Published in final edited form as:
Nat Genet. 2019 July 01; 51(7): 1137–1148. doi:10.1038/s41588-019-0457-0.

Human pancreatic islet 3D chromatin architecture provides insights into the genetics of type 2 diabetes

Irene Miguel-Escalada[#1,2,3,4], **Silvia Bonàs-Guarch**[#1,2,3,4], **Inês Cebola**[#1], **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 García-Hurtado**[2,3,4], **Anthony Beucher**[1], **Ignasi Morán**[1,16], **Lorenzo Pasquali**[4,7,8], **Mireia Ramos-Rodríguez**[8], **Emil V.R. Appel**[9], **Allan Linneberg**[10,11], **Anette P. Gjesing**[9], **Daniel R. Witte**[12,13], **Oluf Pedersen**[9], **Niels Grarup**[9], **Philippe Ravassard**[14], **David Torrents**[15,16], **Josep M. Mercader**[16,17,18], **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**[9], **Marc A. Marti-Renom**[5,15,30,31], **Peter Fraser**[6,32], and **Jorge Ferrer**[1,2,4,*]

[1]Section of Epigenomics and Disease, Department of Medicine, and National Institute for Health Research (NIHR) Imperial Biomedical Research Centre, Imperial College London, London W12 0NN, UK [2]Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Dr. Aiguader 88, Barcelona 08003, Spain [3]Genomic Programming of Beta-cells Laboratory, Institut d'Investigacions August Pi i Sunyer (IDIBAPS), 08036 Barcelona, Spain [4]CIBER de Diabetes y Enfermedades Metabólicas Asociadas, Spain [5]CNAG-CRG, Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology (BIST), Baldiri Reixac 4, Barcelona 08028, Spain [6]Nuclear Dynamics Programme, The Babraham Institute, Babraham Research Campus, Cambridge CB22 3AT, UK [7]Josep Carreras Leukaemia Research Institute, Campus ICO-Germans Trias i Pujol, Ctra de Can Ruti, Camí de les Escoles s/n, Badalona, 08916, Spain [8]Endocrine Regulatory Genomics Lab, Germans Trias i Pujol University Hospital and Research Institute, 08916 Badalona, Spain [9]Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark [10]Center for Clinical Research and Disease Prevention, Bispebjerg and Frederiksberg Hospital, The Capital Region, Copenhagen, Denmark [11]Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark [12]Department of Public Health, Aarhus University, Aarhus, Denmark [13]Danish Diabetes Academy, Odense, Denmark [14]Université Sorbonne, UPMC Univ Paris 06, Inserm, CNRS, Institut du cerveau et de la moelle (ICM) – Hôpital Pitié-Salpêtrière, Boulevard de l'Hôpital, Paris

[*] j.ferrer@imperial.ac.uk.

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.Pi., 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., T.H., I.P., and L.G. provided genetics data. M.R. and L.Pa. created software resources. I.C. and A.B. developed genome-editing methods. M.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 Statement

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

Europe PMC Funders Author Manuscripts

F-75013, France [15]ICREA, Pg. Lluís Companys 23, 08010 Barcelona, Spain [16]Barcelona Supercomputing Center (BSC), Joint BSC-CRG-IRB Research Program in Computational Biology, 08034 Barcelona, Spain [17]Programs in Metabolism and Medical & Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA [18]Diabetes Unit and Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA 02114, USA [19]Diabetes Research Institute (SR-DRI), IRCCS San Raffaele Scientific Institute, Via Olgettina 60, 20132 Milan, Italy [20]Vita-Salute San Raffaele University, Milan [21]Cell Isolation and Transplantation Center, University of Geneva, 1211 Geneva 4, Switzerland [22]Department of Medicine, Leiden University Medical Center, Box 9600, 2300 RC Leiden, Netherlands [23]Hubrecht Institute/KNAW, 85164 3508 AD Utrecht, the Netherlands [24]European Genomic Institute for Diabetes, Inserm UMR 1190, Lille 59800, France [25]Department of Biological Psychology, Vrije Universiteit Amsterdam, The Netherlands [26]Amsterdam Public Health research institute, 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 W12 0NN, UK [29]Department of Clinical and Experimental Medicine, University of Surrey, Guildford, GU2 7XH, UK [30]Universitat Pompeu Fabra (UPF), 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, Florida 32303, USA

[#] These authors contributed equally to this work.

Abstract

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 3D space. Furthermore, their target genes are often unknown. We have now created promoter capture Hi-C maps in human pancreatic islets. This linked diabetes-associated enhancers with their target genes, often located hundreds of kilobases away. It also revealed >1300 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 GWAS signals.

Introduction

Type 2 diabetes (T2D) 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 DNA variants located in the non-coding genome [2,3]. T2D susceptibility variants are enriched in active islet enhancers that cluster in linear genome maps – variably 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 various common diseases[5,7–11]. So far, however, genome-wide maps of

enhancer clusters have been largely defined with unidimensional epigenome maps, which do not necessarily reflect the capacity of enhancers to cluster in three-dimensional (3D) space, as shown for well characterized loci such as *Hbb* (β-globin) and *Hoxd*[12,13]. Linear maps also do not reveal the target genes of enhancers, which are often separated by hundreds of thousands of base pairs. This poses a need to obtain accurate representations or 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 ~1300 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 validate functional interactions between enhancers bearing T2D risk variants and their target genes. Finally, we show that islet hubs are not only enriched for T2D association signals, but 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] (Figure 1a,b and Supplementary Figure 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 Figure 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 Figure 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 ChIP-seq datasets (Figure 1b-d, Supplementary Table 1). This enabled a subclassification of active enhancers according to Mediator, cohesin, and H3K27ac occupancy patterns (Figure 1b-d, Supplementary Data Set 1). Expectedly, promoter-interacting genomic regions were enriched in active enhancers, promoters, and CTCF-bound regions (Figure 1e, Supplementary Figure 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 Figure 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 and are often invariant across tissues (Supplementary Figure 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 (Figure 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 Figure 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 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 was defined as the space containing all pcHi-C interactions that stem from a promoter bait (Supplementary Figure 3g,h). We observed that PATs that had one high-confidence enhancer-promoter interaction were more likely to show other enhancer-promoter interactions, and exhibited chromatin features that distinguished them from other PATs (Supplementary Figure 3 i-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 (Figure 2a; see Supplementary Figure 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 Figure 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 (Figure 2a, Supplementary Data Set 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 (Figure 2b). Importantly, this was true for both high-confidence and imputed assignments (Figure 2b). Islet-selective expression was expectedly enriched in enhancer-assigned genes but not in unassigned genes from the same TAD (Supplementary Figure 3n). Furthermore, we determined 1,091 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

Europe PMC Funders Author Manuscripts

with distance-matched regions (odds ratio 3.18 and 4.36; P = $3.05x10^{-09}$ and $9.01x10^{-23}$, respectively) (Figure 2c).

We further tested enhancer-promoter assignments in a dynamic perturbation model. We exposed human islets from 7 donors to moderately low (4 mM) or high (11 mM) glucose 72 hours, 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 Figure 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 mRNA, compared with distance-matched active control genes from the same TAD (odds ratio 2.7 and 2.6, Fisher's P = 4.9 x 10^{-16} and 6.4 x 10^{-12}, for high-confidence or imputed assignments, respectively) (Figure 2d). Likewise, genes assigned to glucose-induced enhancers showed greater glucose-induction of promoter H3K27ac than distance-matched promoters in the same TAD (Figure 2e). Collectively, these 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 Figure 4a, Supplementary Data Set 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 Figure 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 (Figure 3a, Supplementary Table 3). Some of these target genes were expected based on their linear proximity to the variants (e.g. *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*) (Figure 3a, Supplementary Table 3, see isletregulome.org or www.chicp.org). Selected unexpected targets, including *ABCB9* and *STARD10,* were additionally supported by concordant eQTLs (Supplementary Figure 4c-d).

We used genome editing to validate target genes of 10 enhancers bearing T2D or FG-associated variants from 8 loci (Figure 3b, 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 (Figure 3c, Supplementary Figure 5a,b, Supplementary Table 3). This variant was previously proposed to be a regulatory variant based on 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* (Figure 3c, Supplementary Figure 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 (Figure 3d, Supplementary Figure 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 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 lays in a class I enhancer with unusually high Mediator occupancy (Supplementary Figure 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* [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 Figure 6a). Consistently, targeted deletion, functional inhibition, or stimulation of the enhancer caused selective changes in *TCF7L2* mRNA (Supplementary Figure 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 8 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 (Figure 3b, Supplementary Table 4, Supplementary Data Set 4). This included functionally validated imputed target genes, such as *VEGFA* as well as *MDGA1* and *ZFAND3* (Supplementary Figure 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 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 three-dimensional 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, Figure 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 Figure 8a and Methods). In total, we identified 2,623 enhancer-rich PATs (Supplementary Figure 8b). As noted above, many active enhancers (~40%) had interactions with ≥1 promoter (Supplementary Figure 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* (Figure 4a, Supplementary Figure 8c). Compared to alternate enhancer hub definitions, this definition maximized the enrichment of islet cell functional annotations and the number of mapped hubs (Supplementary Figure 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 Figure 8d). They are often tissue-selective interaction domains, because hub promoters had 2.8-fold higher fraction of islet-selective interactions than non-hub promoters (Wilcoxon's $P = 2.8 \times 10^{-36}$) (Supplementary Figure 8e, examples in Figure 1b, 5a, Supplementary Figures 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 (Figure 4b,c, Supplementary Table 5, Supplementary Data Set 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 ($P = 6.3 \times 10^{-8}$)(Figure 4d). Moreover, hub enhancers showed higher H3K27ac correlations with their target promoters than when were paired with non-hub promoters from the same TAD ($P = 2.2 \times 10^{-16}$) (Figure 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 vs. high glucose concentrations. Glucose-induced enhancers and mRNAs were highly enriched in hubs, compared with non-hub counterparts (Fisher's $P = 1.1 \times 10^{-7}$ and 2.2×10^{-16}, respectively). Of 297 promoters that showed glucose-induced H3K27ac, 94 were contained in hubs, and 65% of these showed glucose-induced mRNA (Supplementary Tables 6,7). 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 (Figure 4f-h, 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 Figure 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 Figure 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 Figure 8j-l) [4,7]. In fact, many hubs contained several inter-connected enhancer clusters or super-enhancers (Supplementary Figure 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 lncRNA *HI-LNC57* (Figure 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 (Figure 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 pcHiC 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 co-localization 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 (Figure 5b,c, Supplementary Figure 10b,c, Supplementary Videos 1-2). Quantitative analysis of *ISL1* and six other T2D-relevant hubs showed analogous tissue-specific aggregation of hub enhancers and promoters (Supplementary Figures 10d-I, 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) (Figure 6a, Supplementary Figure 11a, 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 (Figure 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 Figure 11b,c), *VPS13C, C2CD4A* and *C2CD4B* (Supplementary Figure 12) and *GLIS3*

(Supplementary Figure 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 (Figure 7a), and in islet enhancer hub class I enhancers, rather than in other active enhancers (Figure 7b, Supplementary Figure 9, 14a-f, 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 (Figure 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 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 (Figure 7e, Supplementary Figure 15a, 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 (Figure 7e, Supplementary Figure 15a-f, 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 based on oral glucose tolerance tests[38] (Supplementary Table 10). In sharp contrast, islet hub variants showed no enrichment for HOMA-IR, an estimate of insulin resistance (Supplementary Figure 15e). Of note, significant heritability enrichments were generally also observed for enhancer clusters, stretch enhancers, or super-enhancer annotations, yet estimates were consistently larger for hub enhancers (Figure 7e, Supplementary Figure 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 if 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 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 [43–45]. 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% (Figure 7f).

Next, we created PRS models that contained DNA variants from either (a) islet hub enhancers and promoters (1.6% of the genome), (b) all other islet open chromatin regions (5.0% of the genome), and (c) the rest of the genome. Despite that islet hub regions encompass < 2% of the genome, the T2D risk ratio –defined as the T2D frequency in the top vs. 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)(Figure 7f, 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 (Figure 7e).

Although the genome-wide PRS model expectedly shows higher risk ratios than islet hub PRS models (Figure 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 at 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 1000 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 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 Figure 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 Figure 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 < 60 years of age and BMI < 35 (Figure 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 (Figure 7g, see Supplementary Figure 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 that did not differ from pseudo-hub genomic regions (odds ratio = 1.52, Z = 0.37) (Figure 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.

Europe PMC Funders Author Manuscripts

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 show 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 >1300 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 frequencies46, with numerous well-characterized chromatin hubs[12,13], and with evolutionary conserved noncoding sequence blocks[47,48]. 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 common diseases[35]. Because T2D pathophysiology is heterogeneous and multiorganic [37,49,50], 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This research was supported by the National Institute for Health Research (NIHR) Imperial Biomedical Research Centre. Work was funded by grants from the Wellcome Trust (WT101033 to J.F. and WT205915 to I.P.), Horizon 2020 (Research and Innovation Programme 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 (789055 to J.F., 609989 to M.A.M-R.). Marató TV3 (201611, to J.F., M.A.M-R.), Ministerio de Ciencia Innovación y Universidades (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 (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 (2017/1641 to I.P.), Biobanking and Biomolecular Resources Research Infrastructure (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 (SEV-2012-0208). Human islets were provided through the European islet distribution program for basic research supported by JDRF (3-RSC-2016-160-I-X). We thank Natalia Ruiz-Gomez for technical assistance, Rodrigo Liberal Fernandes, Thomas Thorne (University of Reading), and Alvaro Perdones-Montero (Imperial College London) for helpful discussions regarding Machine Learning approaches, Boris Lenhard and Matthias Merkenschlager (London Institute of Medical Sciences, Imperial College London), Ferenc Müller (UoB) and José Luis Gómez-Skarmeta (CABD) for critical comments on the draft; the CRG Genomics Unit, and the Imperial College High Performance Computing Service.

References

1. Chatterjee S, Khunti K, Davies MJ. Type 2 diabetes. Lancet. 2017; 389:2239–2251. [PubMed: 28190580]
2. Flannick J, Florez JC. Type 2 diabetes: genetic data sharing to advance complex disease research. Nat Rev Genet. 2016; 17:535–49. [PubMed: 27402621]
3. Fuchsberger C, et al. The genetic architecture of type 2 diabetes. Nature. 2016; 536:41–47. [PubMed: 27398621]
4. Whyte WA, et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell. 2013; 153:307–19. [PubMed: 23582322]
5. Parker SC, et al. Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. Proc Natl Acad Sci U S A. 2013; 110:17921–6. [PubMed: 24127591]
6. Gaulton KJ, et al. A map of open chromatin in human pancreatic islets. Nat Genet. 2010; 42:255–9. [PubMed: 20118932]
7. Pasquali L, et al. Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. Nat Genet. 2014; 46:136–143. [PubMed: 24413736]
8. Cohen AJ, et al. Hotspots of aberrant enhancer activity punctuate the colorectal cancer epigenome. Nat Commun. 2017; 8
9. Farh KK, et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. Nature. 2015; 518:337–43. [PubMed: 25363779]
10. Hnisz D, et al. Super-enhancers in the control of cell identity and disease. Cell. 2013; 155:934–47. [PubMed: 24119843]
11. Vahedi G, et al. Super-enhancers delineate disease-associated regulatory nodes in T cells. Nature. 2015; 520:558–62. [PubMed: 25686607]
12. Montavon T, et al. A regulatory archipelago controls Hox genes transcription in digits. Cell. 2011; 147:1132–45. [PubMed: 22118467]
13. Patrinos GP, et al. Multiple interactions between regulatory regions are required to stabilize an active chromatin hub. Genes Dev. 2004; 18:1495–509. [PubMed: 15198986]
14. Javierre BM, et al. Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. Cell. 2016; 167:1369–1384 e19. [PubMed: 27863249]
15. Cairns J, et al. CHiCAGO: robust detection of DNA looping interactions in Capture Hi-C data. Genome Biol. 2016; 17:127. [PubMed: 27306882]
16. Schofield EC, et al. CHiCP: a web-based tool for the integrative and interactive visualization of promoter capture Hi-C datasets. Bioinformatics. 2016; 32:2511–3. [PubMed: 27153610]
17. Mularoni L, Ramos-Rodriguez M, Pasquali L. The Pancreatic Islet Regulome Browser. Front Genet. 2017; 8:13. [PubMed: 28261261]
18. Dixon JR, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature. 2012; 485:376–80. [PubMed: 22495300]
19. Nora EP, et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature. 2012; 485:381–5. [PubMed: 22495304]
20. Benazra M, et al. A human beta cell line with drug inducible excision of immortalizing transgenes. Mol Metab. 2015; 4:916–25. [PubMed: 26909308]
21. Fogarty MP, Cannon ME, Vadlamudi S, Gaulton KJ, Mohlke KL. Identification of a regulatory variant that binds FOXA1 and FOXA2 at the CDC123/CAMK1D type 2 diabetes GWAS locus. PLoS Genet. 2014; 10:e1004633. [PubMed: 25211022]
22. Thurner M, et al. Integration of human pancreatic islet genomic data refines regulatory mechanisms at Type 2 Diabetes susceptibility loci. Elife. 2018; 7
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. 2015; 11:e1005694. [PubMed: 26624892]
24. Varshney A, et al. Genetic regulatory signatures underlying islet gene expression and type 2 diabetes. Proc Natl Acad Sci U S A. 2017; 114:2301–2306. [PubMed: 28193859]

25. Scott RA, et al. An Expanded Genome-Wide Association Study of Type 2 Diabetes in Europeans. Diabetes. 2017; 66:2888–2902. [PubMed: 28566273]
26. Wood AR, 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. 2017; 66:2296–2309. [PubMed: 28490609]
27. Lyssenko V, et al. Mechanisms by which common variants in the TCF7L2 gene increase risk of type 2 diabetes. J Clin Invest. 2007; 117:2155–63. [PubMed: 17671651]
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. 2016; 59:2360–2368. [PubMed: 27539148]
29. Nobrega MA. TCF7L2 and glucose metabolism: time to look beyond the pancreas. Diabetes. 2013; 62:706–8. [PubMed: 23431017]
30. Bau D, et al. The three-dimensional folding of the alpha-globin gene domain reveals formation of chromatin globules. Nat Struct Mol Biol. 2011; 18:107–14. [PubMed: 21131981]
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. 2017; 13:e1005665. [PubMed: 28723903]
32. Gaulton KJ, et al. Genetic fine mapping and genomic annotation defines causal mechanisms at type 2 diabetes susceptibility loci. Nat Genet. 2015; 47:1415–25. [PubMed: 26551672]
33. Boyle EA, Li YI, Pritchard JK. An Expanded View of Complex Traits: From Polygenic to Omnigenic. Cell. 2017; 169:1177–1186. [PubMed: 28622505]
34. Wood AR, et al. Defining the role of common variation in the genomic and biological architecture of adult human height. Nat Genet. 2014; 46:1173–86. [PubMed: 25282103]
35. Khera AV, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nat Genet. 2018; 50:1219–1224. [PubMed: 30104762]
36. Finucane HK, et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat Genet. 2015; 47:1228–35. [PubMed: 26414678]
37. DeFronzo RA, et al. Type 2 diabetes mellitus. Nat Rev Dis Primers. 2015; 1
38. Gjesing AP, et al. Genetic and phenotypic correlations between surrogate measures of insulin release obtained from OGTT data. Diabetologia. 2015; 58:1006–12. [PubMed: 25660259]
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. 2018; 50:1505–1513. [PubMed: 30297969]
40. Khera AV, et al. Polygenic Prediction of Weight and Obesity Trajectories from Birth to Adulthood. Cell. 2019; 177:587–596 e9. [PubMed: 31002795]
41. Richardson TG, Harrison S, Hemani G, Davey Smith G. An atlas of polygenic risk score associations to highlight putative causal relationships across the human phenome. Elife. 2019; 8
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. 2018; 9
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. 2015; 12:e1001779. [PubMed: 25826379]
44. Bycroft, F C, Petkova D, Band G, Elliott LT, Sharp K, Motyer A, Vukcevic D, Delaneau O, O'Connell J, Cortes A, et al. Genome-Wide Genetic Data on ~ 500,000 UK Biobank Participants. BioRxiv. 2017
45. Bycroft C, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature. 2018; 562:203–209. [PubMed: 30305743]
46. Schmitt AD, et al. A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome. Cell Rep. 2016; 17:2042–2059. [PubMed: 27851967]
47. Harmston N, et al. Topologically associating domains are ancient features that coincide with Metazoan clusters of extreme noncoding conservation. Nat Commun. 2017; 8:441. [PubMed: 28874668]
48. Akalin A, et al. Transcriptional features of genomic regulatory blocks. Genome Biol. 2009; 10:R38. [PubMed: 19374772]
49. 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. 2018

50. Kahn SE, Cooper ME, Del Prato S. Pathophysiology and treatment of type 2 diabetes: perspectives on the past, present, and future. Lancet. 2014; 383:1068–83. [PubMed: 24315620]

Europe PMC Funders Author Manuscripts

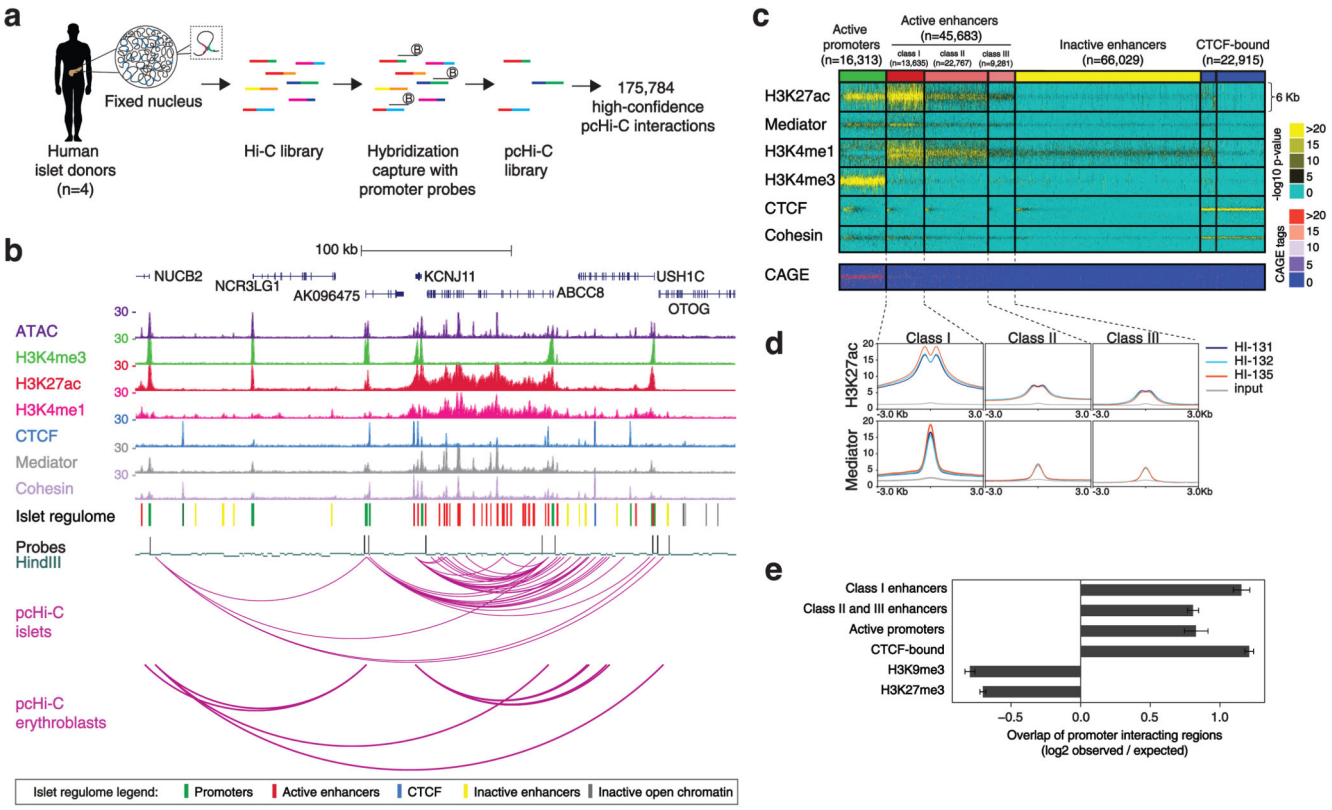


Figure 1. The promoter interactome of human pancreatic islets.

a, Overview of promoter-capture Hi-C (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 was used to define consistent open chromatin regions, which were classified with k-medians clustering based on epigenomic features. Mediator and H3K27ac binding patterns allowed sub-classification 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 Data Set 1. 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 log2 ratios (and 95% confidence intervals) over the overlaps obtained with 100 sets of distance-matched fragments. Error bars show s.d. across control sets.

Europe PMC Funders Author Manuscripts

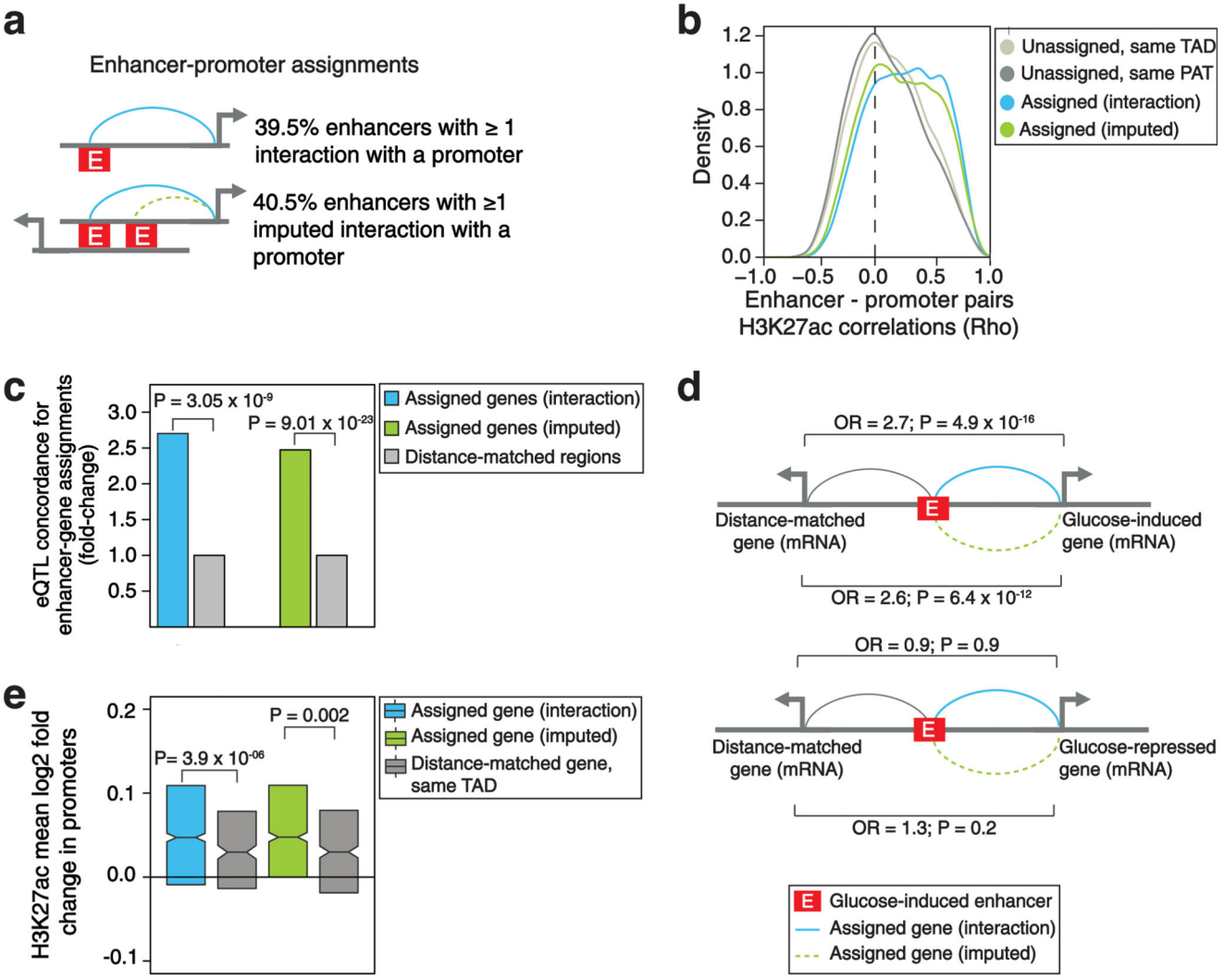
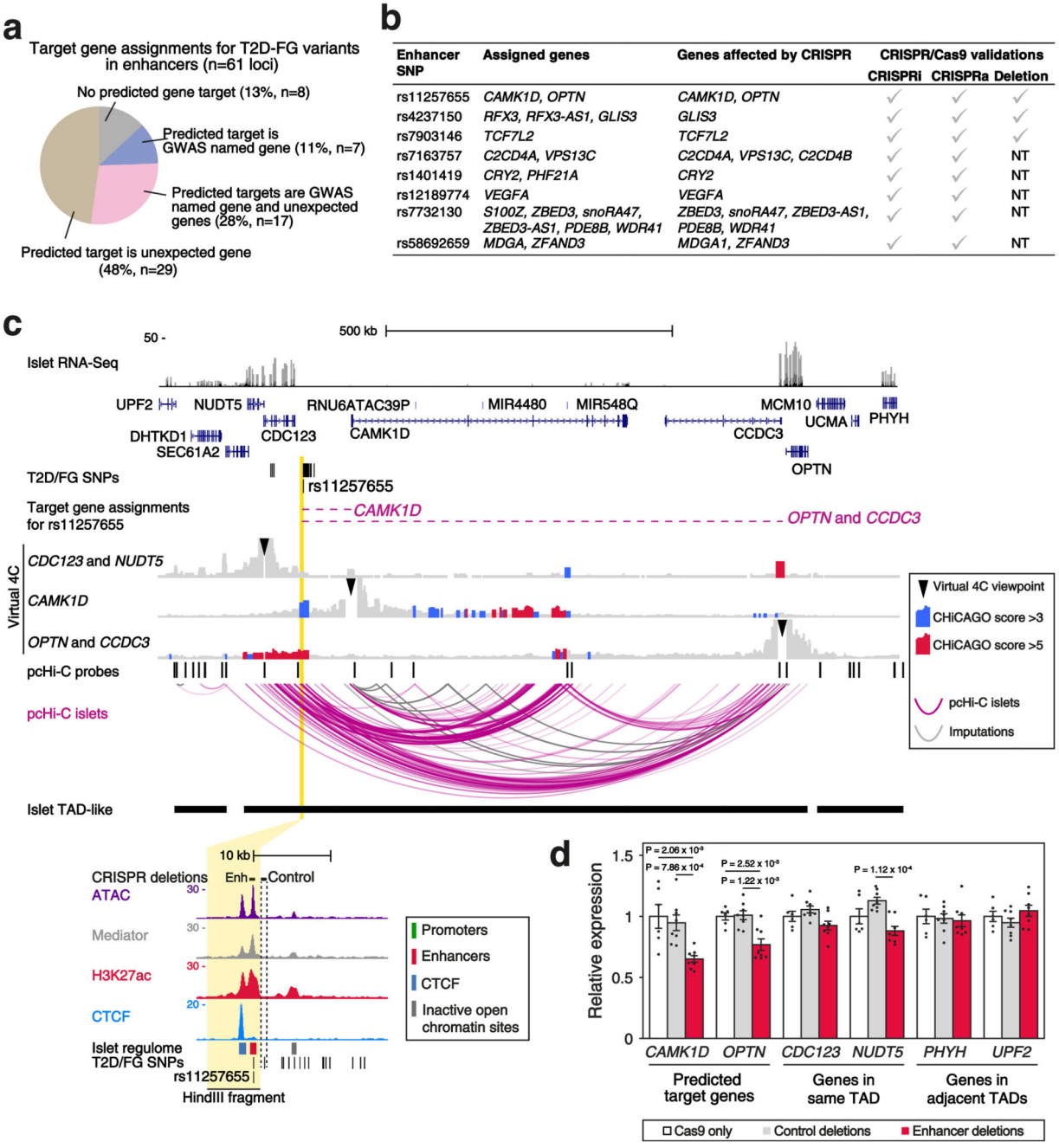


Figure 2. Identification of target genes of islet enhancers.

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 Figure 3l,m for further details and evidence that imputed assigments are enriched in sub-threshold 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. Top: 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. Bottom: 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). 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 IQRs, notches are 95% confidence intervals of median, P values are from Wilcoxon's two-sided signed ranked tests. See also Supplementary Data Set 2.

Europe PMC Funders Author Manuscripts



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

Figure 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 (2 independent experiments with triplicates). Deletions: n = 8 (2 gRNA pairs in 2 independent experiments with biological duplicates). Bars are means ± 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.

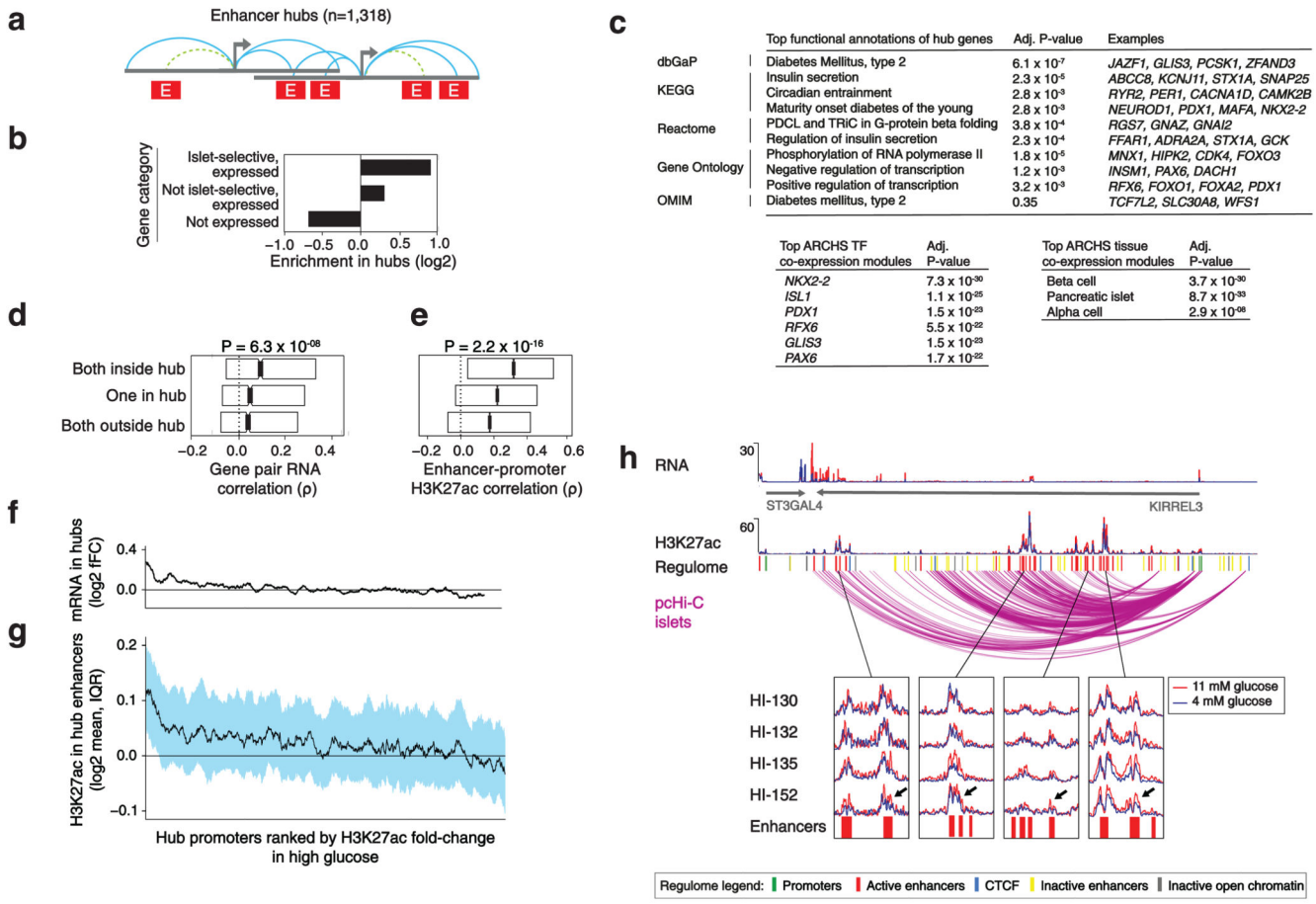


Figure 4. Tissue-specific enhancer hubs regulate key islet genes.

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 Figure 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 7 human islet donor samples at 4 vs. 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 in H3K27ac at high glucose, so that glucose-induced promoters are on the left of the X axis. (f) Median mRNA for genes associated with each hub. (g) Median glucose-dependent fold-

Europe PMC Funders Author Manuscripts

change of H3K27ac in enhancers from hubs connecting with each promoter, IQR values in blue shade. 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) vs. 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, Supplementary Data Set 5.

Europe PMC Funders Author Manuscripts

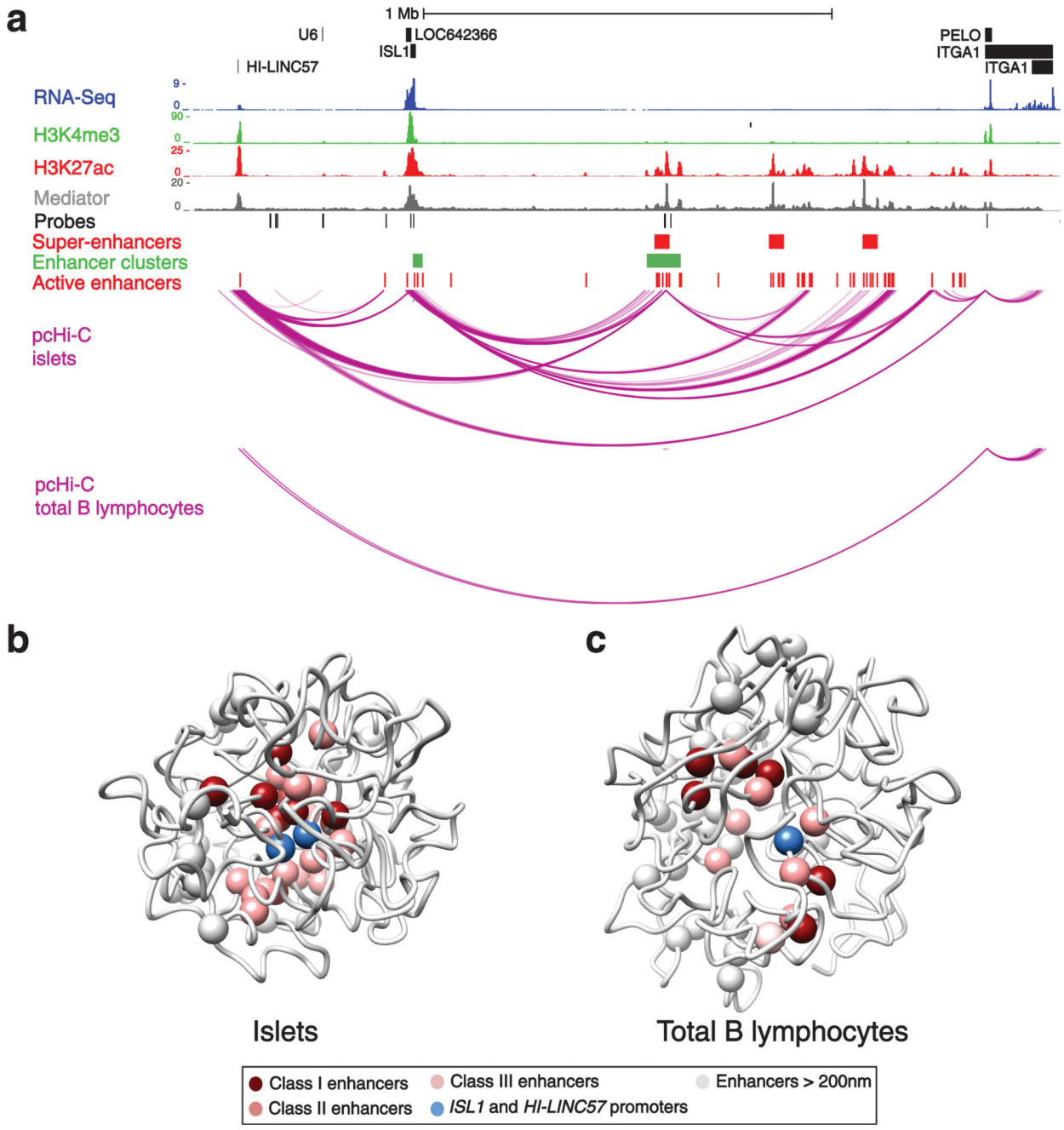


Figure 5. Tissue-specific topology of the *ISL1* enhancer hub.

a, Epigenome 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

Europe PMC Funders Author Manuscripts

in blue if they are within 200 nm of the *ISL1* promoter, or as white spheres if they are further 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 Figure 10b,c and Supplementary Videos 1 and 2.

Europe PMC Funders Author Manuscripts

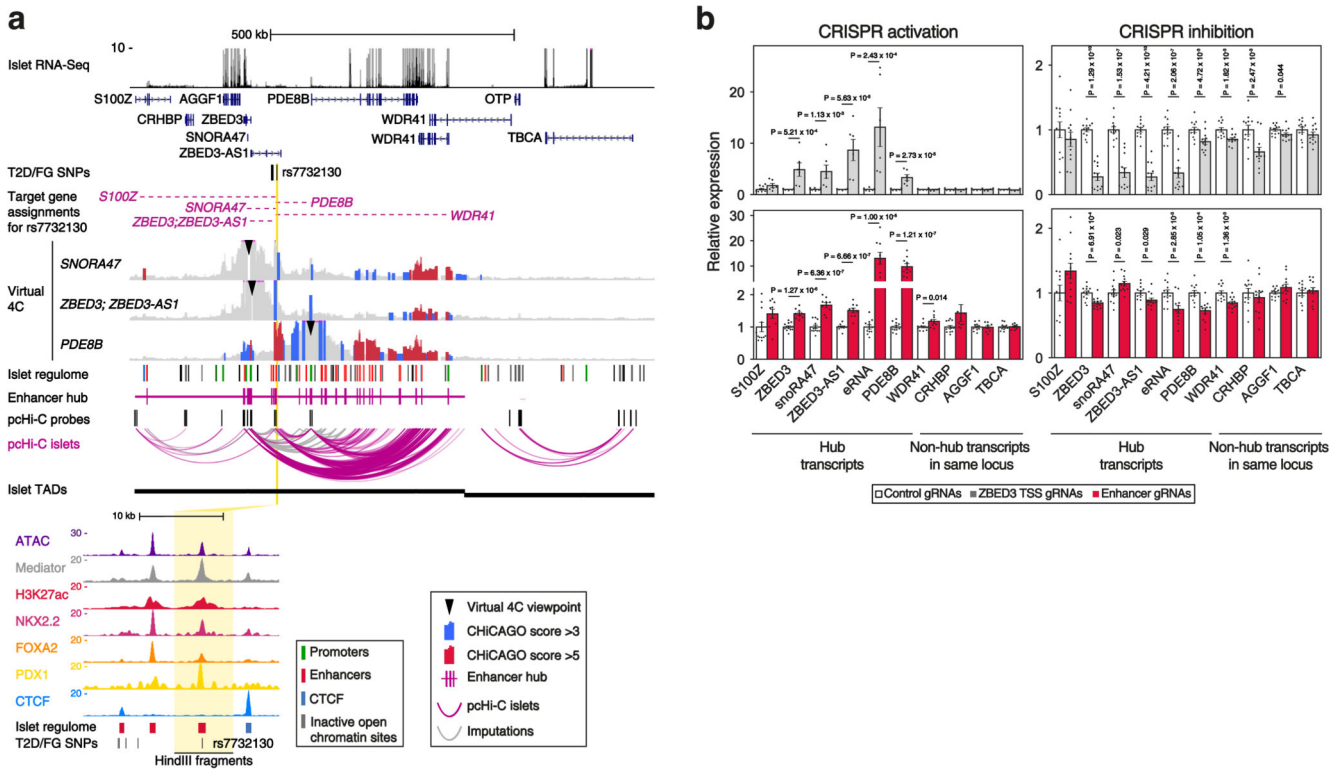


Figure 6. The *ZBED3* enhancer hub links an enhancer bearing a T2D SNP with multiple target genes.

a, pcHi-C and virtual 4C representations from pooled islet samples for three viewpoints (see also Supplementary Figure 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 Figure 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 means ± s.e.m. of all gRNAs combined per target region (enhancer CRISPRa: 3 gRNAs, CRISPRi: 4 gRNAs, all n=3 independent experiments). Statistical significance: two-tailed Student's *t*-test.

Europe PMC Funders Author Manuscripts

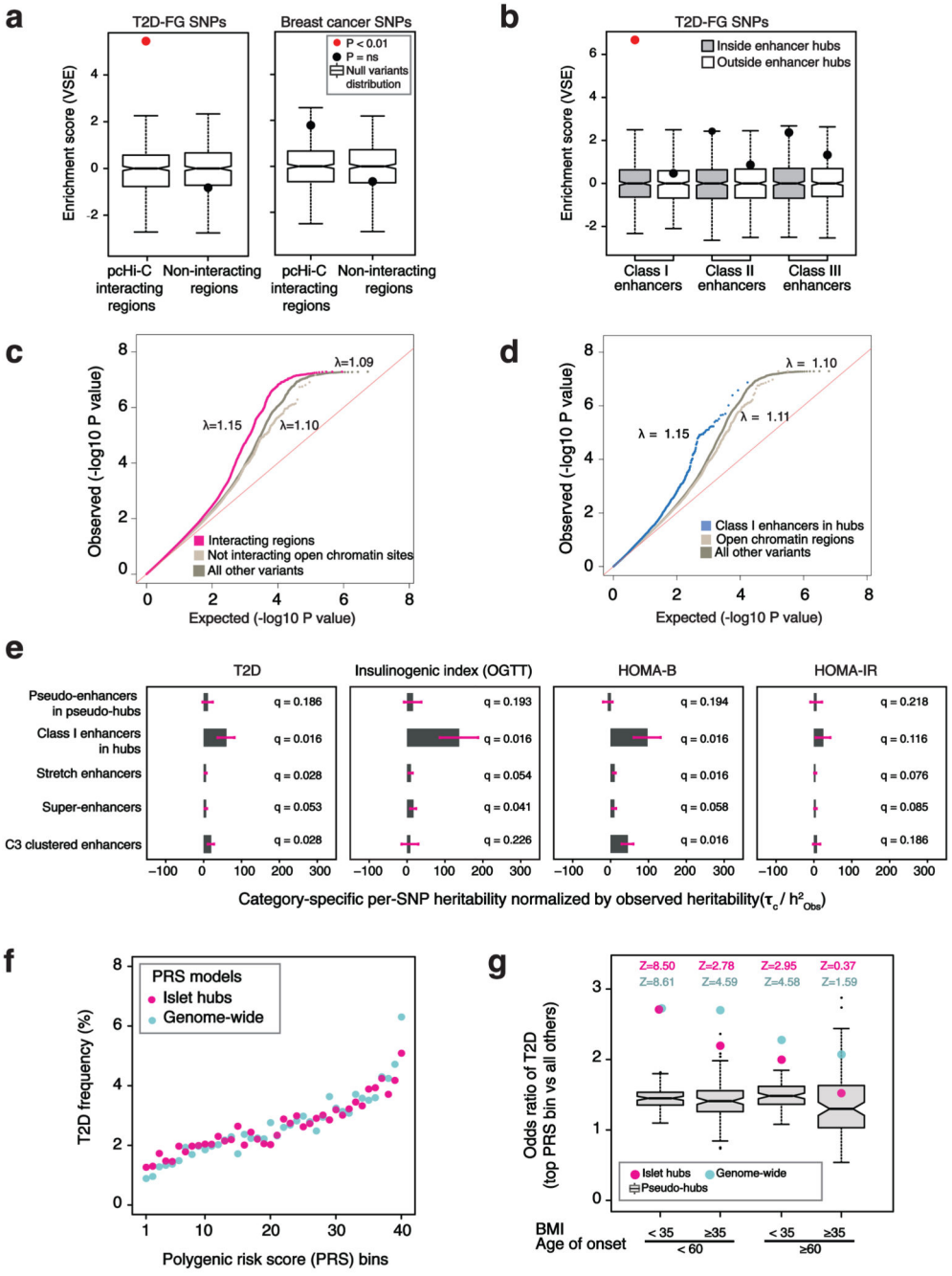


Figure 7. Islet hub variants impact insulin secretion and provide tissue-specific risk scores.

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). **b**, T2D and FG GWAS-significant variants are selectively enriched in hub class I islet enhancers. Boxplots show median and IQR. **c**, Genomic inflation of T2D association P values for non-GWAS significant variants ($P > 5x10^{-8}$) from a T2D GWAS meta-analysis (12,931 cases, 57,196 controls) in islet high-

Europe PMC Funders Author Manuscripts

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 per-SNP heritability coefficients (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 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 vs. all other individuals, using islet hub (pink) or genome-wide models (green), stratified by BMI and T2D age of onset. Boxplots show ORs for PRS from 100 permutations of pseudo-hubs (IQRs). Z-scores are standard deviations of pseudo-hub averages. See also Supplementary Figure 15 and Supplementary Table 17.

Europe PMC Funders Author Manuscripts