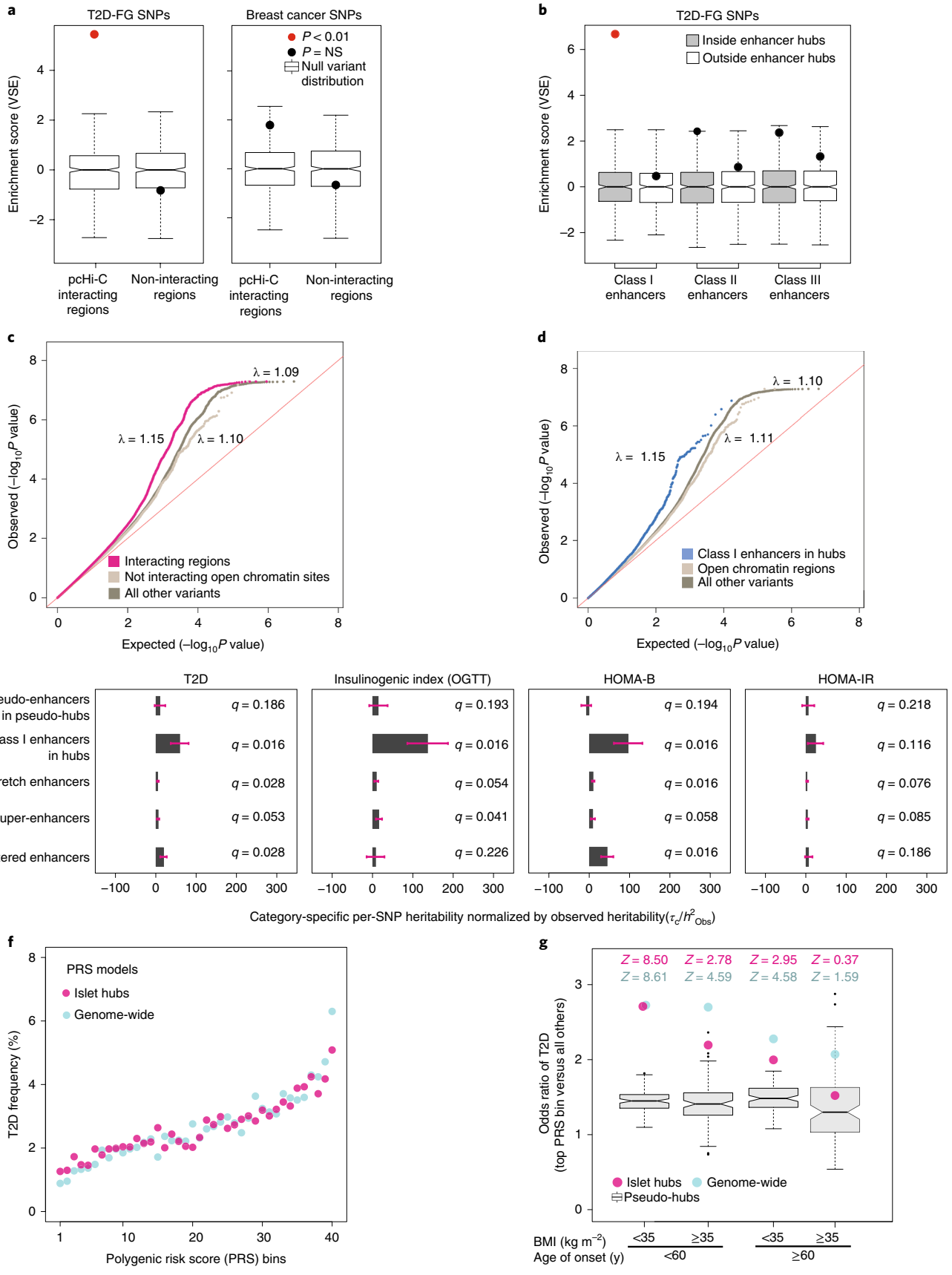
We first created a PRS model using all common variants from a recent body mass index (BMI)-adjusted T2D GWAS meta-analysis[42], and examined the ability of this genome-wide PRS to predict T2D in the UK Biobank population cohort[43,44]. This showed that 2.5% of the UK Biobank individuals with the highest PRS had a 7.11-fold higher frequency of T2D than those with the lowest 2.5% (Fig. 7f).

Next, we created PRS models that contained DNA variants from either: (1) islet hub enhancers and promoters (1.6% of the genome), (2) all other islet open chromatin regions (5.0% of the genome) or (3) the rest of the genome. Despite the fact that islet hub regions encompass <2% of the genome, the T2D risk ratio—defined as the T2D frequency in the top versus bottom risk bins—was 4.02-fold, which was comparable to that observed with variants from the rest of the genome (risk ratio 3.96), and larger than that of other open chromatin regions (risk ratio 3.01) (Fig. 7f and Supplementary Fig. 15g,h). Thus, islet hub variants possess a capacity to predict T2D risk that plausibly reflects their observed impact on the heritability of islet function (Fig. 7e).

Although, as expected, the genome-wide PRS model shows higher risk ratios than islet hub PRS models (Fig. 7e), the latter could potentially define qualitatively distinct T2D risk profiles. Monogenic defects in islet transcription factors typically cause early-onset diabetes in lean individuals, suggesting that islet *cis*-regulatory variants could also predominantly impact T2D risk at an earlier age and lower BMI. We thus compared the effect of hub PRS across BMI and age of onset of diabetes, and considered how it deviated from PRS calculated from genomic regions of similar size and distribution as hubs (100 iterations of 1,000 pseudo-enhancer hubs redistributed across TADs). For hub PRS, this T2D risk ratio showed greatest deviations from pseudo-hub PRS in individuals with BMI < 30 (hub risk ratio = 6.25, $Z = 5.68$) and T2D diagnosed before 50 years (hub risk ratio = 6.67, $Z = 5.27$), but then sharply

Fig. 7 | Islet hub variants impact insulin secretion and provide tissue-specific risk scores. a, Variant set enrichment (VSE) for T2D and FG ($n = 2,771$ variants; Supplementary Table 9) and breast cancer ($n = 3,048$ variants) in high-confidence interacting fragments in islets. Box plots show 500 permutations of matched random haplotype blocks. Red dots indicate significant enrichments (Bonferroni-adjusted $P < 0.01$). NS, not significant. **b**, T2D and FG GWAS significant variants are selectively enriched in hub class I islet enhancers. Box plots show median and IQR. **c**, Genomic inflation of T2D association *P* values for non-GWAS significant variants ($P > 5 \times 10^{-8}$) from a T2D GWAS meta-analysis (12,931 cases, 57,196 controls) in islet high-confidence interacting regions (magenta), non-interacting islet open chromatin (beige) and all other variants (brown). **d**, Genomic inflation of T2D association *P* values for non-GWAS significant variants in hub class I islet enhancers (blue), non-hub islet open chromatin (beige) and all other variants (brown). **e**, Heritability estimates based on GWAS summary statistics for T2D (12,931 cases, 57,196 controls), insulinogenic index (OGTT, 7,807 individuals), homeostasis model assessment of β-cell function (HOMA-B) and insulin resistance (HOMA-IR) (~80,000 individuals), for indicated islet enhancer domains. Bars show category-specific τ_c divided by LD Score heritability (h^2) of each trait. τ_c coefficients were obtained independently for each trait, controlling for 53 functional annotation categories. Values were multiplied by 10^7 and are shown with s.e.m. **f**, T2D frequency across 40 bins, each representing 2.5% of individuals in the UK Biobank test dataset (226,777 controls; 6,127 T2D cases) with increasing PRS, calculated with hub (pink dots) or genome-wide variants (light green). **g**, Odds ratios (OR) for T2D calculated for 2.5% individuals with highest PRS versus all other individuals, using islet hub (pink) or genome-wide models (green), stratified by BMI and T2D age of onset. Box plots show ORs for PRS from 100 permutations of pseudo-hubs (IQRs). *Z*-scores are standard deviations of pseudo-hub averages. See also Supplementary Fig. 15 and Supplementary Table 17.

declined with increasing BMI and age of onset of T2D (BMI ≥ 35, hub risk ratio = 2.67, $Z = 2.98$; T2D onset ≥60 years, hub risk ratio = 3.01, $Z = 2.94$) (Supplementary Fig. 15h). This contrasted from PRS models built with the rest of the genome, which showed greatest deviations from pseudo-hubs in individuals with BMI > 35 and T2D diagnosed after 65 years, or PRS built with other islet open



chromatin regions, which showed modest deviations of risk ratios in all groups (Supplementary Fig. 15h). We further stratified UK Biobank individuals by both BMI and age of onset of diabetes, and found that individuals with 2.5% top hub risk scores had an odds ratio of 2.71 for T2D diagnosed at <60 years of age and BMI < 35 (Fig. 7g). This odds ratio was a major deviation from that observed with pseudo-hub PRS ($Z = 8.50$), and was equivalent to the T2D risk of the highest genome-wide PRS (Fig. 7g; see Supplementary Fig. 15i for other control regions). At the other extreme of the phenotypic spectrum (BMI ≥ 35 and age of onset ≥ 60), individuals with the highest islet hub PRS showed a lower odds ratio, which did not differ from pseudo-hub genomic regions (odds ratio = 1.52, $Z = 0.37$) (Fig. 7g). Taken together, these results indicate that islet enhancer hub variants, which impact islet gene regulation and insulin secretion, can provide distinct T2D risk scores.

Discussion

We have created human islet 3D genome maps that link human pancreatic islet enhancers to gene promoters. We validated them with experimental perturbation models and eQTLs, and showed how they can identify the target genes of diabetes-relevant regulatory elements. This resource can therefore assist efforts to understand the molecular mechanisms that influence T2D susceptibility.

Our study has systematically mapped >1,300 enhancer hubs in human islets. These enhancer domains align with earlier observations derived from lower resolution Hi-C maps, which showed broad genomic regions that exhibit unusually high interaction frequencies[45], with numerous well-characterized chromatin hubs[12,13], and with evolutionarily conserved noncoding sequence blocks[46,47]. We show that enhancer hubs exhibit features of regulatory domains that control genes important for islet cell function, differentiation and diabetes. They also contain DNA variants that have a major impact on the heritability of insulin secretion. Hub elements, therefore, define a genomic space that has direct relevance to islet function and human diabetes. Islet enhancer hubs should thus provide a useful gene-centric framework for genetic studies that aim to discover regulatory variants underlying T2D and monogenic diabetes.

Our work is relevant to the dissection of the polygenic underpinnings of T2D. Recently, genome-wide polygenic risk scores have shown promise for the prediction of common diseases[35]. Because T2D pathophysiology is heterogeneous and multiorganic[37,48,49], it is reasonable to presume that partitioned polygenic risk scores could also provide risk estimates that distinguish mechanisms of susceptibility across individuals. Polygenic scores based on islet hub variants could thus be leveraged to quantify patient-specific genetic risk acting through islet gene regulation and insulin secretion.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of code and data availability and associated accession codes are available at https://doi.org/10.1038/s41588-019-0457-0.

Received: 2 July 2018; Accepted: 29 May 2019;
Published online: 28 June 2019

References

1. Chatterjee, S., Khunti, K. & Davies, M. J. Type 2 diabetes. *Lancet* **389**, 2239–2251 (2017).
2. Flannick, J. & Florez, J. C. Type 2 diabetes: genetic data sharing to advance complex disease research. *Nat. Rev. Genet.* **17**, 535–549 (2016).
3. Fuchsberger, C. et al. The genetic architecture of type 2 diabetes. *Nature* **536**, 41–47 (2016).
4. Whyte, W. A. et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319 (2013).
5. Parker, S. C. et al. Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. *Proc. Natl Acad. Sci. USA* **110**, 17921–17926 (2013).
6. Gaulton, K. J. et al. A map of open chromatin in human pancreatic islets. *Nat. Genet.* **42**, 255–259 (2010).
7. Pasquali, L. et al. Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. *Nat. Genet.* **46**, 136–143 (2014).
8. Cohen, A. J. et al. Hotspots of aberrant enhancer activity punctuate the colorectal cancer epigenome. *Nat. Commun.* **8**, 14400 (2017).
9. Farh, K. K. et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* **518**, 337–343 (2015).
10. Hnisz, D. et al. Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934–947 (2013).
11. Vahedi, G. et al. Super-enhancers delineate disease-associated regulatory nodes in T cells. *Nature* **520**, 558–562 (2015).
12. Montavon, T. et al. A regulatory archipelago controls Hox genes transcription in digits. *Cell* **147**, 1132–1145 (2011).
13. Patrinos, G. P. et al. Multiple interactions between regulatory regions are required to stabilize an active chromatin hub. *Genes Dev.* **18**, 1495–1509 (2004).
14. Javierre, B. M. et al. Lineage-specific genome architecture links enhancers and non-coding disease variants to target gene promoters. *Cell* **167**, 1369–1384.e19 (2016).
15. Cairns, J. et al. CHiCAGO: robust detection of DNA looping interactions in Capture Hi-C data. *Genome Biol.* **17**, 127 (2016).
16. Schofield, E. C. et al. CHiCP: a web-based tool for the integrative and interactive visualization of promoter capture Hi-C datasets. *Bioinformatics* **32**, 2511–2513 (2016).
17. Mularoni, L., Ramos-Rodriguez, M. & Pasquali, L. The pancreatic islet regulome browser. *Front Genet.* **8**, 13 (2017).
18. Dixon, J. R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
19. Nora, E. P. et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381–385 (2012).
20. Benazra, M. et al. A human beta cell line with drug inducible excision of immortalizing transgenes. *Mol. Metab.* **4**, 916–925 (2015).
21. Fogarty, M. P., Cannon, M. E., Vadlamudi, S., Gaulton, K. J. & Mohlke, K. L. Identification of a regulatory variant that binds FOXA1 and FOXA2 at the CDC123/CAMK1D type 2 diabetes GWAS locus. *PLoS Genet.* **10**, e1004633 (2014).
22. Thurner, M. et al. Integration of human pancreatic islet genomic data refines regulatory mechanisms at Type 2 diabetes susceptibility loci. *eLife* **7**, e31977 (2018).
23. van de Bunt, M. et al. Transcript expression data from human islets links regulatory signals from genome-wide association studies for Type 2 diabetes and glycemic traits to their downstream effectors. *PLoS Genet.* **11**, e1005694 (2015).
24. Varshney, A. et al. Genetic regulatory signatures underlying islet gene expression and type 2 diabetes. *Proc. Natl Acad. Sci. USA* **114**, 2301–2306 (2017).
25. Scott, R. A. et al. An expanded genome-wide association study of type 2 diabetes in Europeans. *Diabetes* **66**, 2888–2902 (2017).
26. Wood, A. R. et al. A genome-wide association study of IVGTT-based measures of first-phase insulin secretion refines the underlying physiology of type 2 diabetes variants. *Diabetes* **66**, 2296–2309 (2017).
27. Lyssenko, V. et al. Mechanisms by which common variants in the TCF7L2 gene increase risk of type 2 diabetes. *J. Clin. Invest.* **117**, 2155–2163 (2007).
28. Xia, Q. et al. The type 2 diabetes presumed causal variant within TCF7L2 resides in an element that controls the expression of ACSL5. *Diabetologia* **59**, 2360–2368 (2016).
29. Nobrega, M. A. TCF7L2 and glucose metabolism: time to look beyond the pancreas. *Diabetes* **62**, 706–708 (2013).
30. Bau, D. et al. The three-dimensional folding of the alpha-globin gene domain reveals formation of chromatin globules. *Nat. Struct. Mol. Biol.* **18**, 107–114 (2011).
31. Serra, F. et al. Automatic analysis and 3D-modelling of Hi-C data using TADbit reveals structural features of the fly chromatin colors. *PLoS Comput. Biol.* **13**, e1005665 (2017).
32. Gaulton, K. J. et al. Genetic fine mapping and genomic annotation defines causal mechanisms at type 2 diabetes susceptibility loci. *Nat. Genet.* **47**, 1415–1425 (2015).
33. Boyle, E. A., Li, Y. I. & Pritchard, J. K. An expanded view of complex traits: from polygenic to omnigenic. *Cell* **169**, 1177–1186 (2017).
34. Wood, A. R. et al. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.* **46**, 1173–1186 (2014).
35. Khera, A. V. et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat. Genet.* **50**, 1219–1224 (2018).
36. Finucane, H. K. et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).

37. DeFronzo, R. A. et al. Type 2 diabetes mellitus. *Nat. Rev. Dis. Primers* **1**, 15019 (2015).
38. Gjesing, A. P. et al. Genetic and phenotypic correlations between surrogate measures of insulin release obtained from OGTT data. *Diabetologia* **58**, 1006–1012 (2015).
39. Mahajan, A. et al. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat. Genet.* **50**, 1505–1513 (2018).
40. Khera, A. V. et al. Polygenic prediction of weight and obesity trajectories from birth to adulthood. *Cell* **177**, 587–596.e9 (2019).
41. Richardson, T. G., Harrison, S., Hemani, G. & Davey Smith, G. An atlas of polygenic risk score associations to highlight putative causal relationships across the human phenome. *eLife* **8**, e43657 (2019).
42. Bonas-Guarch, S. et al. Re-analysis of public genetic data reveals a rare X-chromosomal variant associated with type 2 diabetes. *Nat. Commun.* **9**, 321 (2018).
43. Sudlow, C. et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779 (2015).
44. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
45. Schmitt, A. D. et al. A compendium of chromatin contact maps reveals spatially active regions in the human genome. *Cell Rep.* **17**, 2042–2059 (2016).
46. Harmston, N. et al. Topologically associating domains are ancient features that coincide with Metazoan clusters of extreme noncoding conservation. *Nat. Commun.* **8**, 441 (2017).
47. Akalin, A. et al. Transcriptional features of genomic regulatory blocks. *Genome Biol.* **10**, R38 (2009).
48. Ahlqvist, E. et al. Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables. *Lancet Diabetes Endocrinol.* **6**, 361–369 (2018).
49. Kahn, S. E., Cooper, M. E. & Del Prato, S. Pathophysiology and treatment of type 2 diabetes: perspectives on the past, present, and future. *Lancet* **383**, 1068–1083 (2014).

Acknowledgements

This research was supported by the National Institute for Health Research Imperial Biomedical Research Centre. Work was funded by grants from the Wellcome Trust (nos. WT101033 to J.F. and WT205915 to I.P.), Horizon 2020 (Research and Innovation Programme nos. 667191, to J.F., 633595, to I.P., and 676556, to M.A.M.-R.; Marie Sklodowska-Curie 658145, to I.M.-E., and 43062 ZENCODE, to G.A.), European Research Council (nos. 789055, to J.F., and 609989, to M.A.M.-R.). Marató TV3 (no. 201611, to J.F. and M.A.M.-R.), Ministerio de Ciencia Innovación y Universidades (nos. BFU2014-54284-R, RTI2018-095666, to J.F., BFU2017-85926-P, to M.A.M.-R., IJCI-2015-23352, to I.F.), AGAUR (to M.A.M.-R.). UK Medical Research Council (no. MR/L007150/1, to P.F., MR/L02036X/1 to J.F.), World Cancer Research Fund (WCRF UK, to I.P.) and World Cancer Research Fund International (no. 2017/1641 to I.P.), Biobanking and Biomolecular Resources Research Infrastructure (nos. BBMRI-NL, NWO 184.021.007, to I.O.F.). Work in IDIBAPS, CRG and CNAG was supported by the CERCA Programme, Generalitat de Catalunya and Centros de Excelencia Severo Ochoa (no. SEV-2012-0208). Human islets were provided through the European islet distribution program for basic research supported by JDRF (no. 3-RSC-2016-160-I-X). We thank N. Ruiz-Gomez for technical assistance; R. L. Fernandes, T. Thorne (University of Reading) and A. Perdones-Montero (Imperial College London) for helpful discussions regarding Machine Learning approaches; B. Lenhard and M. Merkenschlager (London Institute of Medical Sciences, Imperial College London), F. Müller (University of Birmingham) and J. L. Gómez-Skarmeta (Centro Andaluz de Biología del Desarrollo) for critical comments on the draft; the CRG Genomics Unit; and the Imperial College High Performance Computing Service.

Author contributions

I.M.-E., I.C. and B.M.J. performed and analyzed experiments. I.M.-E. and J.G.-H. processed human islet samples. I.M.-E., S.B.-G., I.C., J.P.-C., D.M.Y.R., G.A., C.C.M. and I.M. performed computational analysis. J.M.-E. and I.F. modeled and analyzed 3D data. L. Piemonti, T.B., E.J.P.d.K., J.K.-C., F.P. and P.R. provided material and reagents. E.V.R.A., A.L., A.P.G., D.R.W., O.P., N.G., J.M.M., D.T., I.O.F., I.P., T.H., and L.G. provided genetics data. M.R.-R. and L. Pasquali created software resources. I.C. and A.B. developed genome-editing methods. M.A.M.-R., P.F. and J.F. supervised analysis. I.M.-E., I.C., S.B.-G., J.P.-C., D.M.Y.R. and J.F. conceived the project. I.M.-E., S.B.-G., I.C. and J.F. wrote and edited the manuscript, which all authors have approved.

Competing interests

P.R. is a shareholder and consultant for Endocells/Unicercell Biosolutions.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41588-019-0457-0.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to J.F.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Methods

Human islets. Human pancreatic islets from organ donors without a history of glucose intolerance were purified using established isolation procedures[50–53], shipped in culture medium and re-cultured at 37 °C in a humidified chamber with 5% CO_2 in glucose-free RPMI 1640 supplemented with 10% fetal calf serum, 100 U ml^{-1} penicillin, 100 U ml^{-1} streptomycin and 11 mM glucose for 3 d before analysis. RNA was extracted from flash-frozen islet pellets using TRIzol Reagent (ThermoFisher Scientific). For glucose regulation studies, islets were cultured in identical time and medium, except that glucose-free RPMI 1640 medium was supplemented with glucose to achieve final concentrations of either 4 or 11 mM glucose. Donor and sample characteristics are provided in Supplementary Table 11.

Compliance with ethical regulations for human research studies is described in Supplementary Note 1.

pcHi-C. From four islet donors, 30–60 million human islet cells per donor were cultured as described above for 3 d before fixation in 2% paraformaldehyde (Agar Scientific) at room temperature for 10 min with mixing, quenched in 125 mM glycine for 5 min at room temperature and 15 min in ice and washed twice in PBS. Dry pellets were flash-frozen and stored at −80 °C.

Hi-C libraries were prepared with in-nucleus ligation and processed to capture 22,076 HindIII fragments containing 31,253 annotated promoters for 18,202 protein-coding and 10,929 non-protein-coding genes (Ensembl v.75), using SureSelect target enrichment (Agilent Technologies), as described previously[14,54]. After library enrichment, a post-capture PCR amplification step was carried out with four PCR amplification cycles.

Twelve sequencing replicates from four human islet donor libraries were processed using a reported pipeline that maps di-tags against the human genome (GRCh37), filters out experimental artifacts, such as re-ligations, and removes PCR duplicates[55]. Reads from replicates from each donor were then pooled. Alignment statistics are shown in Supplementary Tables 12 and 13.

Interaction confidence scores were computed with CHiCAGO[14,15]. High-confidence interactions were defined as CHiCAGO scores >5, as described[14]. pcHi-C datasets from unrelated tissues[14] were processed identically. CHiCAGO analysis is generally performed with pooled libraries as this increases sensitivity and mitigates subsampling in individual libraries[14,15]. We assessed reproducibility across individual samples, and observed that high-confidence interaction calls showed (1) high CHiCAGO scores in individual samples, with limited overlap with distance-matched regions (Supplementary Fig. 1d), (2) pairwise Pearson ρ values of individual sample CHiCAGO scores ranging 0.62 to 0.74 (Supplementary Fig. 1e) and (3) consistent above-background scores in individual samples (Supplementary Figs. 1f,g and 5a).

ChIP-seq and ATAC-seq. ChIP and ATAC were performed as previously described[7,56], with modifications (Supplementary Note 2). Adaptor trimming of ChIP-seq reads was performed with cutadapt v.1.9.1 (options: -m 20)[57]. For ATAC-seq, low quality bases and adaptor trimming were processed using TrimGalore v.0.4.1 (options: --quality 15 --nextera). Trimmed reads were aligned to hg19 using bowtie2 v.2.1.0 (options: --no-unal) allowing no mismatches[58], retaining uniquely mapped reads (MAPQ ≥ 30) using SAMtools v.1.2 (ref. [59]), removing duplicate reads (picard v.2.6.0)[60], blacklisted regions[61] and, for ATAC-seq, mitochondrial reads. Data quality was assessed with SPP.R script from phantompeaktools[62]. ChIP-seq and ATAC-seq information is shown in Supplementary Table 1.

For histone modifications, broad enriched regions were called with MACS2 (ref. [63]) using --g hs --extsize = 300 --keep-dup all --nomodel --broad, and narrow regions were called without using --broad flag. For transcription factors and cofactors, narrow regions were called using --g hs --extsize = 300 --keep-dup all. For ATAC-seq, we used --shift 100 --extsize = 200 --keep-dup all --nomodel.

To obtain a robust set of ChIP-seq peaks, we called peaks in individual human islet samples with relaxed stringency ($P < 0.01$), and in pooled samples using a stringent threshold (false discovery rate (FDR) $q < 0.05$ for Mediator and cohesin; and $q < 0.01$ for histone modification marks). We then identified peaks present in at least three individual samples, or at least two samples if only three replicates were processed, as well as in the pooled set. For accessible chromatin sites, we called peaks at $P < 0.01$ in 13 individual samples, and FDR $q < 0.05$ from pooled samples. We then defined consistent peaks present in at least three samples as well as in the pooled set. Consistent ATAC peaks that showed multiple subpeaks in >3 islet samples were manually split, leading to $n = 241{,}481$ ATAC peaks. A final set of accessible chromatin regions ($n = 249{,}582$) was defined by adding regions lacking ATAC-seq peaks that showed either Mediator or CTCF binding ($n = 1{,}319$, $n = 9{,}596$ respectively) or were bound by at least two islet transcription factors ($n = 1{,}514$)[7]. bigWig files were generated using bamCoverage from deepTools (-e = 300 --normalizeTo1x2451960000).

Classification of human islet-accessible chromatin. We classified 249,582 consistent islet open chromatin regions using *k*-medians clustering of ChIP-seq signal distributions of H3K27ac, H3K4me1, H3K4me3, Mediator, cohesin and CTCF, using islet samples with the greatest signal to noise for these marks. Briefly, $-\log_{10}$ (*P* value) signal was calculated for each mark using 100 base pair bins across a 6-kb window centered on consistent open chromatin regions. *K*-median clustering (flexClust[64]) was used to classify open chromatin regions into 14 clusters, which were manually merged into eight clusters based on the chromatin mark enrichment patterns. Each open chromatin class was ranked by CTCF binding to highlight a subset of CTCF-bound enhancers. Post-hoc analysis showed that human islet transcription start sites defined by cap-analysis of gene expression (CAGE) were markedly enriched in regions classified as active promoters and, to a lesser extent, in class I enhancers (Fig. 1c). See Supplementary Dataset 1 for genomic locations.

PATs and enhancer–promoter assignments. We defined 16,030 promoter-associated three-dimensional spaces (PATs) as the linear space covered by all interactions originating from a pcHi-C bait, within the same islet TAD-like compartment (Supplementary Note 3).

We used PAT features to assign enhancers to promoters, following a stepwise approach such that each step was performed on unassigned enhancers from previous steps. We assigned enhancers to baits with at least one active islet promoter according to our regulome annotations (Supplementary Dataset 1) or ChromHMM analyses (Supplementary Note 6), and report target genes with average human islet RNA expression >1.5 transcripts per million (TPM) (Supplementary Dataset 2), based on the following criteria:

(1) Presence of high-confidence interactions (CHiCAGO > 5) to one or more baits, including those that cross TAD boundaries (also referred to as assignment by interaction).
(2) For enhancers with no high-confidence interactions, we defined PAT(s) in which they were contained. We did not assign enhancers to all overlapping PATs because only some active genes are regulated by enhancers, and instead only imputed orphan enhancers to PAT(s) anchored by an active promoter that already showed high-confidence interactions with other islet enhancers.
(3) For remaining enhancers located <10 kb away from a bait containing active promoter(s), we assumed that (1) this linear distance is more likely to provide functional enhancer–promoter communication than for promoters located more distally that do not show high-confidence interactions, and (2) random collisions are too frequent to detect high-confidence interactions above background noise, and thus we imputed these enhancer–promoter assignments.
(4) For the remaining enhancers that were exclusively contained within a single PAT with an active promoter, we imputed the assignment to expressed genes in that PAT bait. We refer to assignment criteria 2–4 as imputations here.

Enhancer–promoter assignments can be found in Supplementary Dataset 2 and were validated by analysis of (1) CHiCAGO scores in imputations, (2) increased enhancer–promoter correlations, (3) islet-specificity of assigned genes, (4) concordance with eQTLs and (5) coordinated changes after exposure to varying glucose concentrations (Supplementary Note 7).

Candidate target genes of T2D/FG-associated variants. We integrated lists of T2D/FG-associated variants (Supplementary Note 8) with enhancer–promoter assignments to identify candidate target genes. We associated 555 enhancer variants from 51 loci to islet-expressed genes using high-confidence interactions and imputations. Supplementary Table 3 provides a more extensive list of 830 T2D/FG-associated variants overlapping an active enhancer or promoter, with information on connections to candidate target genes through (1) high-confidence interactions (CHiCAGO > 5), (2) moderate-confidence interactions (CHiCAGO = 2.5–5), (3) imputations, (4) indirect connections through a common hub and (5) location of actively expressed gene within 10 kb. This category also included actively transcribed genes from associated variant-containing promoters that overlap pcHi-C baits. Supplementary Table 3 additionally lists T2D/FG variants overlapping a promoter-interacting region that do not overlap an annotated regulatory element.

Cell-based genome and epigenome editing. Experimental validation of T2D-relevant enhancer–promoter assignments in EndoC-βH3 cells[20] is described in Supplementary Note 10 and Nature Protocol Exchange (I.C. and A. Beucher, https://protocolexchange.researchsquare.com/article/nprot-7395).

Classification of PATs based on enhancer content. We defined enhancer-rich PATs as those with three or more class I enhancers (Supplementary Fig. 8b). This was supported by logistic regression analysis (Supplementary Note 11) showing that the number of class I enhancers assigned to a PAT was independently predictive of islet-selective expression of PAT genes. This effect was optimized with PATs with ≥3 assigned class I enhancers (Supplementary Fig. 9).

Enhancer hubs. Enhancer-rich PATs were frequently interconnected through one or more shared enhancers (42.4% of all active enhancers had high-confidence interactions with >1 bait). We thus merged enhancer-rich PATs with other PATs that were connected by one or more common enhancers through high-confidence interactions (CHiCAGO > 5). For 99.5% of hubs, all hub components were restricted to one chromosome. Alternative definitions of hubs were created to test how, (1) the number of enhancers in enhancer-rich PATs, (2) the inclusion of enhancer–gene imputed assignments and (3) criteria to merge PATs, influence definitions of enhancer hubs (Supplementary Fig. 9).

To annotate hub genes, we considered annotated promoters of genes with median RNA expression >1.5 TPM in human islets. In a few cases ($n=426$), pcHi-C bait fragments contained active enhancers that established high-confidence pcHi-C interactions with non-baited fragments containing active islet promoters, which were also considered as constituents of islet hubs. A list of human islet enhancer hubs is presented in Supplementary Dataset 5. Functional enrichments of hub Ensembl genes were performed with Enrichr[65].

The analysis of correlated hub promoter and enhancer activity, and islet selectivity of enhancer interactions, is described in Supplementary Note 12.

3D modeling of hubs. 3D modeling and analysis of enhancer hubs were partly based on previously described methods[66,67], and are described in Supplementary Note 13.

T2D/FG variant enrichments in regulatory annotations. Variant set enrichment[68] was used to compute the enrichment of T2D- and FG-associated variants in regulatory annotations, using lead SNPs from 109 loci (Supplementary Table 9), and is described in Supplementary Note 14.

GWAS meta-analysis of insulin secretion. A total of 7,807 individuals from four population studies were included in these analyses: the Inter99 study (ClinicalTrials ID no.: NCT00289237) ($n=5,305$)[38], the Health2008 cohort ($n=605$)[69], the 1936 Birth Cohort ($n=709$)[70] and the ADDITION-Pro cohort ($n=1,188$)[71]. All study participants gave informed consent and studies were approved by the appropriate ethical committees in accordance with the scientific principles of the Helsinki Declaration II.

In all cohorts, glucose-stimulated insulin secretion was evaluated by measurement of plasma glucose and serum insulin at 0, 30 and 120 min during a 75-g OGTT. We calculated insulinogenic index = (s-insulin at 30 min [pmol l^{-1}] − fasting s-insulin [pmol l^{-1}])/p-glucose at 30 min (mmol l^{-1}). Individuals with known diabetes were excluded.

Two sample sets (Inter99 and Health2008) were genotyped by Illumina OmniExpress array and others by Illumina CoreExome array. Genotypes were called by the Illumina GenCall algorithm. Genotype data were filtered for variants with call rate <98% and Hardy–Weinberg equilibrium $P<10^{-5}$. Samples were excluded if they were ethnic outliers, had mismatch between genetic and phenotypic sex or had a call rate <95%.

Genotype data were imputed to the Haplotype Reference Consortium reference panel v.1.1 (ref. [72]) at the Michigan Imputation Server using Minimac3 after phasing genotypes into haplotypes with Eagle2 (ref. [73]). Post-imputation SNP filtering included exclusion of variants with minor allele frequency (MAF) < 0.01 or info score < 0.70. In each cohort, association analysis was performed by applying a linear regression model including age and sex as covariates via SNPTEST[74]. The phenotype was rank-normalized within each cohort before analysis. A fixed-effects meta-analysis implemented in the R package meta[75] was finally performed.

Heritability estimates. See also Supplementary Notes 16 and 17. To estimate the polygenic contribution of different genomic annotations to GWAS-based heritability of T2D and related traits, we applied the stratified LD Score regression method[36,76]. This method leverages the relationship between LD structure and association test statistics to estimate the average per-SNP contribution to heritability (τ_c coefficient) of functional genomic categories. We used a panel of 53 baseline genomic annotations[36,76], and interrogated a broad range of islet regulatory annotations, including enhancer hubs, as well as control annotation sets, such as central nervous system functional annotations, random non-open chromatin regions and pseudo-enhancer hubs. We provide the per-SNP heritability τ_c coefficient for each regulatory annotation. To facilitate comparisons across traits and annotations, we normalized the τ_c estimates by dividing them by the LD Score heritability for each phenotype, and multiplied by 10^7. To correct for multiple testing, we generated τ_c q values (FDR-adjusted P values calculated from the Z-scores of the τ_c coefficients) with the qvalue R package over 17 functional categories and six traits. The FDR significance threshold was set at 0.05.

PRS. See also Supplementary Note 18. We created PRS based on T2D GWAS summary statistics from 70kfort2d[42] (base dataset). UK Biobank individuals[43] were used as the target datasets, which comprised training and testing datasets.

To select markers for PRS we first considered all genetic markers that were used as input for phasing and genotype imputation by UK Biobank, and filtered for variants with MAF ≥ 5% and imputation quality score >0.8. We then reconciled the base and target datasets by looking at the variant overlap between summary statistics and the imputed UK Biobank data, discarding variants showing allele inconsistency between both datasets. We also removed those located in the major histocompatibility complex region, resulting in a final collection of 5,352,737 variants.

We excluded UK Biobank individuals with: (1) excess of relatives (showing >10 putative third-degree relatives, as provided by UK Biobank), (2) greater than third-degree of relatedness (from each pair of related individuals we excluded the subject with the highest missing rate for a set of high-quality markers, as provided by UK Biobank), (3) no gender information, (4) International Classification of Diseases (ICD-10) codes E10 (insulin-dependent diabetes mellitus), E13 (other specified diabetes mellitus) and E14 (unspecified diabetes mellitus), (5) no BMI information. T2D cases were defined by the E11 ICD-10 code.

The sample size of UK Biobank qualifying individuals was 377,981 controls and 15,764 cases, which were divided into training and testing datasets. For the training dataset, we included only control subjects with age at recruitment ≥55 years and no family history of diabetes mellitus, yielding a final training dataset sample size of 6,305 T2D cases and 73,922 controls. The remaining 236,236 individuals were used as a test dataset, and were not filtered by age or family history.

PRS models were calculated from the above-mentioned base and training datasets using the PRsice software[77] with default settings and clumping parameters (--clump-r2 0.6 --clump-p 0.01). We included 11 covariates in the analysis: the seven principal components provided by UK Biobank investigators as well as BMI, age at recruitment, batch information and sex.

We generated PRS models based on the following common genetic variants: (1) the entire genome-wide set shared by the training and testing dataset (total of 5,352,737 variants; 1,152 qualifying variants in the model), (2) variants overlapping hub pcHi-C baits and enhancers (total variants = 86,158; 179 qualifying variants in the model), (3) variants overlapping islet open chromatin regions, excluding islet hub baits and enhancers and those in LD ($r^2>0.1$) with islet hub index variants (total variants = 269,342; 160 qualifying variants in the model), (4) the remaining genome, excluding variants overlapping islet hub regions or other islet open chromatin regions or those in LD with islet hub index variants (total variants = 4,913,005; 355 qualifying variants in the model).

To enable comparisons of PRS effects in stratified subgroups, we created regions with similar genomic space and distribution as hubs (pseudo-enhancer hubs). Pseudo-enhancer hubs were generated essentially as for LD Score regression analysis, except that they resembled hubs used for PRS, in that they contained all enhancers and baits of hubs. We created 100 sets of ~1,000 pseudo-enhancer hubs by shuffling hub pcHi-C baits and their assigned enhancer fragments across randomly selected size-matched TADs, excluding those in TADs with real hubs or if they crossed TAD boundaries. We then built PRS models using variants overlapping these pseudo-baits and pseudo-enhancers (average of 265 qualifying variants per pseudo-hub PRS model).

To assess PRS, we first stratified the entire UK Biobank test dataset ($n=236,236$) in 40 bins, each one containing 2.5% of individuals ranked by PRS score. To enable assessment of PRS for T2D stratified by BMI and age of diagnosis, all measures of T2D frequency were performed exclusively with the 6,127 T2D cases with known age of diagnosis, and diagnosed after 20 years of age, and all 226,777 controls, which were censored at enrollment to UK Biobank. We calculated either the T2D frequency ratios in top versus bottom bin, or the odds ratio for T2D in individuals with highest PRS scores (top 2.5% bin) versus remaining individuals in the same age and BMI categories, using a logistic regression model adjusted for the first seven principal components of ancestry, sex, age, BMI and batch information. We expressed values as Z-scores relative to the distribution of 100 sets of pseudo-hub PRS to enable comparisons of hub scores in the different stratified subgroups.

Data visualization. Data from this study can be visualized in the following browsers: islet regulome browser (http://isletregulome.org/isletregulome)[17], CHiCP browser (https://www.chicp.org)[16] and WashU Epigenome browser using this session link: http://epigenomegateway.wustl.edu/browser/?genome=hg19&session=62hGf7nfcS&statusId=140947077

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Raw sequence reads from pcHi-C, RNA-seq, ChIP-seq, ATAC-seq and 4C-seq are available from EGA (https://www.ebi.ac.uk/ega), under accession number EGAS00001002917. Processed data files for islet pcHi-C interactions, islet regulome annotations, enhancer–promoter assignments, hub coordinates and components and 3D model videos are provided as supplementary data. The robust set of ATAC-seq peaks, consistent set of Mediator, cohesin, H3K27ac and H3K4me3 peaks, list of islet super-enhancers defined using ROSE algorithm, islet regulome, ChromHMM segmentation model, list of islet TAD-like domains, PATs and the list of high-confidence pcHi-C interactions are provided as Supplementary Datasets and are also deposited at https://www.crg.eu/en/programmes-groups/ferrer-lab#datasets.

Code availability

Custom code in this manuscript is available upon request.

References

50. Melzi, R. et al. Role of CCL2/MCP-1 in islet transplantation. *Cell Transplant.* **19**, 1031–1046 (2010).
51. Kerr-Conte, J. et al. Upgrading pretransplant human islet culture technology requires human serum combined with media renewal. *Transplantation* **89**, 1154–1160 (2010).

52. Bucher, P. et al. Assessment of a novel two-component enzyme preparation for human islet isolation and transplantation. *Transplantation* **79**, 91–97 (2005).
53. Ricordi, C., Lacy, P. E., Finke, E. H., Olack, B. J. & Scharp, D. W. Automated method for isolation of human pancreatic islets. *Diabetes* **37**, 413–420 (1988).
54. Nagano, T. et al. Comparison of Hi-C results using in-solution versus in-nucleus ligation. *Genome Biol.* **16**, 175 (2015).
55. Wingett, S. et al. HiCUP: pipeline for mapping and processing Hi-C data. *F1000Res.* **4**, 1310 (2015).
56. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
57. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, 10–12 (2011).
58. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
59. Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
60. McKenna, A. et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
61. Dunham, I. et al. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
62. Kharchenko, P. V., Tolstorukov, M. Y. & Park, P. J. Design and analysis of ChIP-seq experiments for DNA-binding proteins. *Nat. Biotechnol.* **26**, 1351–1359 (2008).
63. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
64. Leisch, F. A toolbox for K-centroids cluster analysis. *Comput. Stat. Data Anal.* **51**, 526–544 (2006).
65. Kuleshov, M. V. et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–W97 (2016).
66. Baù, D. & Marti-Renom, M. A. Genome structure determination via 3C-based data integration by the Integrative Modeling Platform. *Methods* **58**, 300–306 (2012).
67. Di Stefano, M., Paulsen, J., Lien, T. G., Hovig, E. & Micheletti, C. Hi-C-constrained physical models of human chromosomes recover functionally-related properties of genome organization. *Sci. Rep.* **6**, 35985 (2016).
68. Ahmed, M. et al. Variant Set Enrichment: an R package to identify disease-associated functional genomic regions. *BioData Min.* **10**, 9 (2017).
69. Thuesen, B. H. et al. Cohort Profile: the Health2006 cohort, research centre for prevention and health. *Int. J. Epidemiol.* **43**, 568–575 (2014).
70. Drivsholm, T., Ibsen, H., Schroll, M., Davidsen, M. & Borch-Johnsen, K. Increasing prevalence of diabetes mellitus and impaired glucose tolerance among 60-year-old Danes. *Diabet. Med.* **18**, 126–132 (2001).
71. Johansen, N. B. et al. Protocol for ADDITION-PRO: a longitudinal cohort study of the cardiovascular experience of individuals at high risk for diabetes recruited from Danish primary care. *BMC Public Health* **12**, 1078 (2012).
72. McCarthy, S. et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).
73. Loh, P. R., Palamara, P. F. & Price, A. L. Fast and accurate long-range phasing in a UK Biobank cohort. *Nat. Genet.* **48**, 811–816 (2016).
74. Marchini, J. & Howie, B. Genotype imputation for genome-wide association studies. *Nat. Rev. Genet.* **11**, 499–511 (2010).
75. Schwarzer, G. meta: an R package for meta-analysis. *R. News* **7**, 40–45 (2007).
76. Bulik-Sullivan, B. K. et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
77. Euesden, J., Lewis, C. M. & O'Reilly, P. F. PRSice: polygenic risk score software. *Bioinformatics* **31**, 1466–1468 (2015).

nature research

Corresponding author(s): Professor Jorge Ferrer

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	DIAGRAM Metabochip Summary Statistics: http://diagram-consortium.org/downloads.html 70KforT2D Summary Statistics: http://cg.bsc.es/70kfort2d/ Acute Insulin Response based on intravenous glucose tolerance tests GWAS meta-analysis (http://www.t2diabetesgenes.org/data/ Attention deficit/hyperactivity disorder GWAS meta-analysis: http://ipsych.au.dk/downloads/data-download-agreement-adhd-european-ancestry-gwas-june-2017 NHGRI-EBI GWAS Catalog: https://www.ebi.ac.uk/gwas/ Epigenome Roadmap Consortium H3K27ac ChIPs: egg2.wustl.edu/roadmap/data/byFileType/alignments/consolidated/ The Human BodyMap 2 Project: ArrayExpress ID: E-MTAB-513 Human islet RNA-Seq datasets: GEO accession number GSE50244
Data analysis	CHiCAGO R package v1.0.1: https://bioconductor.org/packages/release/bioc/html/Chicago.html Trimgalore v0.4.1: https://github.com/FelixKrueger/TrimGalore Bowtie2: http://bowtie-bio.sourceforge.net/bowtie2/index.shtml Picard v2.6.0: https://broadinstitute.github.io/picard/ MACS2: https://github.com/taoliu/MACS DeepTools2: https://deeptools.readthedocs.io/en/develop/content/installation.html BedTools v2.25.0: https://bedtools.readthedocs.io ChromHMM v1.11: http://compbio.mit.edu/ChromHMM/ FlexClust: https://cran.r-project.org/package=flexclust

ROSE algorithm: https://bitbucket.org/young_computation/rose/src/master/
Michigan imputation server: https://imputationserver.sph.umich.edu/index.html
STAR v2.3.0: https://github.com/alexdobin/STAR
WASP pipeline v0.2.2: https://github.com/bmvdgeijn/WASP/blob/master/CHANGELOG.md
VerifyBamID v1.1.3: https://github.com/statgen/verifyBamID/releases
Rsubread R package v1.5.0: http://bioconductor.org/packages/release/bioc/html/Rsubread.html
edgeR v3.20.9: https://bioconductor.org/packages/release/bioc/html/edgeR.html
PEER: https://www.sanger.ac.uk/science/tools/peer
Matrix eQTL: https://cran.r-project.org/web/packages/MatrixEQTL/index.html
METAL: http://csg.sph.umich.edu/abecasis/metal/download/
qvalue R package v2.16.0: https://bioconductor.org/packages/release/bioc/html/qvalue.html
DESeq2 v1.10.1: http://bioconductor.org/packages/release/bioc/html/DESeq2.html
PLINK v1.9.0: https://github.com/chrchang/plink-ng/
VSE R package v.0.99: https://cran.r-project.org/web/packages/VSE/index.html
PRSice v1 (February of 2018 release): http://prsice.info
TADdyn: custom code available upon request.
NetworkAnalyzer: http://manual.cytoscape.org/en/stable/Network_Analyzer.html
ClusterMaker2: http://www.rbvi.ucsf.edu/cytoscape/clusterMaker2/
Chimera: http://www.cgl.ucsf.edu/chimera/

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

-Raw sequence reads from pcHi-C, RNA-seq, ChIP-seq, ATAC-seq and 4C-seq are available from EGA (https://www.ebi.ac.uk/ega), under accession number EGAS00001002917.
-Processed data files for islet pcHi-C interactions, islet regulome annotations, enhancer-promoter assignments, hub coordinates and components and 3D model videos are provided as supplementary data. The robust set of ATAC-Seq peaks, consistent set of Mediator, cohesin, H3K27ac and H3K4me3 peaks, list of islet super-enhancers defined using ROSE algorithm, islet regulome, ChromHMM segmentation model, list of islet TAD-like domains, PATs and the list of high-confidence pcHiC interactions are provided as Supplementary Data Sets and also deposited at https://www.crg.eu/en/programmes-groups/ferrer-lab#datasets
-Data from this study can be visualized in the following browsers: Islet regulome browser (http://isletregulome.org/isletregulome), CHiCP browser (https://www.chicp.org) and WashU Epigenome browser using this session link:
http://epigenomegateway.wustl.edu/browser/?genome=hg19&session=62hGf7nfcS&statusId=140947077

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	At least 3 different human islet donors were used for each epigenomic map, (13 for ATAC-Seq, 4 for pcHiC, 3 at least for each ChIP and 7 for each RNA-Seq condition, see Supplementary Table 13), which is consistent with the number of donors used by international consortia for chromatin profiling (p.e. The Roadmap Epigenomics Consortium, Nature 2015). For CRISPR/Cas9 perturbations we designed all experiments with at least two different targeting and non-targeting pairs of guide RNAs (for deletion experiments) or at least three targeting and non-targeting individual guide RNAs (for CRISPRa and CRISPRi) per target region. All genomic regions deleted/perturbed by CRISPR/Cas9 in this study were targeted a minimum of three times.
Data exclusions	Exclusions are detailed in Methods section: for CRISPR/Cas9 experiments, all data acquired is available in Supplementary Data Set 4. Because we targeted TSS regions as internal positive controls, we excluded from the analysis TSS guide RNAs that clearly did not yield upregulation of the target gene, since they were non-informative. Outliers identified by Grubbs' test ($P < 0.05$) were also excluded from the statistical analysis.
Replication	pcHiC high-confidence interactions were called on 4 pooled human islet samples but we also evaluated to what extent interactions detected in the pooled set are detected in individual samples: 89% of interactions all individual samples showed CHiCAGO scores above the 95% confidence interval of random distance-matched regions (Supplementary Figure 1d). Also, CHiCAGO scores in all pairs of samples showed high Pearson correlation values (0.62-0.84) (Supplementary Figure 1e). For CRISPR/Cas9 deletions, we used two different guide RNA pairs per region, each of them tested in at least two independent experiments.

	For CRISPRa and CRISPRi experiments, we always used a minimum of three different targeting guide RNAs per region, which we tested in two to three independent experiments. We observed that the great majority of guide RNAs used in CRISPRa and CRISPRi experiments yielded comparable results.
Randomization	Not applicable. All human islets used in this study came from de-identified cadaveric donors without previous history of glucose intolerance.
Blinding	Not applicable. All human islets used in this study came from de-identified cadaveric donors without previous history of glucose intolerance.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☒	☐ Unique biological materials
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants

Methods

n/a	Involved in the study
☐	☒ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	Rabbit polyclonal Anti-Histone H3 (acetyl K27) antibody - ChIP Grade (ab4729) from Abcam (lot GR125454-2). Concentration: 1µg/ChIP reaction; Rabbit polyclonal to CRSP1/TRAP220 (Mediator) - A300-793A from Bethyl Laboratories (lot #3). Concentration: 3µg/ChIP reaction; Rabbit Polyclonal to SMC1 (Cohesin) - A300-055A from Bethyl Laboratories (lot #5).
Validation	The use of antibody for H3K27ac in human islets for ChiP-Seq was validated in Pasquali et al 2014 (PMID: 24413736). The use of Mediator and cohesin antibodies was validated in Kagey et al 2010 (PMID:20720539) and specificity in beta cells in vitro was tested using Western blot.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	The human pancreatic beta cell lines EndoC βH1 and βH3 were generated and provided by Philippe Ravassard, a co-author of this study. The human hepatocellular carcinoma cell line HepG2 was obtained from ATCC. The transformed human embryonic kidney cell line 293FT was purchased from Invitrogen.
Authentication	Cell lines were authenticated by routinary morphological inspection and qPCR measurement of cell-type-specific markers. The β-cell lines EndoC βH1 and βH3 were further authenticated by comparing their interactome, epigenome and transcriptome to those of human islets (which contain about 70% of β-cells).
Mycoplasma contamination	All cell lines tested negative for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used.

ChIP-seq

Data deposition

☒ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☒ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links *May remain private before publication.*	https://www.crg.eu/en/programmes-groups/ferrer-lab#datasets
Files in database submission	High-confidence pcHiC interactions in human islets (washU format): PI_Merged_washU_text.txt Human islet enhancer hubs: Islet_enhancer_hubs.bed Human islet super-enhancers defined using ROSE algorithm: Islet_super_enhancers.bed Human islet regulome: Islet_regulome_simplified.bed.zip ChromHMM segmentation model (15-states): Islet_ChromHMM.bed.zip Human islet TAD-like domains: Islet_TAD-like_domains.bed Human islet PATs (Promoter Associated Three Dimensional Spaces): Islet_PATs.bed

Robust set of ATAC-Seq peaks in human islets: ATAC_consistent_peaks.bed
Consistent set of Mediator peaks in human islets: MED1_consistent_peaks_q001_r0.5.bed
Consistent set of cohesin peaks in human islets: SMCA1_consistent_peaks_q001_r0.5.bed
Consistent set of CTCF peaks in human islets: CTCF_consistent_peaks_q001_r0.5.bed
Consistent set of H3K27ac peaks in human islets: H3K27Ac_consistent_peaks_q005_r0.5.bed
Consistent set of H3K4me3 peaks in human islets: H3K4ME3_consistent_peaks_q005_r0.5.bed

Genome browser session
(e.g. UCSC)

A) islet regulome (http://isletregulome.org/regulomebeta/), B) CHICP Browser (https://www.chicp.org) and C) - WashU Epigenome browser using this session: http://epigenomegateway.wustl.edu/browser/?genome=hg19&session=62hGf7nfcS&statusId=140947077

Methodology

Replicates

- 4 human islet samples for pcHiC experiment.
- 6 Mediator ChiP-Seq datasets in human islets
- 3 Cohesin ChIP-Seq datasets in human islets
- 17 H3K27ac ChiP-Seq datasets in human islets cultured in high-glucose conditions
- 7 H3K27ac ChIP-Seq datasets in human islets cultured in low glucose conditions
- 13 ATAC-Seq datasets in human islets
- 7 RNA-Seq datasets in human islets cultured in high-glucose conditions
- 7 RNA-Seq datasets in human islets cultured in low-glucose conditions
- 2 4C-Seq datasets in EndoC-BH1.

Sequencing depth

Sequencing depth and read length for all ChIP-Seq, ATAC-Seq and RNA-Seq datasets is presented in Supplementary Table 13.

Antibodies

Rabbit polyclonal Anti-Histone H3 (acetyl K27) antibody - ChIP Grade (ab4729) from Abcam (lot GR125454-2). Concentration: 1µg/ChIP reaction; Rabbit polyclonal to CRSP1/TRAP220 (Mediator) - A300-793A from Bethyl Laboratories (lot #3). Concentration: 3µg/ChIP reaction; Rabbit Polyclonal to SMC1 (Cohesin) - A300-055A from Bethyl Laboratories (lot #5).

Peak calling parameters

For ChIP-seq reads from histone modification marks, broad regions of enrichment were called using the options --g hs --extsize=300 --keep-dup all --nomodel --broad and narrow regions of enrichment were called without using --broad flag. For TF and co-factor ChIP-seq reads, narrow regions of enrichment were called using –g hs –extsize=300 --keep-dup all. For ATAC-seq reads, we used the following options --shift 100 --extsize=200 --keep-dup all --nomodel.

Data quality

For ChIP-Seq we first used MACS2 to call peaks in individual human islet samples with a relaxed stringency threshold ($p < 0.01$). Then, we pooled all the biological replicates for each mark and identified pooled peaks using a stringent threshold (FDR $q < 0.05$ for Mediator and cohesin and $q < 0.01$ for histone modification marks). Finally, we identified a set of consistent peaks when they were present in at least 2 individual human islet samples (out of 3) or at least 3 human islet samples (if we had more than 3 replicates) as well as in the pooled set

Software

Illumina TruSeq adapters were removed from ChIP-seq reads using cutadapt 1.9.1 (options: -m 20). In ATAC-seq reads, low quality bases were trimmed using Trimgalore 0.4.1 (options --quality 15 --nextera), which also removes Nextera transposase adapters (https://github.com/FelixKrueger/TrimGalore). Trimmed reads were aligned to hg19 genome build using bowtie2 2.1.0 (options: --no-unal) allowing no mismatches. Aligned reads were filtered to retain only uniquely mapped reads (MAPQ>=30) using samtools 1.2 14 and duplicate reads were removed using picard 2.6.0 15. Reads mapping to blacklisted regions 16 were also removed using BEDTools 2.13.3. Peaks were called with MACS2.